International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 4 Ver. IV|| April 2018 || PP 26-29

Analysis of Hydrological Data Processing for Hydropower Plant Projects in Georgia

Paata Jokhadze[1], Aleksander Jokhadze[1]
[1]*(Faculty of Informatics and Control Systems, Georgian Technical University, Tbilisi, Georgia)*
Corresponding author; Paata Jokhadze

Abstract: *The article reviews the problems associated with the statistical processing of hydrological data time series, which we found during the study of a number of prospective HPP projects in Georgia. The case study of the projects was to determine the causes of a discrepancies between the actual values and the design values of the HPP parameters (design flow, installed capacity, installed capacity factor, average annual energy production) of already implemented new HPP projects. Due to these discrepancies, the payback period of the project increased, the actual value of the Installed Capacity Factor (ICF) is less than its design value. Studies of existing projects have shown that the reason for this is the ignoring of the main requirements of normative documents and generally accepted recommendations by some designers, due to which the design hydrological characteristics were not determined correctly.*
Keywords – *Data sampling, Design Flow, Entire assembly, Flow Duration Curve (FDC), HPP, Hydrological Data Time Series, Installed Capacity Factor (ICF), Run-of-River HPPs, Statistical Homogeneity.*

Date of Submission: 05-03-2018 Date of acceptance: 20-04-2018

I. Introduction

After collapse of Soviet Union, energy independence has become the main challenge of Georgia. Because of the lack of fossil fuels in the territory of Georgia, the main way to achieve energy independence is to develop hydro-energetics. The Government of Georgia approved the State program "Renewable Energy 2008" and published on the website of the Ministry of Energy a list of prospective HPP projects. This program initiated the next stage of the construction of new, mainly run-of-river HPPs, one of the priority tasks of which is to maximize the effective use of available water resources.

Studies of prospective HPP projects showed that designers made incorrect decisions to solve the problem of effective use of available water resources; the direct evidence of this is the revealed discrepancy between the actual values and the design values of the HPP parameters (installed capacity factor, average annual energy production) of already implemented HPP projects. The designers mainly tried to maximize the use of water flow of rivers during flooding period and they blindly increased the design discharge of turbines (installed capacity), not considering the many interrelated other factors. Furthermore, this approach will lead to increase losses due to low turbine efficiency for a low discharge, or due to hydropower unit downtime for insufficient discharges of turbine. As a result, the average annual production of energy decreases, i.e. profitability of HPP decreases, and the payback period of the project increases. Our studies have shown that the reason for this is the ignorance from some designers of the basic requirements of actual normative documents and the ignorance of general accepted recommendations described in the manuals, therefore, the design hydrological characteristics were not determined correctly. Unfortunately, it should be noted that not only the designers of Georgian companies made this mistake, but also the designers of companies from European countries (we deliberately do not mention the names of companies).

II. Hydrological Data Conversion Methods

Many hydrological characteristics of the river (average annual, maximum and minimum water flow, the depth of runoff for high water or flood period, the duration of flood or low water periods, etc.) depend on a number of factors, the degree of influence of each of which is almost impossible to take into account. Thus, the characteristic under study should be considered as a random variable (or a random process), and methods of probability theory and mathematical statistics can be used to determine it. In the practice of hydrological calculations, there are often problems when it is necessary to determine the most probable values of the hydrological characteristics that will occur in the future, for example, it is very important to know these values for determining the technical parameters of the prospective HPP. On the correct determined results significantly depends profitability of the project. For example, the installed capacity, which depends on the design flow of the HPP, the degree of exceedance of which can be determined using the Flow Duration Curve (FDC). The selection of the design flow depends, primarily, on the available flow (hydrology) at the site. For central-grid connected run-of-river projects the optimum design flow is usually close to the flow that is equaled or exceeded about 30% of the time [I], in Georgia the value of exceedance of design flow usually is equal about 22-25% of the time. The solution of such problems is impossible without the use of probabilistic methods. These methods of processing hydrological data time series are described in normative documents and manuals, which give a clear sequence of procedures.

