Raport Badawczy

Research Report

RB/37/2013

Modeling electricity generation potential of a micro hydroelectric power plant

Z. Nahorski, W. Radziszewska,
M. Osuch

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2013

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

Zbigniew Nahorski[1], **Weronika Radziszewska**[1], **Marzena Osuch**[2]

[1] Systems Research Institute, Polish Academy of Sciences
[2] Institute of Geophysics, Polish Academy of Sciences

Modeling Electricity Generation Potential of a Micro Hydroelectric Power Plant

Warszawa 2013

Projekt badawczy własny Nr N N519 580238
Ministerstwa Nauki i Szkolnictwa Wyższego

Komputerowe zarządzanie energią w ośrodku badawczo-szkoleniowym z rozproszonymi źródłami energii i zmiennym zapotrzebowaniem energetycznym na eksperymenty badawcze

Kierownik projektu:

prof. dr hab. inż. Zbigniew Nahorski

Zadanie:

Opracowanie algorytmów dla agentów reprezentujących poszczególne urządzenia bądź instalacje i prognozujących możliwości generacji lub poboru energii i zgłaszania ich alokacji w systemie bilansowania na bieżąco

Modeling electricity generation potential of a micro hydroelectric power plant

Zbigniew Nahorski, Weronika Radziszewska, Marzena Osuch

Abstract The report focuses on elaborating and testing a method suitable for a synthesis of streamflows, which can be used for simulating large number of streamflows of stochastic nature to be used in statistical examinations of a computer simulation system. The stream inflows to a reservoir, which gathers water in producing electrical energy by a micro hydroelectric power plant. This renewable energy source is connected to an electrical microgrid located in few buildings being premises of a research and education center. The proposed methods of synthesis have been also adopted to generation of wind speed and insolation sequences, another media used in producing renewable energy. A matched block bootstrap method has been chosen to form synthetic sequences. Lall & Sharma method and its modification have been used. Both methods were found to produce synthetic sequences with satisfactory statistical properties as compared to real measured data gathered in periods of several to some dozen of years. Several modifications of the methods are proposed.

Keywords: simulation, renewable energy devices, synthesis of weather data, matched block bootstrap, streamflow, wind speed, irradiation

1 Introduction

Microl hydroelectric power plants are quite frequently used as electricity generation units in microgrids. They use a renewable energy source which can be well predicted

Zbigniew Nahorski and Weronika Radziszewska
Systems Research Institute, Polish Academy of Sciences
e-mail: Zbigniew.Nahorski@ibspan.waw.pl
e-mail: Weronika.Radziszewska@ibspan.waw.pl
Marzena Osuch
Institute of Geophysics, Polish Academy of Sciences
e-mail: marz@igf.edu.pl

in a horizon of several hours to few days, depending on localization of the plant and the volume of the reservoir gathering water from an inflowing stream. Quite often also some yearly probabilistic characteristics of the stremflows are known, which enables approximate planning of the energy production with even longer horizons. However, the energy produced depends not only on the amount of water flowing into the reservoir, which in some seasons may be low, but also on the amount of water gathered there. Thus, besides modeling of the inflow to the reservoir, a simple reservoir operation model is needed to predict possible energy that can be produced from the plant.

Synthetic generation of streamflow is a problem that has been discussed from early 1970s. But it is of a vivid interest in the recent literature due to its usefulness in planing, design, and management of water resources systems. Different methods have been proposed for this purpose. They can be classified as follows.

- **Stochastic hydrological models.** Statistical data to estimate distributions, as well as regression or classic ARMA type models are standard tools applied here. This direction of research was started by the report by Yevjevich [52] in 1972. See also the book by Kottegota [21]. A good example of these approaches is also presented in [22]
- **Nonlinear stochastic parameter models.** Artifitial neural networks have been used [30], as well as different nonlinear time series models, including the threshold autoregression ones, see [43] or [17].
- **Nonparametric models.** The most popular approach here has been application of the bootstrap methods. The main idea of these methods, which were originally invented by Efron [13], is to resample the historical records with replacements. This may be done either directly, shuffling the historical data, or through a model fitted to historical data, which then provides new samples of data, see e.g. [11, 13]. Random resampling in the hydrology has been considered in [28, 42]. The dependent data, more appropriate for the stream flow records, has been considered in [25, 26, 35, 39, 40, 41, 45]. One of main problems here is to recover the statistical characteristics of the series. This is connected with problems of choice of the block lengths and of the way of resampling. The resampling is then usually bounded to a corresponding nearest neighboring blocks [26], with an appropriate measure of neighboring. In particular, the matched-block bootstrap technique has been proposed for it [41]. The choice of the block length has been discussed in [39, 40, 45]. It has been selected from extensive simulations with different block length, being mainly a multiplicity of 1 month. The final choice was 3 month. It should be remembered, however, that this length may depend very much on a river considered, and for a small river, like the one described in this report, the best length may be much smaller.
- **Model based resampling.** The matched-block bootstrap technique is a very convenient and simple method. The resampling idea has been, however, also applied in connection with modeling. The idea of using it in connection with classical modeling techniques aroused in 1980s, [14, 15, 16], and consists in bootstrapping residuals of a model, istead of original data. In particular, it has been used for modeling streamflows in [39, 40]. The final synthetic series is constructed

by summing up corresponding values of the time series model and bootstrapped residuals.

- **Wavelet transform models.** Wavelet analysis is a popular method of modeling non-stationary and nonlinear time series. It has been therefore quite widely utilized in hydrological modeling [4, 5, 24, 38, 31, 46, 47, 49, 50]. The wavelet transformation is particularly useful in analysis of different variabilities appearing in the series [5, 46, 50].

In our simulation model of electricity production it is not crucial to have a high precision of streamflow generation. It is enough to keep the average yearly variability, and reflect a distribution of the streamflows in different years. In this report we use the matched block bootstrap approach, which is a simple and quick technique, giving quite good synthesised sequences coming from real data. Some modifications of a method proposed earlier are introduced in the report. The applied method of streamflow synthesis is presented in Sec. 2.

The inflowing water is stored in a small reservoir. Its model consists basically of a simple balance equation. Some problems may be encounted in modeling the height of the water level or the area of the water surface, needed for example in modeling evaporation. The model is described in Sec. 3.

Section 4 concludes. In the appendix some results of using matched-block bootstrap techniques for synthesiszing wind speed and insolation data are presented.

2 Streamflow synthesis

2.1 Real river historical streamflows

The analyzed flow data were measured at a water gauge at Wólka Mlądzka, on a small river called Świder. It is a right tributary of Vistula river. It has its sources in the Żelechów Heights near the town Stoczek Łukowski, at the altitude 178 meters above see level. Its mouth is located south of Warsaw between the towns Józefów and Otwock, at the altitude 85 meters above see level. The length of Świder is 89.1 km and its basin is 1161,5 km^2. The river is meandering, its valley is rather narrow and not deep. The bottom and banks are mostly sandy.

Świder is a typical lowland river with the maximal flows in springs, caused by melting the snow cover. Additional spates are observed in summers (June, July) due to tempest rains and also in late autumns due to frequent rainy days. The low flows are usual in late summers and early autumns. The streamflows have been measured once a day and the record include 18262 measurements from 50 consecutive years, starting from the beginning of November 1960 and ending at the end of October 2010. The flows are statistically characterized in Figures 1 and 2. Evidently, the flows are highly non-stationary and their distribution is skewed.

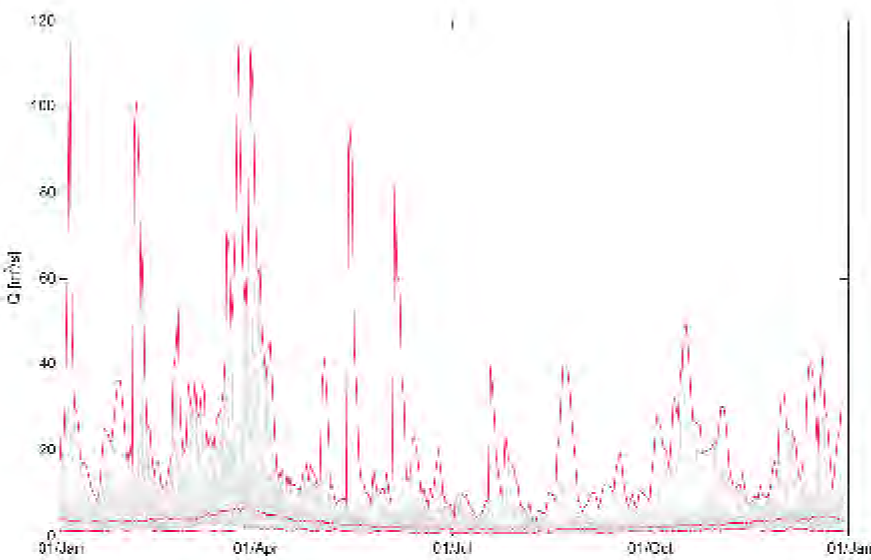

Fig. 1 Yearly minimal, median and maximal flows of Świder river from 1960/11/01 to 2010/10/31.

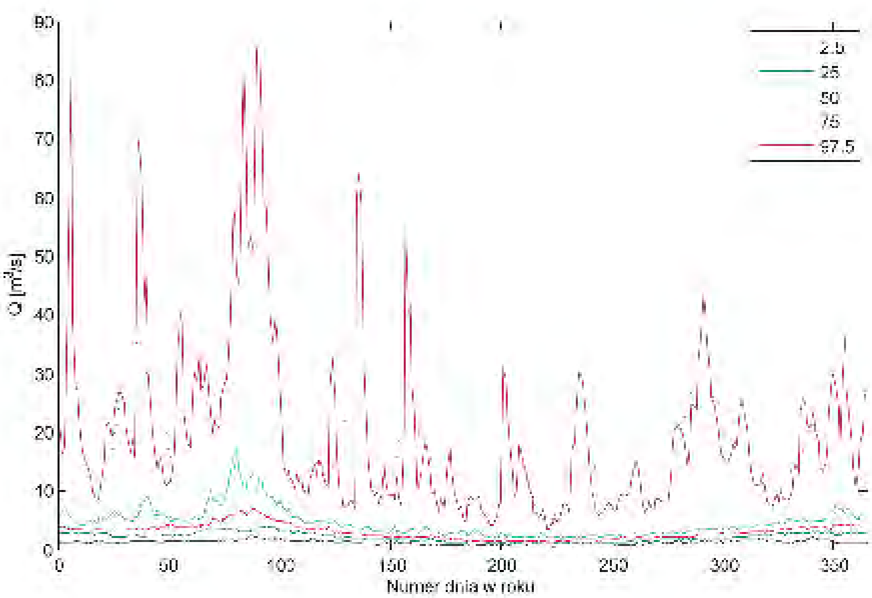

Fig. 2 Percentiles 2.5%, 25%, 50%. 75%, 97.5% of yearly smoothed flows of Świder river from 1960/11/01 to 2010/10/31.

