CHARACTERISTICS OF BASIN INFLOWS A STATISTICAL ANALYSIS FOR LONG-TERM/MID-TERM HYDROTHERMAL SCHEDULING

by

Elis SUTLOVIĆ [a*], Snježana ČUJIĆ ĆOKO [b], and Ivan MEDIĆ [a]

[a] Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia
[b] Croatian Transmission System Operator, Split, Croatia

Original scientific paper
DOI: 10.2298/TSCI1403799S

The presented paper focuses on the characteristics of reservoir inflows and the appropriate inflow model for long-term/mid-term hydrothermal scheduling. The goal was to find the type of distribution that best fits the observed series of monthly and weekly average inflows in most cases for a model which considers the inflows as independent random variables without time correlation. Also, the objective was to explore the correlation between the inflows during time periods (for weekly and monthly intervals, respectively), and to investigate whether the more complex model of reservoir inflow as a dependent random variable is advisable for optimal long-term/mid-term hydrothermal scheduling. Differences in the characteristics of monthly and weekly inflows, which have been noticed during the analysis, are discussed. Numerical results are presented.

Key words: *power system planning, inflow models, statistical analysis, long-term/mid-term hydrothermal scheduling*

Introduction

In a hydrothermal power system with a significant percentage of hydrogenation, the reservoir systems operation has a great impact on the operation cost in the traditional approach or on profits maximizing in the market approach. For optimal hydrothermal scheduling, natural inflows into reservoir systems present one of the major uncertainties of the planning process. The uncertainty increases as longer planning horizons are considered, so stochastic fluctuations of water inflow should be taken into account whenever the study period is greater than the short term (typically one week). In a medium term and in a long term storage regulation, the operating horizon typically spans from several months to several years. A long-term/mid-term hydrothermal scheduling problem is essentially stochastic, due to the uncertainty about future inflows.

Over the past years a number of methods have been developed for solving the long-term/mid-term hydro scheduling problem with different approaches to randomness of inflows. In all these analyses, the study period is divided in discrete time sub-periods, usually one week or one month long. The natural inflows in water basins, as well as river streamflows, can be presented through different models. The simplest one is to treat the inflow at each particular location for each time interval of the study period like a known value, and then proceed with the deterministic procedure [1-5]. Due to inflows uncertainty, this approach

* Corresponding author; e-mail: sutlovic@fesb.hr

requires many calculations with various inflows scenario. Often, it is not sufficient to consider just sequences of average inflows or sequences of some critical (*e. g. worst-case*) values, but a wide range of possible values of inflows that can happen in the future should be taken into account. When the deterministic inflow model is used, it is possible to use less complex optimization methods for a long-term/mid-term hydrothermal scheduling, *e. g.* linear programming, quadratic programming, dynamic programming, *etc*. Simpler optimization methods can take into consideration a number of additional influencing parameters, which in complex optimization methods often make an additional limitation. The disadvantage of a deterministic approach is that every combination of input data produces different results. The analysis of the results requires additional effort and is not always easy and unambiguous. Another way or a stochastic approach tries to incorporate some of the knowledge about the uncertainty into variables and input parameters used in optimization or simulation models. The independent model, which considers the inflows as independent random variables without time correlation, is commonly used. Finally, a more sophisticated model is the dependent model, when the inflows are assumed as correlated in time according to a Markov chain [6-10].

This article emerged during the development of a new long-term/mid-term hydro-scheduling method. The main goal was to find the appropriate inflow model for the long-term/mid-term hydrothermal scheduling that accurately represents the real conditions. The statistical analysis of historical measured inflow data focused on two main issues: firstly, on the selection of the theoretical probability distribution that best describes the observed series of monthly and weekly average inflows in most cases, and secondly, on the investigation of time correlations between different sequences of monthly and weekly average inflows. The differences in the characteristics of monthly and weekly inflows, which have been noticed during the analysis, are discussed. The analysis of observed inflows aims to explore whether a more complex inflow model is really necessary and how well the inflow model describes the actual data.

