

Possibility of the modelling of electricity production from hydropower

Agnieszka Operacz^{1,*}, *Bartosz Szelaż*², and *Mads Grahl-Madsen*³

¹ University of Agriculture in Krakow, Faculty of Environmental Engineering and Land Surveying, Department of Sanitary Engineering and Water Management, Al. Mickiewicza 21, 31-120 Krakow, Poland

² Kielce University of Technology, Al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland

³ 77 Construction USA CORP, 114 Essex St Ste 308 Rochelle Park, NJ 07662 USA

Abstract. In hydropower plants benefits depends on available flow. The paper presents a hybrid model for forecasting the operation of a hydropower plant, including the production of electricity. The possibility of mathematical modeling was chosen to show connections between observed in the past hydrological conditions (available flow) and energy deliver in the future. The available flow which is not enough for start turbines was forecasting by logistic regression model. The opposite situation when the flow starts turbine to produce energy, regression models (the support vector machines SVM, random forest RF, k nearest neighbour k-NN) were used. Results from hybrid model were compared with chosen data-mining methods. The possibility of forecasting of the length of periods when hydropower plant will be working could be very useful. It provides the prognosis of energy value which could be produced from hydropower plant. From the investors' point of view the economic justification for the execution of the project based on the future energy producing could be a main criteria to realize or buy/sell hydropower plant. Also the secondary importance could be a possibility of planning review and maintenance work. Knowledge of power plant working periods could be a base for assessing a potential production from hydropower plant.

1 Introduction

With the vast changes that have taken place in the world over past years there is now a great demand for renewable energy and countries are being required to generate and utilize a specific amount of renewable energy by the year 2020. In recent years, dynamic growth of expenditure on environmental protection has been observed in European Union (EU) countries [1]. Access to EU funds has allowed significant resources for the financing of investment and ecological activities to accumulate in some countries [2]. Renewable power engineering sector investments are recognised, by virtue of the European Parliament and Council Directive 2009/28/EC of 23 April 2009 on the promotion of the use of energy from renewable sources [3] as important sources of energy, supported due to the care about the environment. Thus, the benefits from execution of such investments are highly valued in both regional and global scale, especially in the aspect of climatic changes and progressing pollution of the environment. To confirm the priority of sources of energy recognised as renewable and friendly to the environment, the directive imposes the obligation of generating and managing resources of renewable sources energy, thus defining the objectives for the member countries, including Poland – at 15% of the gross final consumption of energy by 2020

[4]. A large amount of investment has been made during recent years and the advancement of technology has enabled countries to produce renewable energy more cost effectively [5]. An energy production forecast, which gives information about how much energy will be produced effectively by a certain power station in a certain period can be useful for optimising the marketing of a renewable energy [6, 7]. Such regional hydroelectric energy projections can then be used to support energy resource planning and also to evaluate the climate-related risk for long-term power marketing activities in further investigations [8]. Giving the status of a renewable source to such facilities also enables to get preferential funding sources as well as obtain certificates of origin, which greatly improves the economic efficiency of the project [4]. In run-of-the-river power plants benefits depend strictly on the available river flow. The operating principle of hydroelectric power plants is based on utilizing the potential energy of water, which constitutes the basic component of the plant. In other words, water is crucially important in energy production in hydroelectric power plants [9]. On the other hand, the hydropower plants may degraded of water habitats, which is reflected in reduced abundance and diversity of fishspecies and other aquatic organisms [10].

* Corresponding author: a.operacz@ur.krakow.pl

Forecasting production of electricity generated in hydropower plants most often refer to large pumped-storage power stations, i.e. those with two water reservoirs (the upper one and the lower one) along with the possibility of water management [11]. The review of the literature shows that models executed with the methods of artificial neuron networks [12, 13, 14] and their modifications [15, 16] were used for calculations. These models required implementation of complex mathematical algorithms.

However, from a practical point of view, information solely about whether the analysed facility is functioning or not is insufficient. A prospective investor, or the owner of the hydropower plant is interested in a specific value of production of electricity and a period in which he could rationally use excess of the produced energy. Therefore, development of the so-called hybrid models appears to be advisable in modelling operation of run-of-the-river power plants, which constitute the combination of classification models (forecasting of the periods in which the analysed power plant does not generate energy) and a regression model (simulation of discrete values of the values of electricity).

The usefulness and universality of used methods is demonstrated by the fact that they are commonly used in the water-sanitary sector for the development of short-term water usage forecasts for the waterworks and sewerage systems as well as for the sewage treatment plants' optimization [17].

2 Example database

The proposed hydropower plant is located at the existing Dillon Dam, Ohio, United States of America. The dam was built on the Licking River near the town Zanesville. The Dillon Dam near Dillon Falls water-level indicator is located directly below the barrier closing the Dillon Lake (Fig. 1). The information for this water-level indicator available in the database of daily flow

observations cover the years since 1939. The Dillon Dam impounds a reservoir but the proposal hydropower plant will be operated on a strict run of the river mode.

For described analyses, series of daily flows in the Licking River in United States were used as an example database from USGS Water Data.

A sample turbine solution has been chosen, along with simultaneous reduction of production costs of the turbine sets by use of four identical units. The proposed turbine equipment includes 4 turbines with the nominal discharge $20 \text{ m}^3/\text{s}$ and the installed capacity of 1,700 kW each. The simulation conducted within this article is an example, without discussing alternative solutions, if any, that could be economically more justified.