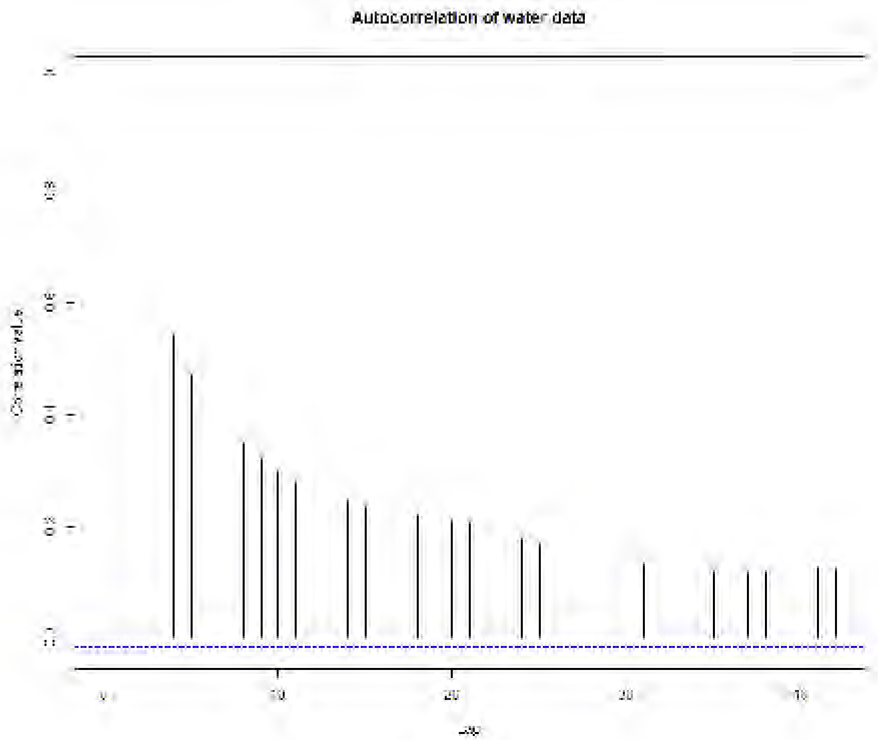

Fig. 3 Correlation of the water data.

2.2 Match block bootstrap resampling scheme

The method of generating flows described in this report is a modification of the matched block bootstrap method invented in [8] and applied for the streamflows in [41]. A general idea of the matched block bootstrap is as follows. The historical flowstream data are divided into non-overlapping subsequent time blocks, i.e. time intervals within a year. The intervals with the same dates from all historical years form sets, which are resampled consequently one by one to obtain a new series of flowstreams. This way each anew generated sequence of streamflows is different, and more exactly a probability of repetition of a sequence is very small (it depends on the lengths of intervals, number of the historical data years, and the way of resampling). Two problems are connected with this method:

- How to choose the blocks (intervals) length?
- How to resample the blocks (intervals) from the sets of the corresponding historical intervals?

Although choice of the block lengths should be prior to the second step, we start discussion with the resampling algorithm. The algorithm used in [41] is based on the k-NN bootstrap method proposed by Lall & Sharma [26]. In their method only k nearest neighbors within the set are only considered for the resampling. The near neighbors are those whose values of so-called 'feature vector' are close. In [41] only one feature was considered, that was the distance between the last and first measurements of flow in the concurrent blocks. The idea behind such choice of a feature is to keep the continuity of the flows between subsequent blocks in the generated sequence. Two strategies of resampling from the k nearest neighbors was considered, pure random, and inverse proportional to the ranks. In the latter case k nearest neighbors are ordered according to the distance from the considered block (in the sense of the distance between the last measurements in the blocks). Assume that the blocks are ordered as $j = 1, 2, \ldots, k$. Then the probability of choosing a j-th block is equal to

$$P_j = \frac{1/j}{\sum_{i=1}^{k} 1/i}$$

That is, the probability of the choice of a year is inversely proportional to the distance between blocks. This is called in the sequel 'the inversion method'.

A modification of the above algorithm has been introduced and examined in this report. In the modified algorithm, all years are considered in choosing the block instead of the k nearest neighbors. Moreover, the probability of choosing a j-th block is calculated as follows

$$P_j = 1 - \frac{D_j}{\sum_{i=1}^{N} D_i}$$

where D_j is the distance between blocks. To allow for choosing the next block from the presently considered year (continuation of the sequence) the next block is selected prior to considering the other years, with the probability equal to the correlation coefficient of the present and the next block in the considered year. In all cases the roulette wheel method is used for the selection with given probabilities.

Lall & Sharma [26] proposed using k equal to the square root of the number of years considered in the historical data. Srinivas & Srinivasan [41] suggested that k can be chosen using cross-validation techniques. No simple method has been given as to the choice of the block lengths. In [41] two values were considered, 1 month and 2 months. The choice has been based on extensive trials with both values and some values of k, using historical data of streamflows from few rivers. Several statistics of the historical streamflows has been compared with the corresponding statistics from the synthesized flows to check quality of the bootstrap method.

Figure 4 presents an example of a synthesized stremflow. Figures 5 - 7 show few results of testing the simulation method.

The above results indicate that the sequences produced by both methods quite well preserve most of statistics: mean value, median, standard deviation, and skewness. This is also visible when statistics for other weather phenomena, like wind speed or irradiation, are compared. To support this visual impression, some statistical tests have been used.

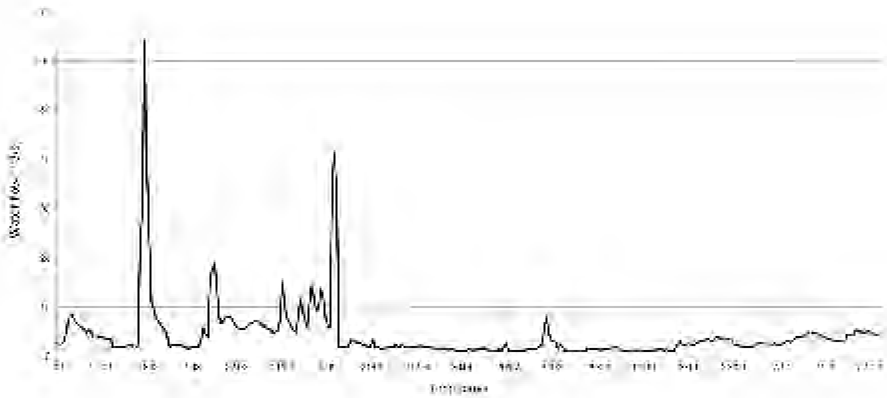

Fig. 4 An example of streamflow synthetized using match block bootstrap.

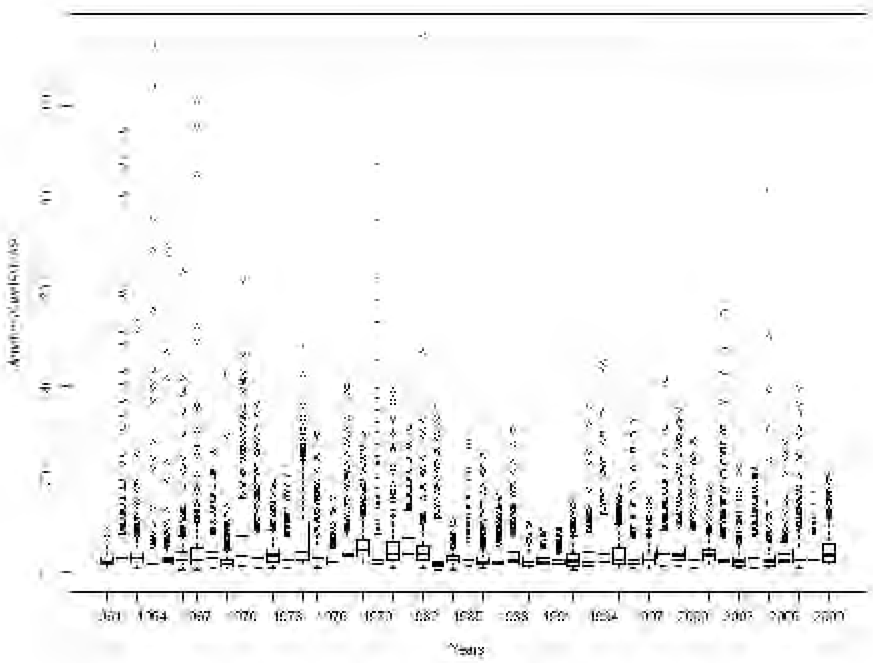

Fig. 5 Yearly box-and-whisker plots of streamflows synthetized using match block bootstrap.