Inflow modeling

The first step in the development of a statistical inflow model is to extract the fundamental information about the joint distribution of flows at different times from historical measured inflow data at pertinent locations.

Fitting a probability distribution function

In general, distributions of hydrological data are positively skewed, having a lower bound near zero and an almost unbounded right-hand tail. One of the goals was to find the type of distribution that best fits particular observations data sets in most cases. The intention was to fit several simple continuous distributions, commonly used in water resources planning, to a set of observed values of the random inflow data. The probability density function (PDF) of chosen distributions corresponds to characteristics of the histogram of recorded inflows. Also, for each chosen distribution there is a well-known method which can be used to estimate distribution parameters based on available sample data [9, 11-14].

Normal distribution. the PDF of a normal random variable x is:

$$f(x) = (1/\sigma\sqrt{2\pi})\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } (-\infty < x < +\infty) \tag{1}$$

with two parameters, mean (μ) and standard deviation (σ).

Lognormal distributions. The random variable x has a two-parameter lognormal distribution (LN2) if the natural logarithm of x, $y = \ln(x)$, has a normal distribution. The PDF of x is:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2}, \text{ for } x > 0 \tag{2}$$

,

The parameter μ_y determines the scale of the distribution, whereas σ_y determines the shape of the distribution.

The three-parameter lognormal distribution (LN3), obtained when $\ln(x - \tau)$, is described by a normal distribution, where τ is location parameter and $x \geq \tau$. The PDF of x is:

$$f(x) = \frac{1}{(x-\tau)\sigma_y\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(x-\tau) - \mu_y}{\sigma_y}\right)^2}, \text{ for } x \geq \tau \tag{3}$$

Gamma distribution (2-Parameter). For a gamma random variable x, PDF is:

$$f(x) = x^{\alpha-1}\left[\frac{\mathrm{e}^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\right], \text{ for } x > 0 \text{ and } \alpha, \beta > 0 \tag{4}$$

where $\Gamma(\alpha)$ is the gamma function:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\mathrm{e}^{-x} dx \tag{5}$$

Gamma function $\Gamma(\alpha)$, for positive integer α, can be substituted for $\Gamma(\alpha) = (\alpha - 1)!$ The parameter α determines the shape of the distribution while β is the scale parameter. If $(x - \gamma)$ has a gamma distribution, where γ is constant, the distribution of x is a three-parameter gamma distribution or the Pearson type 3 distribution.

Log-Pearson type 3 distribution. The log-Pearson type 3 distribution (LP3) describes a random variable with logarithms having a Pearson type 3 distribution. The LP3 distribution has a PDF given by:

$$f(x) = \frac{1}{x|\beta|\Gamma(\alpha)}\left(\frac{\ln(x) - \gamma}{\beta}\right)^{\alpha-1} \mathrm{e}^{-\frac{\ln(x)-\gamma}{\beta}}, \text{ for } \gamma \leq \ln(x) < \infty \text{ and } \alpha, \beta > 0 \tag{6}$$

Gumbel Max distribution. The Probability density function of the Gumbel Max or the Extreme Value type I distribution is given by:

$$f(x) = \frac{1}{\sigma} \mathrm{e}^{-\frac{x-\mu}{\sigma}} \mathrm{e}^{-\mathrm{e}^{-\frac{x-\mu}{\sigma}}} \tag{7}$$

where μ is the location parameter and σ is the scale parameter.

Weibull distribution (2-parameter). The two-parameter Weibull distribution PDF is given by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \mathrm{e}^{-\left(\frac{x}{\beta}\right)^{\alpha}} \text{ for } x \geq 0 \text{ and } \alpha, \beta > 0 \tag{8}$$

where α is the shape parameter and β – the scale parameter.