The first step is the selection of hydrometric observation station on the river or, if there is not one, on the river-analogue. In both cases, we will assume that the hydrometric observation station is located on the river-analogue for the projected site. In accordance with the requirements of actual normative document (Set of Rules 33-101-2003 *Determination of design hydrological performance* [II]), when choosing a river-analogue, the following conditions must be taken into account:

- Uniformity of type of flow of the river-analogue and the river under investigation;
- Geographical proximity of the watersheds location;
- Homogeneity of flow formation conditions, similarity of climatic conditions, homogeneity of soils (grounds) and hydrogeological conditions, a close degree of lacustrine, forestation, swampiness and plowing of catchments;
- The average height of the watersheds should not differ significantly, for mountainous and semi-mountainous areas, slope exposition and hypsometry should be taken into account;
- The catchment areas of rivers should not differ more than 10 times and their average heights (for mountain rivers) - not more than 300 m;
- Absence of factors significantly distorting the natural river flow (regulation of flow, water discharge, withdrawal of runoff for irrigation and other needs).

In the projects that we studied, these requirements were not met, which is one of the reasons for the undesirable results.

The second step is the preprocessing of initial hydrological data time series for the cross-section of the selected hydrometric station, before determining the design hydrological characteristics, it is necessary to check the initial hydrological data time series and select from them a sample of data in accordance with the defined order of actions listed below:

- Determination by average daily flows of average monthly and average annual flows;
- Assessment of the representativeness of a series of hydrological observations for the selected n years;
- Assessment of the sufficiency of the duration of the hydrological observation period;
- Checking the randomness of the model, which describes the structure of time series of hydrological data;
- Assessment of homogeneity and stationarity of the initial hydrological data time series.

These actions are very important for the accuracy of determined values of hydrological characteristics; their ignoring may cause undesirable results.

III. Method of Preprocessing of Initial Hydrological Data according to Normative Documents

Below are given procedures for preprocessing of initial hydrological data. The average values of the flows (average monthly, average annual, etc.) can be determined from the following equation:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{1}$$

Where n is sample size and x_i - sample unit.

In the time series of the annual flow of rivers, there is an alternation of groups of low-water and high-water years. So-called "water flow cycles" are formed. One cycle of water flow includes low-water years and high-water years phases of the river. Duration of cycles on rivers is not equal - it varies. If the series of hydrological observations does not contain complete cycles, then the values of the calculated hydrological characteristics will be incorrect, the initial hydrological data time series is not representative. Thus, the verification of the representativeness of a series of hydrological observations is a necessary condition.

To verify the representativeness of a series of hydrological observations, it is possible to apply the method of difference integral curves. This method is a simple graphical method and provides a visual representation of the cycles of fluctuations in annual flow of river. For construction the graph of the difference integral curve, it is necessary to calculate the norm of river flow and the modular ratio of average annual flow, using the following equations:

$$\overline{Q}_0 = \frac{\sum_{i=1}^{n} \overline{Q}_i}{n}, \tag{2}$$

$$k_i = \frac{\overline{Q_i}}{\overline{Q_0}}, \tag{3}$$

where: $\overline{Q_i}$ is an average annual flow, $\overline{Q_0}$ is the norm of river flow and k_i is the modular ratio of average annual flow. The graph of the difference integral curve can be constructed according to the following equation:

$$\varphi(t) = \sum_{i=1}^{t}(k_i - 1). \tag{4}$$

When $\varphi'(t) > 0$, then $\overline{Q_i} > \overline{Q_0}$ and this part of the graph corresponds to the high-water years period of the river and when $\varphi'(t) < 0$, then $\overline{Q_i} < \overline{Q_0}$ and this part of the graph corresponds to the low-water years period of the river. From the constructed graph, it is necessary to select a hydrological observation period that contains complete cycles, the resulting sample will satisfy the requirements of representativeness.

The representativeness of the hydrological observation period can be verified by an analytical method for which the average modular ratio is calculated in accordance with the equation given below:

$$k_{period} = 1 + \frac{I_{end} - I_{start}}{n_{period}}, \tag{5}$$

where I_{end} and I_{start} are the ordinates of the difference integral curve at the beginning and at the end of the selected observation period, n_{period} is the number of years in this period. If the selected observation period contains complete cycles, then the following condition is satisfied:

$$k_{period} \approx 1 \tag{6}$$

and the selected sample is representative.