Table 1 Comparison of descriptive statistics of real and generated flows

	Mean	Median	Autocorrelation	Skewness	Std. deviation
Inversion	4.39	2.85	0.88	7.58	6.23
Negation	4.30	2.72	0.87	7.38	5.76
Real measurements	4.25	2.84	0.91	7.28	5.56

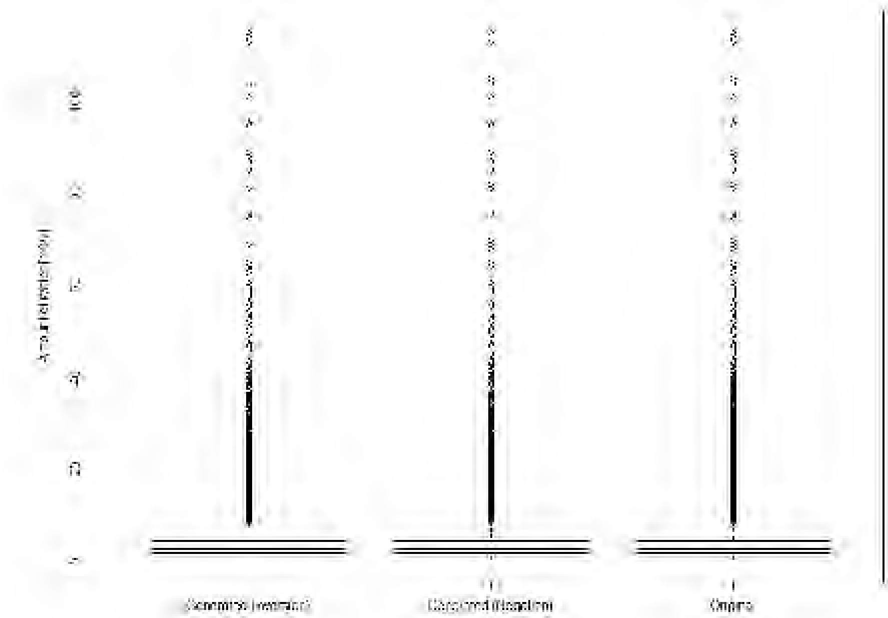

Fig. 6 Box-and-whisker plots of original and synthesized streamflows, for two match block bootstrap methods.

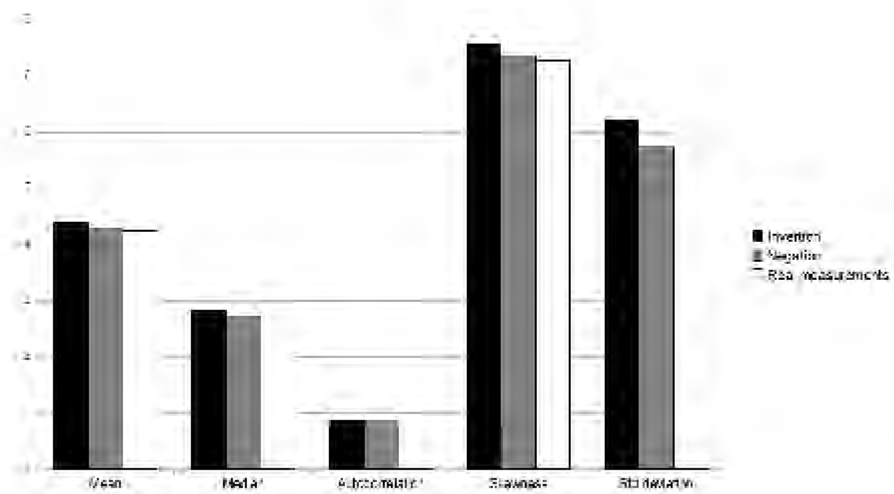

Fig. 7 Comparison of descriptive statistics of the original and synthetized streamflows.

First, tests for assessing the statistical significance of the difference between two sample means was performed. It is assumed that the underlying populations have normal distributions with the same variances. The null and alternative hypotheses are

$$H_0: \quad \mu_1 = \mu_2 \qquad H_1: \quad \mu_0 \neq \mu_1$$

where μ_1 and μ_2 are the means of the first and the second sequences. This hypothesis is evaluated with a two-tailed test.

The test statistic for independent samples of equal lengths is

$$t_i = \frac{\bar{x}_i - \bar{x}_0}{\sqrt{\frac{s_i^2}{n} + \frac{s_0^2}{N}}}$$

with $i = 1$ for the inversion method and $i = 2$ for the negation method. This meaning of i will be also used in the sequel. The above statistic has the Student's t distribution.

It is rather difficult to approve the independency assumption, due to seasonality of the data. Assuming dependence of the sequences, with other assumptions sustained, the matched-pairs test can be used. Let us consider the differences $d_{in} = x_{in} - x_{0n}$, where x_{0n} are the elements of the measurement sequence and x_{in} are elements of a synthesized sequence. The test statistics in this case is

$$t = \frac{\bar{d}_i}{\sigma_{d_i}}$$

where $\bar{d}_i$ is the mean value of the sequence $d_{in}, n = 1, \ldots, N$ and σ_{d_i} is its standard deviation. This is a one-tail test.

Results of both tests are presented in Table 2. Only in one case, for the inversion method and independence assumption, the zero hypothesis on equal mean values is rejected on the significance level 0.05. In all other cases the tests fail to reject H_0.

Table 2 The Student's t test statistics for comparison of means between the original and two synthesized sequences under assumption of independent or dependent samples.

	independent			dependent		
	t statistic	p-value	decision	t statistic	p-value	decision
Inversion	-2.28	0.023	nonequal	0.018	>0,99	equal
Negation	-0.63	0,521	equal	0.005	>0.99	equal

If the normality distribution can not be assumed, Mann-Whitney U test (or Wilcoxon-Mann-Whitney test) for independent samples is evaluated, while for the dependent samples the Wilcoxon matched-pairs signed-ranks test can be applied. Both tests use medians. In the former equality of medians (H_0) against inequality (H_1) is tested. In the latter it is tested if the median of the differences is zero (H_0) versus the alterrnative hypothesis that it is different of zero (H_1). Both are two-tailed tests. Table 3 presents the test results. In this case H_0 is rejected for the negation method in the Mann-Whitney test, while all other cases fail to reject it.

Table 3 The Mann-Whitney U test and Wilcoxon matched-pairs signed-ranks test statistics for comparison of medians for the original and two synthesized streamflow sequences.

	Mann-Whitney			Wilcoxon		
	statistic	*p*-value	decision	statistic	*p*-value	decision
Inversion	158366885	0.11	equal	78341228	0.53	equal
Negation	162134431	0.024	nonequal	76250087	0.79	equal

It may be concluded that the tests are in favour of equality of the examined statisctics for real and synthesized sequences. It is often so for the Wilcoxon test for dependent samples, that can be supposed true from the way of the generation of the sequences. In some cases where tests for independent samples are used, this hypothesis is rejected. But the reason may be most posssibly wrong assumption on independence of the real and synthesized sequences.

In the last test variances of two samples are compared. The hypotheses are:

$$H_0 : \sigma_0^2 = \sigma_i^2 \qquad H_1 : \sigma_0^2 \neq \sigma_i^2$$

The following statistic is used

$$F = \frac{s_i^2}{s_0^2}$$

It is assumed that the populations have normal distributions and the sequences are independent. Then the above statistic has the F Snedecor distribution.

The results of testing are depicted in Table 4. In both cases the zero hypothesis on equality of variances is rejected. But the assumptions of this test can hardly be accepted.

Table 4 The F test statistics for comparison of variances between the original and two synthesized sequences.

	F statistic	*p*-value	decision
Inversion	0.8004	$< 2.2\text{e-}16$	nonequal
Negation	0.9358	6.762e-06	nonequal

In the sequel, extensions of the method are discussed. The modifications are expected to improve the autocorrelation properties of the generated sequences. The main goal of the research described in this report was to elaborate a method of synthesis of streamflows with statistical properties similar to that obtained from the historical records, and useful for simulation of the electric energy produced by a hydraulic plant. As the sequences obtained by using the above described methods have been found to fit well to this purpose, the extensions proposed below have not been tested, and only recorded for further examination in the future works.

2.3 Two-element feature vector

In their method, Lall & Sharma [26] designed a feature vector, which may consists of more than one element. Following their idea, two features are considered here, called v_1 and v_2. The first feature is, as earlier, the absolute value of the difference between the last measurement of the previous block and the first measurement of the consecutive block. The second feature is the absolute value of the difference between the slopes of the end of the previous block and the beginning of the consecutive block. Then, the distance between the blocks is defined as follows

$$r_j = \sqrt{w_1 v_{1j}^2 + w_2 v_{2j}^2}$$

where j is the number of the consecutive block and w_1, w_2 are weights. Then, the rest of the procedure is as before.

The problem in the above procedure is in finding the slopes of the curves, known only by noisy measurements in the discrete times. The method used follows the Savitzky-Golay idea [33]. It is proposed to fit a linear function

$$P(t) = b_0 + b_1 t \tag{1}$$

to the $2L+1$ last or first measurements of the consecutive blocks, respectively. To simplify the notation, let us call the consecutive extreme measurements in the blocks $u_{-L}, u_{-L+1}, \dots, u_0, \dots, u_{L-1}, u_L$. It is also assumed that the measurements are made with a constant period of time Δ.

The parameters b_0 and b_1 are calculated to minimize the sum of squares

$$S = \sum_{l=-L}^{L} (u_l - b_0 - b_1 \Delta l)^2$$

Differentiating S with respect to b_0 and b_1, and equaling them to 0, we get a set of two linear equations to be solved for b_0 and b_1

$$b_0 \sum_{l=-L}^{L} \Delta^j l^j + b_1 \sum_{l=-L}^{L} \Delta^{j+1} l^{j+1} = \sum_{l=-L}^{L} \Delta^j l^j u_l \qquad j = 0,1$$

It is easy to notice that the sums for odd powers of the summands on the left hand side are equal to zero. Therefore, it is easy to get the following expressions for the parameters

$$b_0 = \frac{1}{2L+1} \sum_{l=-L}^{L} u_l$$

$$b_1 = \frac{\Delta \sum_{l=-L}^{L} l u_l}{\Delta^2 \sum_{l=-L}^{L} l^2} = \frac{3}{\Delta L(L+1)(2L+1)} \sum_{l=-L}^{L} l u_l$$

We want to have an estimate of the derivative at the points ΔL and $-\Delta L$, respectively for the previous and the next blocks. As the function (1) is linear, its slope is constant and equal to

$$\frac{dP(t)}{dt} = b_1$$

We calculate the slope only for simplest cases of $L = 1$ or $L = 2$, that is using extreme 3 or 5 points of the block sequence. It provides the following expressions

$$\frac{dP(t)}{dt} = \frac{1}{2\Delta}(-u_{-1} + u_1) \qquad \text{for} \qquad L = 1$$

$$\frac{dP(t)}{dt} = \frac{1}{10\Delta}(-2u_{-2} - u_{-1} + u_1 + 2u_2) \qquad \text{for} \qquad L = 2$$

As mentioned earlier, the indices $-2, -1, 1, 2$ correspond to the consecutive extreme measurements in adjacent blocks.