Inflow models

We are looking for a simple and robust model to generate inflows sequences in different time stages that describe selected characteristics of the historical flows. The selected model should correctly reproduce the seasonality of hydrological processes at a single site. By the lag-1 autoregressive Markov model, commonly used dependent model for long-term hydro scheduling [6, 9, 10], the inflow model for normally distributed monthly inflows can be produced by:

$$i_{j,m} = \mu_m + r_m \left(\frac{\sigma_m}{\sigma_{m-1}} \right) (i_{j,m-1} - \mu_{m-1}) + z_m \sigma_m \sqrt{1 - r_m^2} \tag{9}$$

where z_m is a standard normal random variable with zero mean and unit variance. Equation (9) generates inflows for a particular location. The inflow $i_{j,m}$ for months m = 1, 2, ...,12 and years j = 1, …, n can be evaluated if the previous inflows $i_{j,m-1}$, the normal parameters (μ_m and σ_m) for each months, and the lag-1 correlation r_m between two successive months are known. These parameters used in eq. (9) can be estimated using the historical data, according to (10)-(12):

$$\mu_m = \frac{1}{n} \sum_{j=1}^{n} i_{j,m} \tag{10}$$

$$\sigma_m = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} \left(i_{j,m} - \mu_m \right)^2} \tag{11}$$

$$r_m = \frac{\sum_{j=1}^{n} (i_{j,m} - \mu_m)(i_{j,m-1} - \mu_{m-1})}{\sigma_m \sigma_{m-1}} \tag{12}$$

The inflow $i_{j,m}$ will also be normally distributed because sums of independent normally distributed random variables ($i_{j,m-1}$ and z_m) are normally distributed. The relationship between normal distributed variables is linear. The conditional mean of inflow in month m, given the inflow in month m – 1:

$$\mu_{m|m-1} = \mu_m + r_m \left(\frac{\sigma_m}{\sigma_{m-1}} \right) (i_{j,m-1} - \mu_{m-1}) \tag{13}$$

is the deterministic component in eq. (9), while the conditional standard deviation:

$$\sigma_{m|m-1} = \sigma_m \sqrt{1 - r_m^2} \tag{14}$$

multiplied by z_m is a random component which represents the random deviation of inflows independent of the correlation with the previous month's inflow. For r_m = 0, eq. (9) will represent the independent model of a monthly inflow.

If inflows are log-normally distributed random variable, eq. (9) can be applied to the logarithms of the inflows, with random numbers (z_m) from the normal distribution:

$$\ln(i_{j,m}) = \mu'_m + r_m'\left(\frac{\sigma'_m}{\sigma'_{m-1}}\right)[\ln(i_{j,m-1}) - \mu'_{m-1}] + z_m\sigma'_m\sqrt{1-r'^2_m} \tag{15}$$

where μ_m', σ_m' and μ_{m-1}', σ_{m-1}' are the mean and standard deviation of $\ln(i_{j,m})$ and $\ln(i_{j,m-1})$, respectively. They can be expressed using the mean and standard deviation of the monthly flow $i_{j,m}$ and $i_{j,m-1}$ [15, 16] with the equations:

$$\mu'_i = \ln\frac{\mu_i^2}{\sqrt{\sigma_i^2+\mu_i^2}}, \qquad \sigma'^2_i = \ln\left(1+\frac{\sigma_i^2}{\mu_i^2}\right),\ i = m-1, m \tag{16}$$

The correlation coefficient r_m' of $\ln(i_{j,m})$ and $\ln(i_{j,m-1})$ can be estimated, at least in theory, from the correlation coefficient r_m. However, when dealing with lognormal random variables, whose logs are normally distributed and linearly correlated on the log scale, then the correlation of lognormals is no longer linear on the original scale. The relationship between the variables in general becomes a curve. Thus, more efficient estimates of linear correlation of lognormal variables are generally obtained by measuring the correlation on the log scale [9].

If the normal variates $\ln(i_{j,m})$ follow the autoregressive Markov model given by eq. (15), then the corresponding inflows $i_{j,m}$ follow the model:

$$i_{j,m} = (i_{j,m-1})^{r_m'\frac{\sigma'_m}{\sigma'_{m-1}}}\exp\left[\mu'_m - r_m'\frac{\sigma'_m}{\sigma'_{m-1}}\mu'_{m-1}\right]\exp\left[z_m\sigma_m'\sqrt{1-r'^2_m}\right] \tag{17}$$

The same autoregressive models given by eqs. (9) and (17) can be applied to weekly inflows, if we assume that m is the index of the week (m = 1, 2, ..., 52), and mean, standard deviation and the lag-1 correlation between two successive weeks are known.