The next step is to assess the sufficiency of the duration of the hydrological observation period. The duration of the hydrological observation period is considered sufficient if it is representative and the relative standard error of calculated average value of the hydrological characteristic does not exceed 10% for annual and seasonal flows, 20% for maximum and minimum flows and 15% for the coefficient of variation. According to the method of moments, the biased estimates of the coefficients of variation and skewness can be determined from the following equations:

$$\tilde{C}_v = \sqrt{\frac{\sum_{i=1}^{n}(k_i - 1)^2}{n-1}}, \tag{7}$$

$$\tilde{C}_s = \frac{n \cdot \sum_{i=1}^{n}(k_i - 1)^3}{\tilde{C}_v^3 \cdot (n-1) \cdot (n-2)}. \tag{8}$$

If $\tilde{C}_v < 0{,}6$ and $\tilde{C}_s < 1{,}0$, then the values of the coefficients of variation and skewness can be determined in accordance with equations (7) and (8), otherwise their values should be refined by the maximum likelihood method (the essence of this method in this article will not be concretized).

The relative standard errors in calculating the average annual flow rate and coefficient of variation can be determined from the following equations:

$$\varepsilon_{\overline{Q}} = \frac{C_v}{\sqrt{n}} \cdot 100\%, \tag{9}$$

$$\varepsilon_{C_v} = \sqrt{\frac{(1 + C_v^2)}{2 \cdot n}} \cdot 100\%. \tag{10}$$

To check the randomness of the model describing the structure of hydrological data time series, several criteria are used for which the hypothesis that the structure of a given hydrological time series corresponds to a model of a random variable is taken as the null hypothesis. All the applied criteria equally confirm or disprove the null hypothesis, but in the practice of determining hydrological characteristics, the criterion for the significance of the autocorrelation coefficient are most often used. According to this criterion, the null hypothesis is confirmed if the following condition is satisfied:

$$r(1) < \sigma_{r(1)} \cdot t_{2\alpha}, \tag{11}$$

where $t_{2\alpha}$ is the normalized ordinate of the normal distribution law with a significance level of 2α (usually $2\alpha = 5\%$ or $2\alpha = 10\%$); The coefficient of autocorrelation and the absolute error of its calculation are determined by the following equations:

$$r(\tau) = \frac{\sum_{i=1}^{n-\tau}\left[(x_i - \bar{x}) \cdot (x_{i+\tau} - \bar{x})\right]}{(n-\tau) \cdot \sigma^2}, \tag{12}$$

$$\sigma_{r(\tau)} = \frac{1 - r^2(\tau)}{\sqrt{n-\tau-1}}. \tag{13}$$

The last and most important step in the preprocessing of hydrological data is the verification of the selected sample for statistical homogeneity. If the structure of the selected hydrological time series corresponds to a random variable model, then each element of this series should belong to the entire assembly (the general population), which describes the unchanged conditions for the formation of river flow. In other words, this means that the conditions for the formation of each element of any sample have not changed due to natural disasters and / or have not changed due to anthropological influences. If the statistical characteristics (for example average values of river flow and dispersion) of different parts of the hydrological time series differ significantly from each other, then the entire hydrological time series is not homogeneous. This means that the conditions for the formation of elements of different parts of the hydrological time series are different and the conditions for the formation of river flow have changed. It is obvious that the use of such a hydrological time series to determine the hydrological characteristics is incorrect, since the results obtained will differ from the actual values. Thus, before determining the hydrological characteristics it is necessary to check the homogeneity of the selected sample. The homogeneity of the sample of hydrological data is verified by so-called parametric and nonparametric criteria, the Student and Fisher's parametric criteria are most often used in hydrology.

According to the Student's criterion (t-criterion), the entire sample is divided into two subsamples $(x_1, x_2, ..., x_m)$ and $(y_1, y_2, ..., y_n)$ by the number of elements m and n, if we assume that both subsamples belong to the same general population then there should be an insignificant difference between these subsamples, in other words, the difference between their average values should be statistically insignificant. Based on this difference, the empirical value of Student's statistics is calculated according to the following equation:

$$t^* = \frac{\bar{x} - \bar{y}}{\sigma_{\bar{x}-\bar{y}}}, \tag{14}$$

where $\bar{x}$ and $\bar{y}$ are the average values of the subsamples, and $\sigma_{\bar{x}-\bar{y}}$ is the standard deviation of the difference $(\bar{x} - \bar{y})$, which is determined by the following equation:

$$\sigma_{\bar{x}-\bar{y}} = S \cdot \sqrt{\frac{(m+n)}{m \cdot n}}, \tag{15}$$

where S is the empirical estimate of the standard deviation $\sigma_{\bar{x}-\bar{y}}$ which is defined by the following equation:

$$S = \sqrt{\frac{(m-1) \cdot \sigma_x^2 + (n-1) \cdot \sigma_y^2}{m+n-2}}, \tag{16}$$

where:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m}}, \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}. \tag{17}$$