Higher order polynomials can by used as well. Although the derivations are more complicated then, the final expression is also in a form of the weighted sum of consecutive measurements. However, the solution may be then more sensitive to errors and perhaps not suitable for very volatile sequences.

2.4 Model based resampling

Srinivas & Srinivasan [39, 40] developed a model based resampling method, which was an adaptation of the idea presented in [11], where it has been called the post-blackening approach. Two time series models have been considered, a classical ARMA model [39] and periodic ARMA (PARMA) model [40]. To estimate the model parameters Box-Jenkins [6] procedure is applied.

The main idea of applying the model is to get a sequence of independent residuals, which can be then simply bootstrapped to form new sequences. Srinivas & Srinivasan [40] proposed to use a simple autoregressive model of the first order AR(1) in order to only prewhiten the original sequence, and then using a match block bootstrap resampling for residuals, which form a much weaker dependent sequence.

So, the algorithm is as follows.

1. Standardize the elements of the original sequence y_n to form a new sequence

$$x_n = \frac{y_n - \bar{y}}{s_n}$$

   where $\bar{y}$ is the mean and s_n is a 'local' standard deviation of the sequence y_n.
2. Fit an AR(1) model

$$x_n = ax_{n-1}$$

to the standardized data. The least-squares estimate of the parameter a is given by the expression

$$\hat{a} = \frac{\sum_{n=2}^{N} x_n \sum_{n=2}^{N} x_{n-1}}{(\sum_{n=2}^{N} x_{n-1})^2}$$

3. Pre-whiten the standardized data x_n by calculating the residuals

$$\varepsilon_n = x_n - \hat{a}x_{n-1}$$

4. Bootstrap the obtained sequence of residuals $\{\varepsilon_{n-1}\}_{n=1}^{N}$ using a matched block bootstrap resampling scheme. The time dependence of the prewhitened sequence is much weaker, so it should be easier to obtain good statistical properties of the generated sequences.
5. Post-blacken a bootstrapped series by inverting the earlier pre-whitening calculations, i.e. use the following expression

$$z_n = \hat{a}z_{n-1} + \varepsilon_n$$

with $z_0 = 0$.
6. Inverse standardization, i.e. calculate

$$\tilde{y}_n = s_n z_n + \bar{y}$$

to get a new synthesized sequence of stremflows.

It is worth trying to differentiate the sequence before standardization, to get even a weaker dependent sequence at the start of the post-blackening algorithm. It is very likely that the mean value of the differentiated series will be close to zero, and then the standardization and inverse standardization steps may be abandoned. In any case, the final sequence of the modified procedure is obtained by summing the consecutive elements of $\tilde{y}_n$.

3 Model of a reservoir

In our model a stream inflows to a small reservoir, from which water is fed to a turbine of a micro hydroelectric power plant of run-of-the-river type. Operation of reservoir, together with iy modeling, has a long history in the water management literature, see e.g. [51], but it is mostly connected with large reservoirs and optimization of its water usage. To simplify calculations in our simulation, the reservoir is modeled as the inverse of a frustrum (parallel cutting) of a wedge (a shape of a baking pan) sketched in Figure 8, where we assume that $a > b$. The volume of this polyhedron is given as follows [7]

$$V_{\max} = \frac{h}{6}[ab + a_1 b_1 + (a + a_1)(b + b_1)]$$

In futher derivations it will be useful to specify a_1 and b_1 by angles of the river bank slope, which is denoted by α, for every riverside. From simple triginometric conditions in the triangle presented in Figure 9 we have

$$\tan\alpha = \frac{H}{b/2} \qquad \tan\alpha = \frac{H-h}{b_1/2}$$

Eliminating H we get

$$b_1 = b - \frac{2h}{\tan\alpha}$$

and similarly

$$a_1 = a - \frac{2h}{\tan\alpha}$$

Using above, the volume can be expressed as

$$V_{\max} = h\big[ab - \frac{h}{\tan\alpha}(a+b) + \frac{4h^2}{3\tan^2\alpha}\big] \tag{2}$$

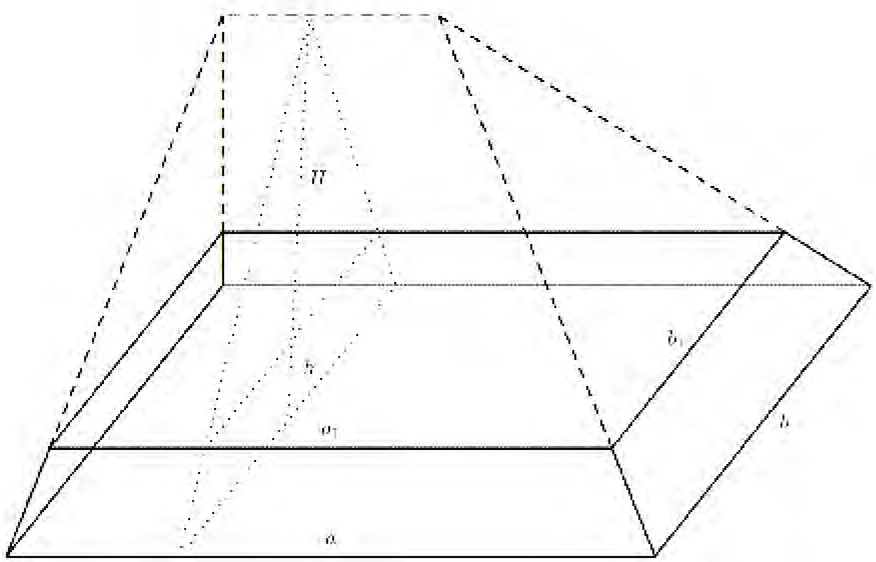

Fig. 8 A sketch of a wedge being a model of a reservoir, after turning it upside down.

The volume of water in the reservoir is changing due to the inflow from the stream, inF_t, the outflow to the power station or overfilling of the reservoir, $outF_t$, evaporation from the reservoir surface, E_t, as well as groundwater inflow and outflow, precipitation on the reservoir surface, and surface runoff from a watershed into a lake. It is actually impossible to take into account all these reservoir water budget components in our modeling due to lack of appropriate data. These data are neither

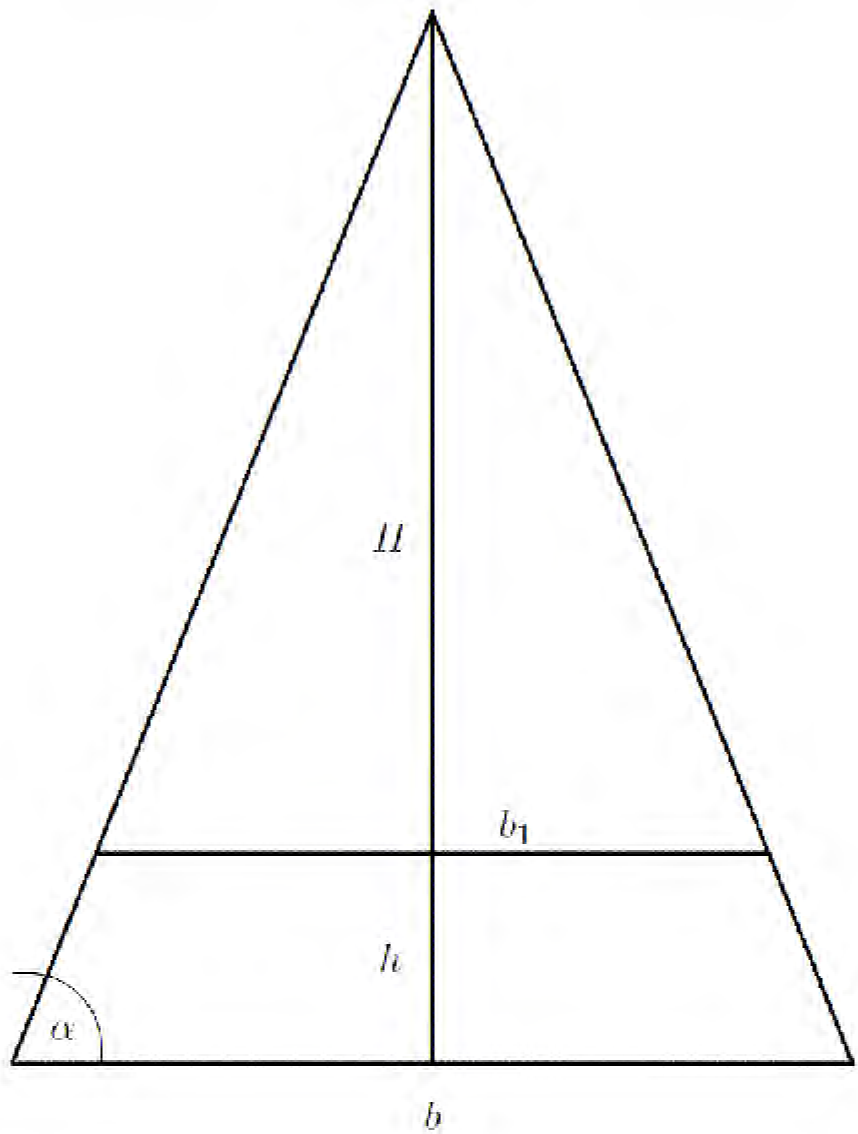

Fig. 9 The triangle from Figure 8.

available in traditional modeling of reservoir water level, which makes this task to be quite a big problem, requiring application of sophisticated techniques, see [44]. So, we only consider three first components. Let us denote the net inflow of the water to the reservoir during a considered period by

$$\Delta F_t = inF_t - outF_t - E_t$$

This value may be positive, negative or equal to zero. The net inflow causes the change of the water lever, denoted by Δp_t. The new values of the open water surface dimensions in the next period are then equal to

$$a_{t+1} = a_t - \frac{2\Delta p_t}{\tan\alpha} \qquad b_{t+1} = b_t - \frac{2\Delta p_t}{\tan\alpha}$$

and the corresponding change of the water level in the reservoir Δp_t can be found from the equation (2) to be as follows

$$\Delta F_t = \Delta p_t \big[a_t b_t - \frac{\Delta p_t}{\tan\alpha}(a_t + b_t) + \frac{4\Delta p_t^2}{3\tan^2\alpha}\big]$$

Let us denote $\frac{\Delta p_t}{\tan\alpha} = x$ and $\frac{\Delta F_t}{\tan\alpha} = Y_t$. Then we get the following expression

$$Y_t = \frac{4}{3}x^3 - (a_t + b_t)x^2 + a_t b_t x \tag{3}$$

This is a dependence of Y_t on x. However, we actually need the reverse dependence, showing how the level changes due to an inflow to the reservoir. This dependence can be found by solving above third order linear equation in x. Although theoretically possible, solution of this equation is too complicated for practically useful analysis. So we use another approach.