Required length of observation in dependence on time interval

A long-term/mid-term hydrothermal scheduling analysis requires a reasonably long period of observation of hydrological data in order to make use of historical data as a representation of the hydrological characteristics of the river basins. When statistical methods are applied in hydrology, it is advisable to use the data series of at least 30 years. When using shorter sequences of data it is necessary to evaluate the representativeness of the given series, and the reality of the results.

Whether the available series of hydrological data are long enough can be evaluated according to the standard error of the coefficient of variation. The standard error of the variation coefficient for homogeneous set of observed data can be expressed [17] as:

$$\sigma_{cv} = cv\sqrt{\frac{1+2cv^2}{2n}} \tag{18}$$

or slightly stricter [18] as:

$$\sigma_{cv} = \frac{cv}{\sqrt{2(n-1)}}\sqrt{1+3cv^2} \tag{19}$$

where cv is the coefficient of variation (σ/μ) for the sequence of n members. The sample estimate of the coefficient of variation for any time interval m can be evaluated from eq. (10) and (11). If σ_{cv} is lower than 0.10, available data series can be considered as being sufficiently long for a relevant analysis [18].

It is interesting to notice how the choice of base time interval (month or week) affects the required length of observation. As the basic period is longer, the standard deviation of a series of mean values of seasonal inflows is lower. The average standard deviation decreases as the base time interval increases. Considering that the average mean value of seasonal inflows is independent of the basic time step, it follows that the increase of the base period of time reduces the average coefficient of variation. By reducing the coefficient of variation, with the same total period of observation n in the formula (18) or (19), the standard error of the coefficient of variation is reduced. It follows that, with the same allowed threshold level of standard error of the variation coefficient, the increase of the base period reduces the total required observation period n. Thus, if a month is chosen as the basic time unit, a smaller series of input data are required as opposed to a week being the basic time unit. The presented conclusions can be clearly illustrated by the conducted case study.

Coefficient of skewness in dependence on time interval

The asymmetry of a distribution is often measured by its coefficient of skewness. The sample estimate of the skewness coefficient of seasonal inflows for time interval m is:

$$cs_m = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{i_{j,m}-\mu_m}{\sigma_m}\right)^3 \tag{20}$$

The choice of the basic computational time interval affects the distribution asymmetry. As the basic period is longer, the coefficient of skewness is lower. Therefore, the expected skewness coefficient of the distribution of weekly average inflows is slightly higher than the expected skewness coefficient of distribution of monthly average inflows.

Case study

Observed data

The Cetina River is a typical karst river which, due to the water quantities and built hydroelectric facilities on the river basin, constitutes the most important hydropower system in the Republic of Croatia [19]. The artificial lake Peruca is the biggest water reservoir in the Republic of Croatia which considerably affects the Cetina flow regulation at the downstream power plants. The gaged inflows in the upstream accumulation lake Peruca are natural data, while many other downstream inflows and streamflows are, in some way, influenced by discharges from the reservoir Peruca or other storage reservoirs in the Cetina basin.

Since 1960, when the accumulation lake Peruca was built, monitoring and measurement of hydrologic data have been carried out every day in the same manner at the gaging station. The reservoir inflow is estimated from the water budget computation which, beside gaged releases and storage level, takes into account estimated losses due to evaporation and other reasons. However, due to various occasions, primarily because of a few large construction works in the wider catchment area during the seventies and eighties of the previous century, inflows data recorded in the whole period cannot be treated as homogeneous and stationary series. Stationarity means that statistical properties of the analysed data do not change over time. Since significant changes of the observed inflow data caused by some construction projects could not be clearly identified, only the last sequence of mean daily inflow recorded from 1994 to 2012 will be considered as a homogeneous, stationary and continuous series.