Taking into account the last three equations, the statistics can be determined by the following equation:

$$t^* = \frac{(\bar{x} - \bar{y})}{S} \cdot \sqrt{\frac{m \cdot n}{m+n}}. \tag{18}$$

The hypothesis of homogeneity of the sample (null hypothesis) is not refuted if the following inequality is valid:

$$\left|t^*\right| < t_{2\alpha}, \tag{19}$$

where $t_{2\alpha}$ is the theoretical value of Student's statistics with a significance level of 2α (usually $2\alpha = 5\%$ or $2\alpha = 10\%$).

The Fisher's criterion (F-criterion) is a criterion for estimating the equality of two variances. According to this criterion, like the Student's criterion, the entire sample is divided into two subsamples $(x_1, x_2, ..., x_m)$ and $(y_1, y_2, ..., y_n)$ by the number of elements m and n, if we denote the largest value between two variances σ_x^2 and σ_y^2 by D_1 and the smallest value by D_2, then the empirical value of Fisher's statistics can be determined by the following equation:

$$F^* = \frac{D_1}{D_2}. \quad (20)$$

The hypothesis of homogeneity of the sample (null hypothesis) is not refuted if the following inequality is valid:

$$F^* < F_{2\alpha}, \quad (21)$$

where $F_{2\alpha}$ is the theoretical value of Fisher's statistics with a significance level of 2α (usually $2\alpha = 5\%$ or $2\alpha = 10\%$) And with degrees of freedom $m-1$ and $n-1$.

All of the above procedures for preprocessing hydrological data time series sometimes need to be performed several times, until the selected sample satisfies all of the above requirements; only after this, it is possible to begin to determine the design hydrological characteristics. All this requires hard work and a lot of time; this explains that in most cases, designers skip this stage of data processing.

IV. Conclusion

In fact, the formation of river flow is a stochastic process that depends on many factors, it is impossible to predict when and what natural conditions and phenomena will affect the formation of river flow. Thus, it is impossible to determine the probabilistic values of the hydrological characteristics without applying the methods of mathematical statistics, the use of which in the practice of hydrological calculations began in the 40s of the XX century. Over the years, the methodology has been improved, normative documents and methodological recommendations have been developed, especially the publications of the methodological recommendations of the State Hydrological Institute of the Russian Federation, which clearly describe the sequence and importance of each step of the data processing procedures. Unfortunately, we could not find similar documents in English, which were developed and published in other countries.

In presented article, we discussed the mistakes made by some designers that we found in the part of preprocessing of hydrological data time series. We are confident that interested parties will also be able to detect such mistakes in HPP projects if they have well-written methodologies and normative documents. During the study of prospective HPP projects, we also found mistakes in the part of determination of the main design hydrological characteristics for the cross-section of the selected hydrometric station and in the part of conversion of characteristics for project cross-section of HPP; also, in the part of determining the technical parameters of HPPs. We do not discuss them in this article, since we believe that in its framework this topic refers to another article.

Taking into account the current situation, we came to a decision to create software for complete hydrological calculations from preprocessing of the hydrological data time series to the determination of the main technical parameters of HPPs taking into account the requirements of normative documents and methodological recommendations. We hope that this software will significantly increase the reliability of the calculated results and maximally reduce the deviation of the estimated values of important technical parameters of HPPs from the actually expected values, such as the installed capacity, the probable average annual energy production and the installed capacity factor.

References

I. Retscreen®International, Clean Energy Decision Support Centre *Retscreen® Software Online Manual* (Minister Of Natural Resources Canada 1997-2004)

II. СП 33-101-2003 Определение основных расчетных гидрологических характеристик (Госстрой РФ, Москва 2004) - (Set of Rules 33-101-2003 Determination of design hydrological performance (Gosstoroy RF, Moscow 2004))

Paata Jokhadze "Analysis of Hydrological Data Processing for Hydropower Plant Projects in Georgia" International Journal of Engineering Science Invention (IJESI), vol. 07, no. 04, 2018, pp 22-26