Let us consider the function

$$f(x) = \frac{4}{3}x^3 - (a_t + b_t)x^2 + a_t b_t x - Y \tag{4}$$

Interesting solutions (real ones - there must be at least one real solution) of the equation (3) are the zeros of the above function. The derivative of this function is

$$f'(x) = 4x - 2(a_t + b_t)x + a_t b_t$$

This derivative has two zeros

$$x_1 = \frac{b_t}{2} \qquad x_2 = \frac{a_t}{2}$$

It is easily found that $f(b_t/2) - f(a_t/2) = (a_t - b_t)^3/12 > 0$. That is $f(a_t/2) < f(b_t/2)$. As we have

$$f(\frac{a_t}{2}) = \frac{a_t^2}{4}\big(b_t - \frac{a_t}{3}\big) - Y_t$$

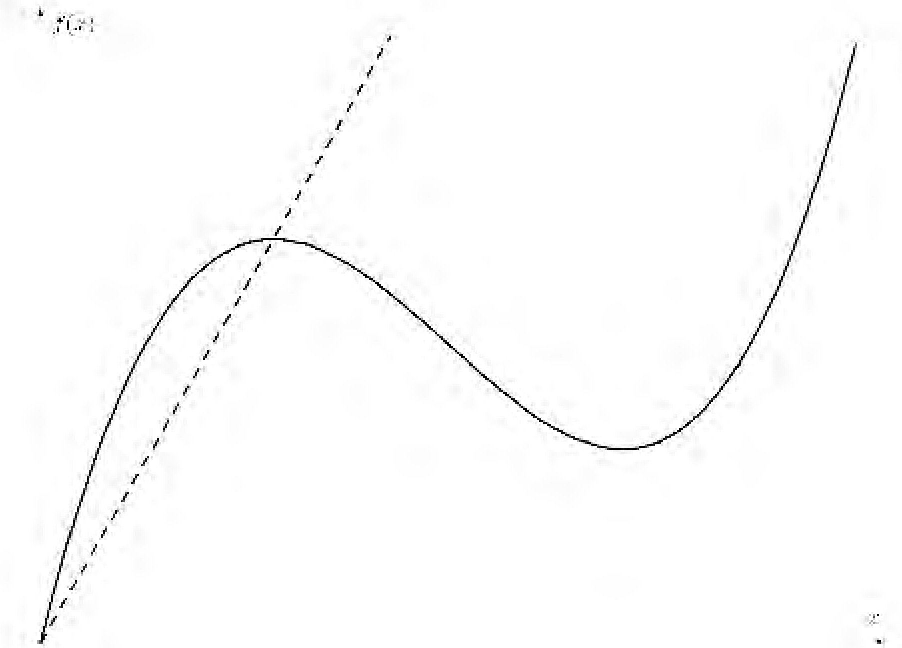

Fig. 10 A cubic function (3) with explanation of an upper bound calculation.

then this function has only one zero when

$$\frac{a_t^2}{4}\left(b_t - \frac{a}{3}\right) > Y_t$$

that is for sufficiently high values of the surface area dimension. We assume in the sequel that the above condition is true. A function $f(x)$ is sketched in Figure 10. The zero x_0 of this functions can be bound by the condition $0 < x_0 < b_t/2$.

The upper bound can be easily sharpened. Let us draw a line by the points $(0, f(0))$ and $(b_t/2, f(b_t/2))$, that is $(0, -Y_t)$ and $(b_t/2, b_t^2(3a_t - b_t)/12 - Y_t$. The equation of this line is $y = \frac{1}{6}b_t(3a_t - b_t)x - Y_t$ and therefore it cuts the x axis in the point

$$x_{01} = \frac{6Y_t}{b_t(3a_t - b_t)}$$

which is a sharper upper bound. This bound is used to assess the time period in modeling. Assuming $a_t = 100$m and $b_t = 40$m the denominator is equal to 22400 m^2. The streamflow of the river can reach the value of almost 100 m^3/s. This gives the value $p_t \approx 27$ mm/s or 160 cm per minute. This means that a reasonable period for modeling changes of the water level in the reservoir would be only dozens of seconds. On the other hand, for the small streamflow of about 2 m^3/s this value is only 32 mm per minute, so the period can be equal to a dozen of minutes. Taking

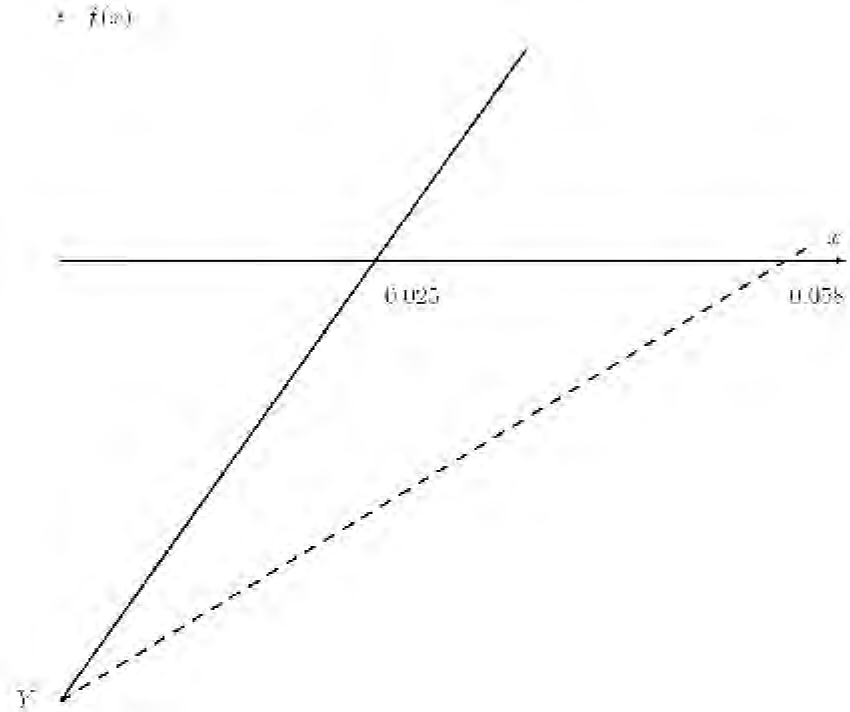

Fig. 11 An initial part of the the Figure 10 showing accuracy of an upper bound.

into account that the bound is around twice greater than the zero of the function (3), see Figure 11, the above estimates are perhaps too conservative and can allowed to be greater.

Evaporation from the reservoir surface is worth to examine as a component of the water budget in summers. Study of the evaporation has been started by Dalton [10], who assumed that it is proportional to the difference of the saturation vapour pressure and the vapour pressure in the air. The proprtionality coefficient has been assumed in further studies to be dependend on the wind speed, temperatures, humidity, etc., see e.g. [32, 37] for discussions of different formulas. These formulas are often related to specific conditions or require measurements of different conditions. In spite of the large documentation, their accuracy may be not sufficiently good, particularly when they are applied in other sites then those, for which their parameters were fitted. On the other hand, pan evaporation is often measured, and this result can be used for estimation of evaporation from the reservoirs after some modifications. Here, we form a simple model inspired by data for Minnesota presented in [9]. Minnesota State lies on similar lattitude as Poland, slightly more shifted to south. Average daily evaporation in July, which is the month of highest evaporation, has been estimated there to be in the range 3.5 - 6 mm/day. The lowest evaporation is in

December and amounts to about 0.3 - 0.5 mm/day. The shape of average monthly evaporations resembles a sinusoid. This inspired us to propose a simple model for daily evaporation in the dth day of the year as

$$e_d = 5\big(\sin(\frac{2\pi}{365}d+1\big)+0.5 \qquad \text{[mm/day]}$$

It assumes the highest evaporation about 5 mm/day in the days 182 and 183 (1st and 2nd July), and the lowest equal to 0.5 mm/day in the day 365 (31st December). For the leap years the number 366 instead 365 may be used.

However, even peak evaporation rates are only of the order of few milimiters per day [9, 32]. This may be meaningful when the net inflow (difference of inflow and outflow) is close to zero. But it is not much when it is compared to change of the water level due to inflow from the stream. So, it seems that evaporation can be ignored in our modeling.

Finally, change of the water volume in the reservoir is given by

$$V_{t+1} = V_t + \Delta F_t$$

starting with an assumed initial value V_0. Additionally, if $V_{t+1} > V_{\max}$, then in the next step the maximum reservoir volume should be taken $V_{t+1} = V_{\max}$. A minimal level of water is required for the turbines to work. If it is denoted by $h_{\min}$, than the minimal volume is the difference of the maximal volume and the volume corresponding to the change $h - h_{\min}$

$$V_{\min} = h\big[ab - \frac{h}{\tan\alpha}(a+b) + \frac{4h^2}{3\tan^2\alpha}\big] -$$

$$-(h-h_{\min})\big[ab - \frac{h-h_{\min}}{\tan\alpha}(a+b) + \frac{4(h-h_{\min})^2}{3\tan^2\alpha}\big]$$

that is

$$V_{\min} = h\big[ab - \frac{2h-h_{\min}}{\tan\alpha}(a+b) + \frac{4(3h^2-3hh_{\min}+h_{\min}^2)}{3\tan^2\alpha}\big] \tag{5}$$

This model requires only inspection of the minimal lever. With the values of $a =$ 100m and $b = 40$m assumed earlier 1 m^3/s of the net inflow causes 16 mm change of the water level per minute, that is almost 25 cm per quater of hour. As inspection of the minimal level is important for small inflows, it seems that a quater to half an hour may be a reasonable time period in modeling.