From the continuous time series of daily average inflows in the period of 19 years, a 12 series of monthly average inflows and a 19 series of weekly average inflows were created. Thus, the monthly inflows are described by the 12 × 19 matrix, and weekly inflows by the 52 × 19 matrix. In addition, we analyzed and compared the data of the monthly average inflow into the reservoir Peruca recorded in a period of 53 years (described by the 12 ×53 matrix), having in mind the suspicion of the homogeneity of these sequences.

Statistical analysis of data series

Firstly, we checked if statistical methods can be applied to the resulting series of monthly and weekly average inflows. This included the testing of the parameters of the data series, as follows: the homogeneity, the detection of non-randomness, the checking of the mutual independence of the members, the verification of the series length requirements, *etc*.

Secondly, we quantitatively described the main features of the sets of monthly and weekly average inflows with descriptive statistics (mean, variance, standard deviation, coefficient of variation, coefficient of skewness, *etc*.). Table 1 shows the average value of means, standard deviations, coefficients of variations and coefficients of skewness of weekly average inflows and of monthly average inflows into the reservoir Peruca for the period of 19 years. For comparison, tab. 1 also displays the mean, standard deviation, coefficient of variation and coefficient of skewness for a series of annual average inflows for the same period.

Table 1. Summary results of the statistical analysis of the data series of weekly, monthly and annual inflows into the reservoir Peruca

Time interval	Week	Month	Year
n	19	19	19
Mean [m^{3}s^{-1}]	36.509	36.977	36.977
St. deviation [m^{3}s^{-1}]	20.119	16.109	8.527
Coeff. of variation	0.545	0.445	0.231
Coeff. of skewness	1.247	0.895	0.233
σ_{cv} eq. (18)	0.121	0.089	0.039
σ_{cv} eq. (19)	0.137	0.099	0.041

The average standard deviation (20.119) and the average coefficient of variation (0.545) of weekly average inflows were higher than the average standard deviation (16.109) and the average coefficient of variation (0.445) of monthly average inflows. The standard error of the coefficient of variation according to both equations is higher for the weekly average inflows then for the monthly average inflows. For example, the observed period of 19 years is long enough for monthly average inflows (σ_{cv} is less than 0.1) but is not long enough for weekly average inflows (σ_{cv} is greater than 0.1). For annual average inflows during the same period of 19 years the standard deviation, coefficient of variation and standard error of the coefficient of variation were significantly lower (σ_{cv} is approximately 0.04), which means that the period of 19 years is a sufficiently long period of observation for relevant statistical analysis of the annual inflows.

From tab. 1 it can be observed that the average skewness coefficient of 52 weekly mean inflows of the reservoir Peruca is 1.247, and the average skewness coefficient of 12 monthly mean inflows is 0.895. The skewness coefficient of the single series of annual mean inflows is 0.233. If we accept the common classification of coefficient of skewness, the average distribution of weekly mean inflows can be interpreted as highly skewed (*cs* greater than + 1), the average distribution of monthly mean inflows can be interpreted as moderately skewed (*cs* between 0.5 to 1), whereas the distribution of annual mean inflows is approximately symmetric (*cs* less than 0.25). The obtained values of the coefficient of variation and the coefficient of skewness of the data series of weekly, monthly and annual inflows into the reservoir Peruca are in accordance with the corresponding values obtained in other rivers and reservoirs [20-24].

Fitting the probability distribution function

Testing the compatibility of empirical data with the theoretical probability distribution function was performed with the Chi-Squared test (with significance level of 0.05) [11]. According to the results of the Chi-Squared test we rank the probability distribution for each period of observation.

Analysing monthly inflows we noticed that the empirical data in 10 months can be described with lognormal distributions (both with LN2 and LN3) and the LP3 distribution. In a few months, empirical data can also be described with one of the other tested distributions (Gamma, GumbelMax and Weibull). Figure 1 shows one typical graph comparing the probability density of the fitted distributions with a histogram of the data (frequencies are normalized).

Since the Chi-square statistics depends on the number of the histogram class in which the data are grouped, the width of classes has been varied. The modification of the classes width sometimes caused a slight change in the distribution rank, but generally, LN2, LN3 and LP3 were fit to a set of observations well enough to accept the hypothesis that the data follow the specified distribution.

Almost the same conclusions can be obtained when analysing the weekly inflows. The available data set for each week is smaller, so the fitting probability distributions are less precise. Nevertheless, both lognormal distributions and logPearson3 distribution, with well-estimated parameters, can be used to represent the observed values in most cases.

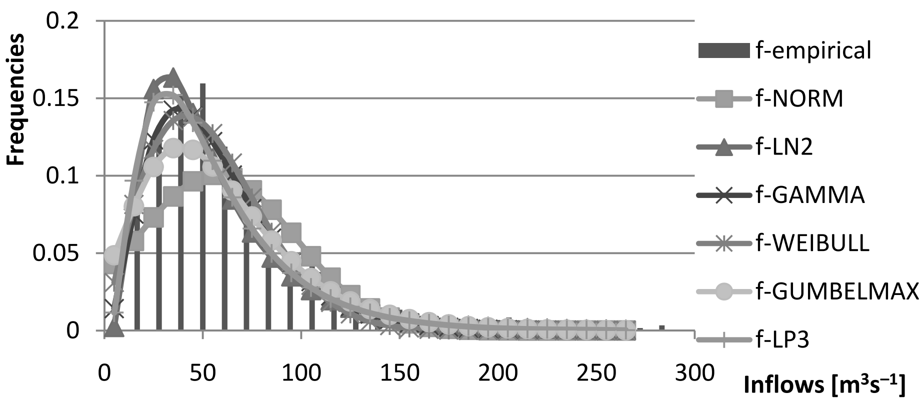

Figure 1. Fitting probability distributions to a set of observed values in January

Correlation matrices

According to eq. (12), but with time lag from 1 to *n*-1, the matrix (12 × 12) of the Pearson product-moment correlation coefficients of monthly mean inflows and the matrix (52 × 52) of the Pearson product-moment correlation coefficients of weekly mean inflows into the reservoir Peruca in the period 1994-2012 were calculated. The symmetric correlation matrix of monthly mean inflows is shown in tab. 2. Table 3 displays the average values of the Pearson's correlation coefficients of monthly mean inflows and of weekly mean inflows for time lags 1 to 4.

Table 2. Matrix of correlation coefficients of monthly mean inflows into Peruca reservoir in the period 1994-2012

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0.575	0.088	0.202	0.298	0.046	0.094	–0.173	0.223	0.000	0.396	0.005
2	0.575	1	0.487	0.286	0.377	0.311	0.515	0.266	0.172	–0.031	0.180	0.011
3	0.088	0.487	1	0.596	0.436	0.400	0.324	0.173	0.008	–0.334	–0.161	0.092
4	0.202	0.286	0.596	1	0.664	0.357	0.183	0.057	–0.036	–0.006	0.073	0.334
5	0.298	0.377	0.436	0.664	1	0.604	0.580	0.281	0.396	0.136	0.006	0.335
6	0.046	0.311	0.400	0.357	0.604	1	0.809	0.178	0.115	–0.192	–0.179	0.537
7	0.094	0.515	0.324	0.183	0.580	0.809	1	0.411	0.295	0.024	–0.248	0.274
8	–0.173	0.266	0.173	0.057	0.281	0.178	0.411	1	0.487	0.452	–0.001	–0.096
9	0.223	0.172	0.008	–0.036	0.396	0.115	0.295	0.487	1	0.721	0.073	–0.077
10	0.000	–0.031	–0.334	–0.006	0.136	–0.192	0.024	0.452	0.721	1	0.341	0.003
11	0.396	0.180	–0.161	0.073	0.006	–0.179	–0.248	–0.001	0.073	0.341	1	0.320
12	0.005	0.011	0.092	0.334	0.335	0.537	0.274	–0.096	–0.077	0.003	0.320	1

Considering the values of the Pearson correlation coefficients of monthly mean inflows, we calculated the partial correlation coefficients between successive months. A partial correlation coefficient is a measure of the linear dependence of a pair of random variables from a collection of random variables in the case where the influence of the remaining variables is eliminated. The average lag-1 Pearson correlation coefficient and the average lag-1 partial correlation coefficient of monthly mean inflows were almost equal.