4 Conclusions

Block matched bootstrap techniques proved to be a powerful tool for synthesizing large amount of data with statistically similar properties to the original measurements. These techniques do not require learning of a sophisticated theory, nor implementing of complicated algorithms. The tests presented in this report confirmed good statistical properties of the synthesized series of streamflows, and to some extent also of wind speed and insolations. The properties of obtained sequences did not differ substantially for both applied methods: Lall & Sharma (inversion) and a novel one (negation), proposed in this report as a modification of the former one.

Synthesized series are often produced to test statistical properties of simulated objects, either existing ones or designed, to get some better knowledge of their possible behavior. This was also a main goal of the sequences synthesized using the methods elaborated in this report. They are to be used as inputs to renewable electric energy sources: a micro hydroelectric power plant, a wind power generator, and photovoltaic panels. The electric power produced by them supplies a microgrid, which is simulated in a computer as a model of a microgrid for a designed research and education center, with renewable energy units.

Few modifications and extensions of the methods have been proposed in the report. They may be the subjects of further research in order to improve quality of the methods. However, the sequences obtained up to now are good enough for testing of microgrid behavior. That is why the elaboration of the methods have been tentatively stopped at this point.

Appendix: Results for wind speed and insolation

Wind speed

Wind speed synthesis has not been as popular subject of research as in the sreamflow case, as no extensive simulations similar to those required for testing water systems models were needed. In these applications where the wind speed is needed, most often historical data have been used. Much more attention has been put to wind speed prediction and different methods for this purpose have been developed, starting from general weather forecasting methods [27], with possibly downscaling to a small space grid using different models, like regression, artificial neural network or support vector machines [18], through different time series forecasting methods [1, 3, 53], possibly using on-line measurements in nearby weather stations, up to different wavelet and fuzzy approaches [53]. In many of them diurnal wind speed is only considered.

In this report the matched block bootstrap methods has been used, applied earlier for the flowstream syntheses. The measurements used for the modeling were gathered in weather station located at 52°10‘53“N 20°52‘13“E in Reguły near Warsaw[1]. The wind speed has been measured every 10 minutes by consecutive 10 years, from 2003 to 2012. The total number of data contains 526032 points. The block length of 5 hours (30 points) has been fixed, following examination of few ad hoc chosen

[1] The data on the wind speed and insolation measurements were provided by LAB-EL Elektronika Laboratoryjna s.j, Reguły, Poland.

values. The autocorrelation diagram is presented in Fig. 12. Figure 13 shows an example of a synthetic wind speed sequence, while Figures 14 - 15 present comparison of some statistics for the original and synthesized data. The statistics for synthesized series are similar to the statistics for the original ones.

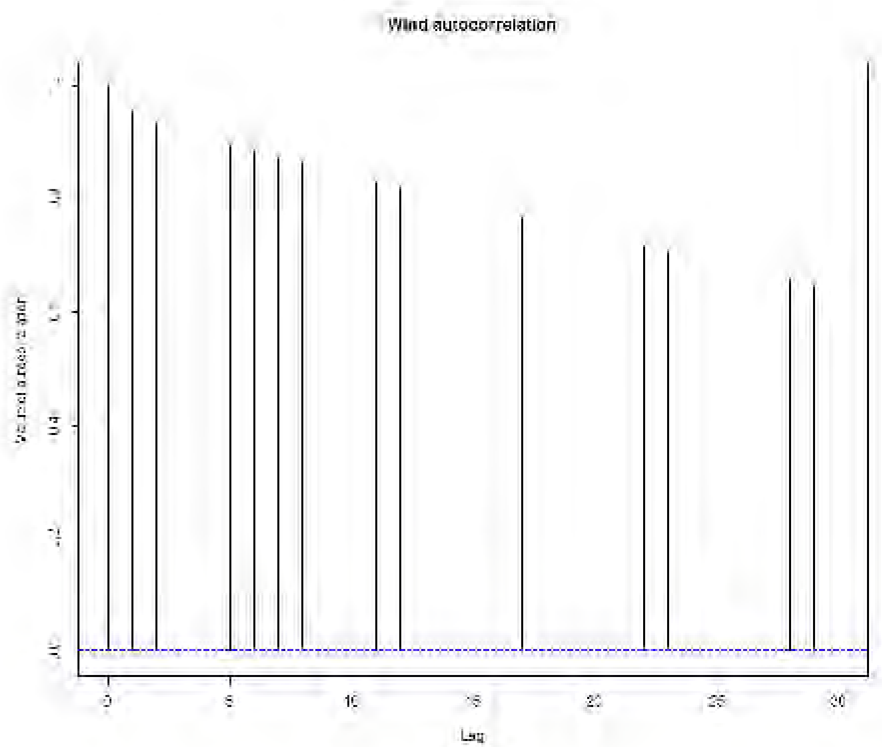

Fig. 12 Diagram of autocorrelations calculated for lags from 0 to 30, which shows correlation between measured data and 5 hour shift.

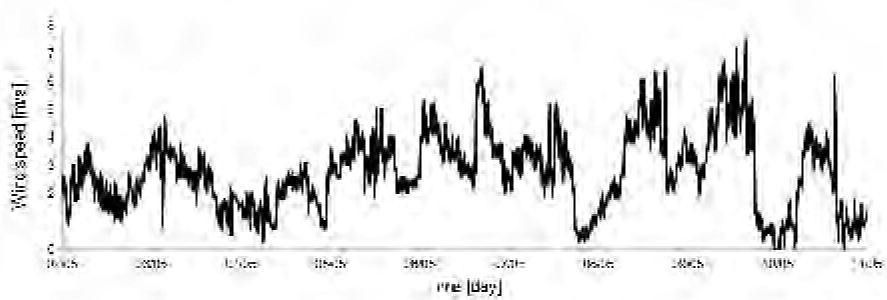

Fig. 13 An example of wind speed synthesized using match block bootstrap.

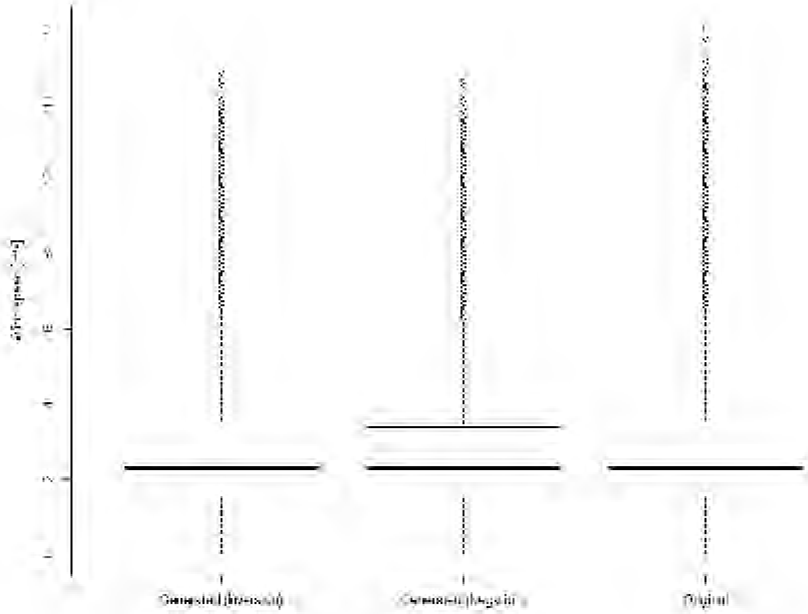

Fig. 14 Box-and-whisker plots of original and synthesized wind speed, for two match block bootstrap methods.

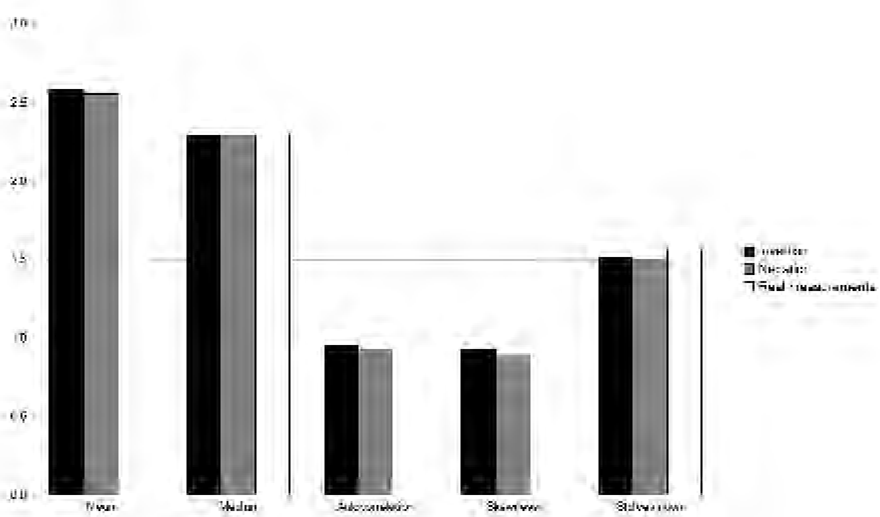

Fig. 15 Comparison of descriptive statistics of the original and synthetized wind speed.

Finally, Figure 16 depicts fractions of occurrence of wind speeds in the original and synthesized series. It is clear from it that the fitness proportionate selection with negation method of determining the probabilities of choosing next block is a bit more randomized and the most extremely values appear less frequent.