Finally, the matrix (12 × 12) of the Pearson correlation coefficients of logarithms of the monthly mean inflows and the matrix (52 × 52) of the Pearson correlation coefficients of logarithms of the weekly mean inflows were calculated.

Table 3. Average values of Pearson's correlation coefficients of the mean inflows for different time lags

Lag	Average correlation coefficient between two months	Average correlation coefficient between two weeks
1	0.502	0.678
2	0.263	0.497
3	0.148	0.429
4	0.101	0.356

Table 4. Average values of Pearson's correlation coefficients of logarithms of the mean inflows for different time lags

Lag	Average correlation coefficient between two months	Average correlation coefficient between two weeks
1	0.505	0.745
2	0.252	0.575
3	0.166	0.495
4	0.121	0.406

The average values of these correlation coefficients for time lags 1 to 4 are shown in tab. 4. Stronger correlations between weekly inflows than between monthly inflows were expected.

According to the obtained results it can be concluded that the simple autoregressive model presented by eq. (9) is not suitable for modeling monthly and especially weekly inflow due to the significant asymmetry of the inflow distribution. Furthermore, the results show that for modeling the monthly inflows it is quite reasonable to neglect the correlations with time lag greater than one, *i. e.* the simple lag-1 autoregressive model given by eq. (17) is an appropriate dependent model for producing monthly inflows in our case. Due to the stronger correlations between the weekly inflows, the neglect of the relationships with time lag greater than one, sometimes are not acceptable. Using the partial correlation coefficients for consecutive weeks instead of the Pearson correlation coefficients may be a suitable solution.

Conclusions

For an optimal long-term/mid-term scheduling of the hydrothermal power system the study period is divided in discrete computational time intervals, usually one week or one month long. The selection of a time step, among other reasons, depends on data availability and characteristics of inflows. This article has clearly shown:

- The choice of the base time interval affects the required length of hydrological data observation in order to make use of historical data as a representation of the hydrological characteristics of the river basins. As the basic period is longer, the total required observation period is shorter.
- The choice of the basic computational time interval affects the asymmetry of a distribution. For the same set of observations inflows, the expected skewness coefficient of the distribution of weekly inflows is slightly higher than the expected skewness coefficient of the distribution of monthly inflows.
- The inflows are generally highly correlated for the shorter time intervals.
- The simple lag-1 autoregressive model for log-normally distributed inflows is an appropriate dependent model for producing monthly inflows in our case.
- The simple lag-1 autoregressive model for log-normally distributed weekly inflows is sometimes not acceptable. Using partial correlation coefficients for consecutive weeks instead of the Pearson correlation coefficients may be a suitable solution.

Nomenclature

cv – coefficient of variation [–]
cs – coefficient of skewness [–]
i – inflow [m^3s^{-1}]
r – correlation coefficient [–]
z – normal random variable with zero mean and unit variance [–]

Greek symbols

μ – mean [–]
σ – standard deviation [–]

Subscripts

m – month
j – year
cv – coefficient of variation

Acronyms

PDF – probability density function
LN2 – two-parameter lognormal distribution
LN3 – three-parameter lognormal distribution
LP3 – log-Pearson type 3 distribution

References

[1] Yu, Z., A New Long-Term Hydro Production Scheduling Method for Maximizing the Profit of Hydroelectric Systems, *IEEE Transactions on Power Systems, 13* (1998), 1, pp. 66-71