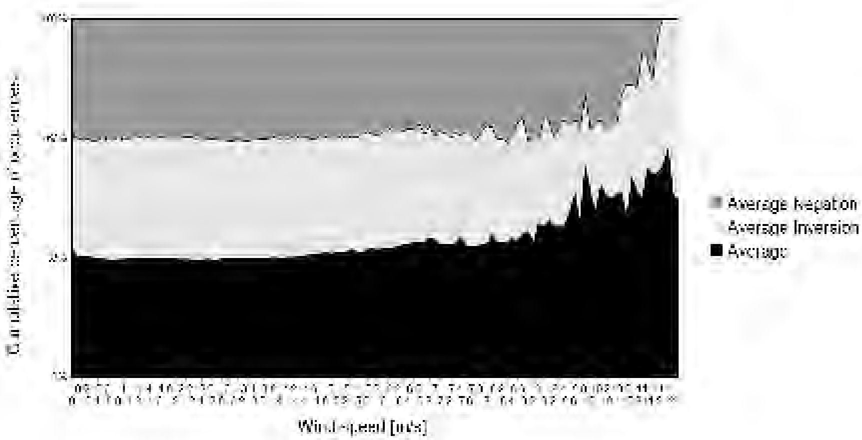

Fig. 16 Fractions of occurrences of wind speeds in two generated and real measurement sequences.

Table 5 Comparison of descriptive statistics of real and generated wind speed

	Mean	Median	Autocorrelation	Skewness	Std. deviation
Inversion	2.59	2.30	0.95	0.94	1.52
Negation	2.56	2.30	0.93	0.90	1.51
Real measurements	2.60	2.30	0.96	0.98	1.57

As in the case of streamflow data, also here some tests have been performed to compare distributions of the original and synthesized series. The results are depicted in Tables 6 and 7. Table 6 contains results of the t-test, under assumption of normality of the data, for independent and dependent sequences. Table 7 shows results for the Mann-Whitney and Wilcoxon tests, respectively for the independent and dependent data, without normality assumption. All these results suggest acceptation of the alternative hypothesis of different distributions of the real and synthesized sequences. This conclusion may a little surprising in light of comparison results of Table 5. Possible reasons of this descrepancy may be relatively short blocks and lack of regularity in wind speed values accross the years (no clear seasonality). This question needs, however, more thorough examination.

Table 6 The Student's *t* test statistics for comparison of means between the original and two synthesized sequences under assumption of independent or dependent samples.

	independent			dependent		
	t statistic	*p*-value	decision	*t* statistic	*p*-value	decision
Inversion	16.51	<0.01	nonequal	16.51	< 2.2 10^{1}6	nonequal
Negation	13.26	<0,01	nonequal	0.69	0.489	equal

Table 7 The Mann-Whitney U test and Wilcoxon matched-pairs signed-ranks test statistics for comparison of medians for the original and two synthesized wind speed sequences.

	Mann-Whitney		Wilcoxon			
	statistic	*p*-value	decision	statistic	*p*-value	decision
Inversion	139866932286	< 2.2e-16	nonequal	54979777950	< 2.2e-16	nonequal
Negation	139319528634	1.812e-14	nonequal	54553238682	< 2.2e-16	nonequal

Insolation

To simulate insolation usually astronomical data and physical principles have been considered, like position of the sun in the sky, the resulting flux of the sunshine coming to the earth surface, etc., see [29]. Up to the knowledge of the present authors, no influence of the clouding on the running energy production has been considered. An assessement of insolation in Poland in different months has been presented in [23]. In this report, matched block bootstrap methods have been used to synthesize the insolation sequences, together with clouding effect.

The insolation data have been gathered in the same weather station as those for wind speed. It contains measurements in every 10 minutes from 2004 to 2012. They form a sequence of 473472 values.

Choice of the block length was self-imposed. One-day-long blocks have been considered, which means that they cover the length of 24 hours from midnight to midnight. They include intervals of the zero values connected with the nights. The length of nonzero values in the days along the year varies: the longest during the break of the spring and summer, and the shortest at the break of autumn and winter.

Because the adjoining points of all blocks are always the same, equal to zero insolation, another measure of neighborhood of consecutive blocks have been used. It was the correlation coefficient of the daily insolation sequences.

Figure 17 present an example of synthetic daily insolation sequence. Figures 18 - 20 show comparison of statistics, similar to those presented earlier. Figure 21 shows the statistics of the original and generated data but with zero values excluded. As before, quite good similarity of statistics for the synthetic and real data is observed, including the autocorrelation coefficients. This suggests satisfactory generation of synthetic series by both matched block bootstrap methods.

It is obvious that no assumptions on stationarity or normality for these data can be supposed, which questions any use of standard tests, as performed in for the earlier data. It seems that other approach is needed to solve this problem.

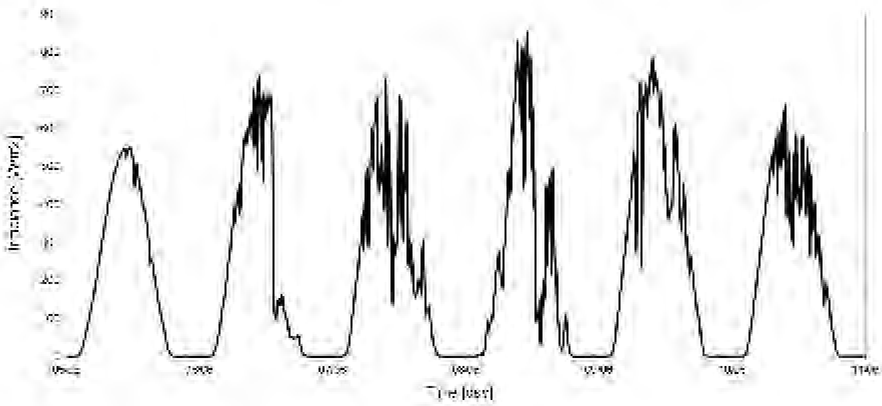

Fig. 17 An example of irradiance synthesized using match block bootstrap.

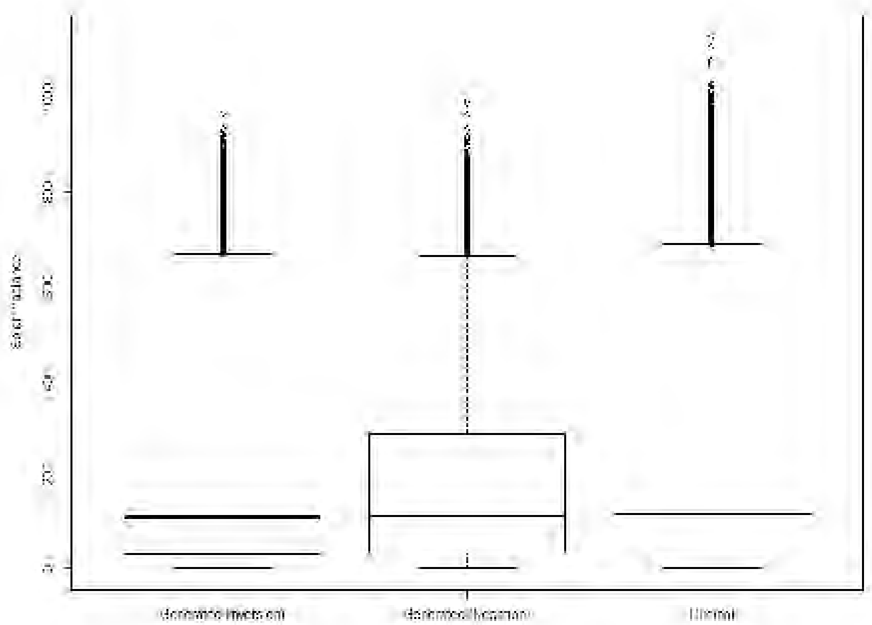

Fig. 18 Box-and-whisker plots of original and synthesized irradiance, for two match block bootstrap methods.

Table 8 Comparison of decriptive statistics of real and generated irradiance, with excluded 0 values.

	Mean	Median	Autocorrelation	Skewness	Std. deviation
Inversion	189.2	118.7	0.958	1.03	190.0
Negation	190.1	120.2	0.956	1.05	190.8
Real measurements	192.0	119.1	0.954	1.09	195.0

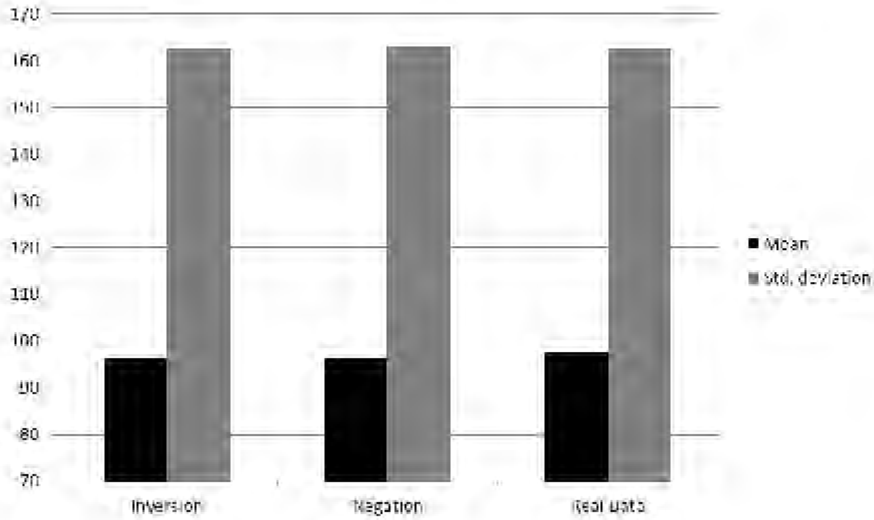

Fig. 19 Comparison of means and standard deviations of the original and synthetized irradiance.

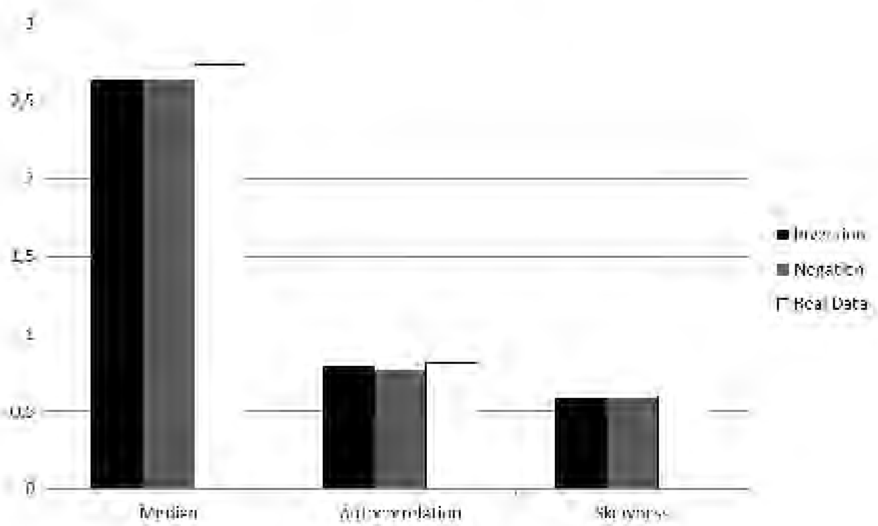

Fig. 20 Comparison of medians, autocorrelation coefficients and skewness of the original and synthesized irradiance.

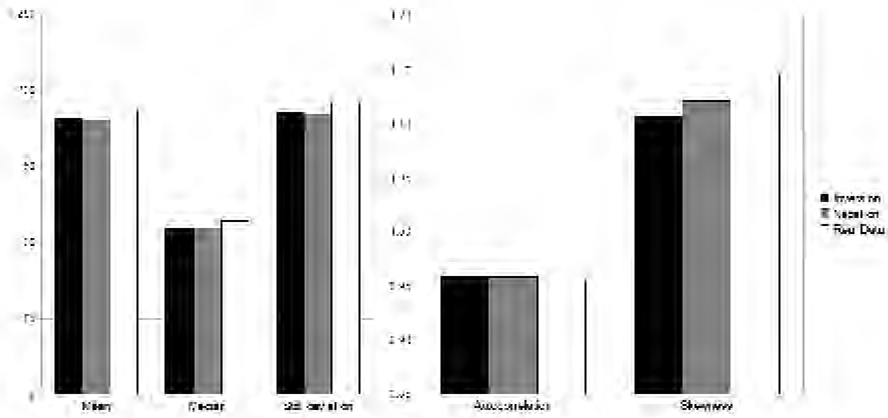

Fig. 21 Comparison of medians, autocorrelation coefficients and skewness of the original and synthesized irradiance, with exclusion of zero values of irradiance.

Acknowledgments

The research was supported by the Polish Ministry of Science and Higher Education under the grant N N519 580238, and by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund. The streamflow data were provided by the Institute of Meteorology and Water Management (IMGW), Poland.

References

1. Agrawal M.R., Boland J., Ridley B. (2013) Analysis of wind farm output: estimation of volatility using high-frequency data. *Environmental Modeling Assessment*, 18:481-492.
2. Aksoy H., Bayazit M. (2000) A model for daily flows of intermittent streams. *Hydrological Processes*, 14:1725-1744.
3. Alexandridis A., Zapranis A. (2013) Wind derivatives: modeling and pricing. *Computationl Economics*, 41:299-326.
4. Anctil F., Coulibaly P. (2004) Wavelet analysis of the inter-annual variability in southern Quebec stream flow. *Journal of Climate*, 17:163-173.
5. Bayazit M., Onoz B., Aksoy H. (2001) Nonparametric streamflow simulation by wavelet or Fourier analysis. *Hydrological Science Journal*, 46(4):623-634.
6. Box G.E.P., Jenkins G.M. (1976) Time Series Analysis: Forecsting and Control. Holden-Day, San Francisco.
7. Bronshtein I.N., Semendyayev K.A., Musiol G., Mühlig H. (2007) *Handbook of Mathematics*. Springer.
8. Carlstein E., Do K.-A., Hall P., Hestenberg T., Künsch H.R. (1998) Matched-block bootstrap for dependent data. *Bernoulli*, 4(3):305-326.
9. Dadaser-Celik F., Sefan H.G. (08) Lake evaporation response to climate in Minnesota. Project Report No. 506. University of Minnesota.

10. Dalton J. (1802) Experimental essays on the constitution of mixed gases; on the force of steam or vapor from water and other liquids in different temperatures, both in a Torricellian vacuum and in air; on evaporation and on the expansion of gases by heat. *Mem. Manchester Liter. and Phil. Soc.*, 5–11:535–602.
11. Davison A.C., Hinkley D.V. (1997) Bootstrap Methods and their Application. Cambridge University Press, Cambridge, MA.
12. Dutta D., Welsh W.D., Vaze J., Kim Sh.S.H., Nicholls D. (2012) A comparative evaluation of short-time streamflow forecasting using time series analysis and rainfall-runoff models in ewater source. *Water Resources Management*, 26:4397-4415.
13. Efron B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1):1-26.
14. Efron B., Tibshirani R.J. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statistical Science*, 1:54-96.
15. Freedman D.A. (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals od Statistics*, 12:827-842.
16. Freedman D.A., Peters S.C. (1984) Bootstrapping a regression equation: some empirical results. Journal of American Statistical Association. 79:97-106.
17. Helsel D.R., Hirsch R.M. (1992) *Statistical Methods in Water Resources.* Elsevier, New York.
18. Ishak A.M., Remesan R., Srivastava P.K., Islam T., Han D. (2013) Error correction modelling of winf speed through hydro-meteorological parameters and mesoscale model: a hybrid approach. *Water Resources Mnagement*, 27:1-23.
19. Jeong C., Shin J.-Y., Kim T., Heo J.-H. (2012) Monthly precipitation forecasting with a neuro-fuzzy model. *Water Resources Management*, 26:4467-4483.
20. Kisi O. (2011) Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resources Management*, 25:579-600.
21. Kottegota N.T. (1980) *Stochastic water resources technology*. Macmillan Press.
22. Koutsoyiannis D. (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resorces Research*. 36(6):1519-1533.
23. Koźmiński C., Michalska B. (2004) Zmienność usłonecznienia rzeczywistego w Polsce. *Acta Agrophysica*, 291-305.
24. Kumar P., Foufoula-Georgiou E. (1993) A multicomponent decomposition of spatial rainfall fields 1. Segregation of large- and small-scale features using wavelet transforms. *Water Resources Management*, 29(8):2515-2532.
25. Kumar D.N., Lall U., Petersen M.R. (2000) Multisite disaggregation of monthly to daily streamflow. *Water Resources Research*, 36(7):1823-1833.
26. Lall U., Sharma A. (1996) A nearest neighbor bootstrap for resampling hydrological time series. *Water Resources Research*, 32(3):679-693.
27. Lange M., Focken U. (2006) Physical Approach to Short-Term Wind Power Prediction. Springer.
28. Moss M.E., Tasker G.D. (1991) An intercomparison of hydrological network-design technologies. *Hydrological Sciences Journal*, 36(3):209-221.
29. Myers D.R. (2003) Solar radiation modeling and measurements for renewable energy applications: data and model quality. *NREL Report NREL/CP-560-33620*.
30. Raman H., Sunilkumar N. (1995) Multivariate modeling of water resources time series using artificial neural network. *Hydrological Science Journal*, 40(2):145-163.
31. Sang Y.-F. (2013) Improved wavelet modeling framework for hydrologic time series forecasting. *Water Resources Management*, 27:2807-2821.
32. Sartori E. (2000) A critical review on equations employed for the calculation of the evaporation rate from free water surfaces. *Solar Energy*, 68(1):77-89.
33. Savitzky, A.; Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36 (8): 1627–1639. doi:10.1021/ac60214a047.
34. Senthil kumar A.R., Goyal M.K., Ojha C.S.P., Singh R.D., Swamee P.K., Nema R.K. (2013) Application of ANN, fuzzy logic and decision tree algoritms for the development of reservoir operating rules. *Water Resources Management*, 27:911-925.

35. Sharma A., Lall U. (1999) A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, 48:361-371.
36. Sharma A., Tarboton D.G., Lall U. (1997) Stream simulation: a nonparametric approach. *Water Resources Research*, 33(2):291-308.
37. Singh V.P., Xu C.Y. (1997) Evaluation and generalization of 13 mass-transfer equations for determining free water evaporation. *Hydrological Processes*, 11:311-323.
38. Smith L.C., Turcotte D.L., Sacke B.I. (1998) Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes*, 12:233-249.
39. Srinivas V.V., Srinivasan K. (2000) Post-blackening approach for modelling dependent annual streamflows. *Journal of Hydrology*, 230(1-2):86-126.
40. Srinivas V.V., Srinivasan K. (2001) Post-blackening approach for modelling periodic streamflows. *Journal of Hydrology*, 241(3-4):221-269.
41. Srinivas V.V., Srinivasan K. (2005) Matched block bootstrap for resampling multiseason hydrologic time series. *Hydrological Processes*. 19:3659-3682.
42. Tasker G.D. (1987) Comparison of methods for estimating low flow chracteristics of streams. *Water Resources Bulletin*, 23(6):1077-1083.
43. Tong H. (Ed.) (1990) Non-linear time series: a dynamical system approach. Clarendon Press, Oxford.
44. Valizadeh N., El-Shafie A. (2013) Forecasting the level of reservoirs using multiple input fuzzification in ANFIS. *Water Resources Management*, 27:3319-3331.
45. Vogel R.M., Shallcross A.L. (1996) The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, 32(6):1875-1882.
46. Unal N.E., Aksoy H., Akar T. (2004) Annual and monthly rainfall data generation schemes. *Stochastic Environmental Research and Risk Assessment*, 18:245-257.
47. Wang W.S., Ding J. (2003) Wavelet network model and its application to the prediction of hydrology. *Nature and Science*, 1(1):67-71.
48. Wang W.S., Ding J. (2007) A multivariate nonparametric model for synthetic generation of daily streamflow. *Hydrological Processes*. 21:1764-1771.
49. Wang W.S., Jin J.L., Li Y.Q. (2009) Prediction of inflow at Three Gorges Dam in Yangtze River with wavelet network model. *Water Resources Management*, 23:2791-2803.
50. Wang W.S., Hu Sh., Li Y. (2011) Wavelet transform method for synthetic generation of daily streamflow. *Water Resources Management*, 25:41-57.
51. Yang N., Mei Y., Zhou Ch. (2012) An optimal reservoir operation model based on ecological requirement and its effect on electricity generation, *Water Resources Management*. 26:4019-4028.
52. Yevjevich V. (1972) *Stochastic processes in hydrology*. Water Resources Publications, Colorado.
53. Zhu X., Genton M.G. (2012) Short-term wind speed forecasting for power system operation. *International Statistical Review*, 80(1):2-23.