[2] Ikura, Y., Gross, G., Efficient Large-Scale Hydro System Scheduling with Forced Spill Conditions, *IEEE Transactions on Power Apparatus and Systems*, *103* (1984), 12, pp. 3502-3520
[3] Ngundam, J. M., *et al*., Optimal Scheduling of Large-Scale Hydrothermal Power Systems Using the Lagrangian Relaxation Technique, *Electrical Power and Energy Systems*, *22* (2000), 4, pp. 237-245
[4] Ferrero, R. W., *et al*., A Dynamic Programming Two-Stage Algorithm for Long-Term Hydrothermal Scheduling of Multireservoir Systems, *IEEE Transactions on Power Systems*, *3* (1998), 4, pp. 1534-1540
[5] Sutlovic, E., *et al*., A Method for the Long-Term Scheduling of Hydrothermal Power System with Multiple User Reservoirs, *Thermal Science*, *11* (2007), 3, pp. 75-83
[6] Siqueira, T. G., *et al*., Stochastic Dynamic Programming for Long-term Hydrothermal Scheduling Considering Different Streamflow Models, *Proceedings,* 9th International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, 2006
[7] Pereira, M., *et al*., Long-Term Hydro Scheduling Based on Stochastic Models, *Proceedings*, International Conference on Electrical Power Systems Operation and Management, Zurich, Switzerland, 1998
[8] Gjelsvik, A., *et al*., Long- and Medium-Term Operations Planning and Stochastic Modelling in Hydro-Dominated Power Systems Based on Stochastic Dual Dynamic Programming, in: *Handbook of Power Systems I* (Ed. P. Pardalos *et al*.), Energy Systems, Springer, 2010, pp. 33-55
[9] Loucks, D. P., *et al*., *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*, UNESCO, Paris, 2005
[10] Wurbs, R. A., *Modeling and Analysis of Reservoir System Operations*, Prentice Hall PTR, New York, USA, 1996
[11] Pavlić, I., *Statistical theory and application* (in Croatian), Technical book, Zagreb, Croatia, 1988
[12] ***, Engineering Statistics*, NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, 2/02/2013
[13] Yue, S., *et al*., A Review of Bivariate Gamma Distributions for Hydrological Application, *Journal of Hydrology, 246* (2001), 1, pp. 1-18
[14] Al-Fawzan, M. A., Methods for Estimating the Parameters of the Weibull Distribution, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia, 2000
[15] Yue, S., The Bivariate Lognormal Distribution to Model a Multivariate Flood Episode, *Hydrological Processes, 14* (2000), 14, pp. 2575-2588
[16] Thomopoulos, N. T., Johnson, A. C., Some Measures on the Standard Bivariate Lognormal Distribution, *Proceedings,* Decision Sciences Institute Annual Meeting, Boston, Mass., USA, 2004
[17] McMahon, T. A, Diaz Arenas, A., Methods of Computation of Low Streamflow – a Contribution to the International Hydrological Programme, UNESCO Report, 1982
[18] Žugaj, R., *Hydrology* (in Croatian), University of Zagreb, Zagreb, Croatia, 2000
[19] Bonacci, O., *Karst Hydrology with Special Reference to the Dinaric Karst*, Springer-Verlag, Berlin, Germany, 1987
[20] Wiche, J. G., Vecchia, V. A., Lake-Lavel Frequency Analysis for Devils Lake, N. D., USA, Geological Survey Water-Supply paper; 2469, US Department of the Interior, Washington D. C., USA, 1996
[21] Gupta, V. L., Fordham, J. W., Multisite Streamflow Simulation of Truckee River, Nevada, *Proceedings*, Symposium on statistical hydrology, Tuscon, Ariz., USA, 1971, pp. 26-46
[22] Wang, Y., Guo, S., Chen, H., Comparative Study of Monthly Inflow Prediction Methods for the Three Gorges Reservoir, *Stochastic Environmental Research and Risk Assessment, 28* (2014), 3, pp. 555-570
[23] Luna, I., *et al*., Fuzzy Inference Systems for Synthetic Monthly Inflow Time Series Generation, EUSFLAT-LFA 2011, *Proceedings*, The 7th Conference of the European Society for Fuzzy Logic and Technology, Aix-Les-Bains, France, 2011, pp. 1060-1065
[24] McMahon, T. A., *et al*., Global Streamflows – Part 1: Characteristics of Annual Streamflows, *Journal of Hydrology, 347* (2007), 3-4, pp. 243-259

Paper submitted: November 28, 2013
Paper revised: March 29, 2014
Paper accepted: May 5, 2014