For the Dillon Dam near Dillon Falls water-level indicator, daily observations of the flow of water are available for the period from 01.10.1939 to 30.09.1991 [18]. The Dillon Reservoir was completed in 1961 for flood control because the initial database for further calculations was limited to the this year, so that the developed forecast model could be applied to the power plant operated in the run-of-river system without the possibility of managing water in the reservoir.

The value of the power head was assumed in accordance with the characteristics of the existing ground dam as equal to 10.36 m. In reality, the value of the head is strictly dependent on the flow. The run-of-river hydropower plant will not be operated under flood conditions due to the necessity of maintaining free passage of flood waters.

The hydropower plant starts to produce electricity when the flow in the Licking River achieves $8 \text{ m}^3/\text{s}$, which activates the first turbine. The maximum discharge of a single turbine is $24 \text{ m}^3/\text{s}$. When the river flow achieves the level of $32 \text{ m}^3/\text{s}$, the second turbine is activated, and the third and fourth turbines are activated at the flow of $56 \text{ m}^3/\text{s}$ and $80 \text{ m}^3/\text{s}$, respectively. In case of flows below $8 \text{ m}^3/\text{s}$ and above $140 \text{ m}^3/\text{s}$, the power plant will not be operated (Fig. 2).

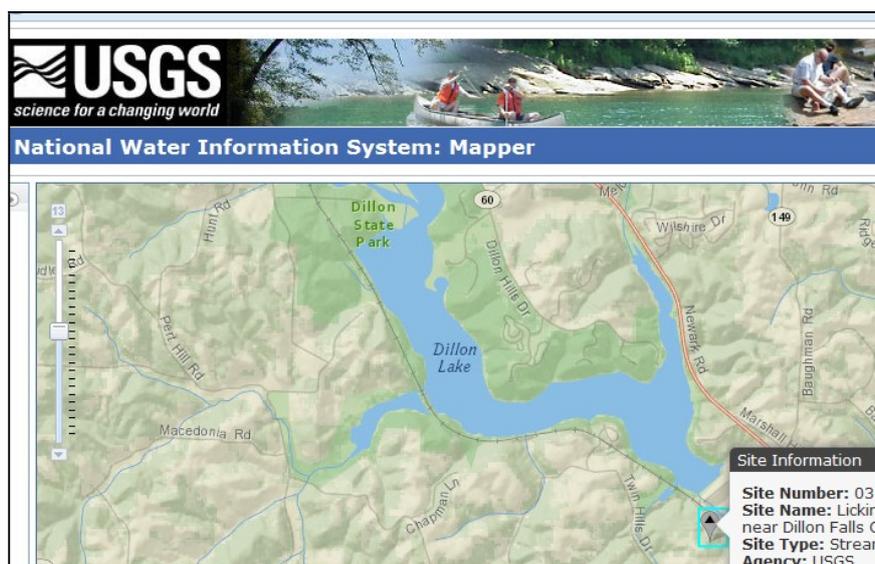


Fig. 1. Watergauge „Dillon Dam near Dillon Falls” on Licking River (USGS Water Data)

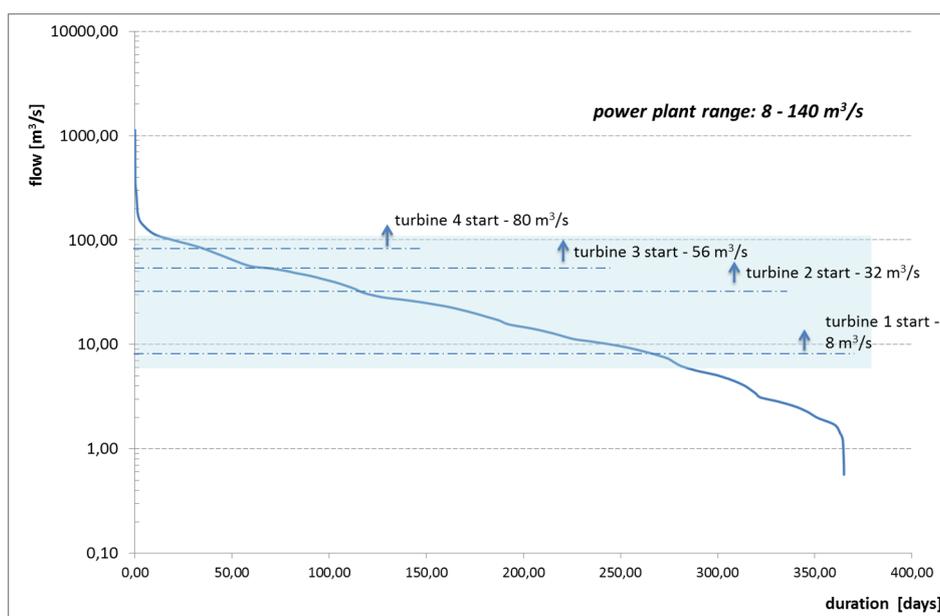


Fig. 2. The curve of the sums of duration along with the higher for the Dillon profile on the Licking River along with characteristics of the power plant operation.

Due to the fact the analysed power plant has the concept nature, the temporary achieved power of the turbine P [kW] is calculated from the formula:

$$P = 9.81 \cdot Q \cdot H \cdot \eta \quad (1)$$

where: Q – disposable temporary flow [m^3/s], H – head, η – turbine performance according to Table 1, 9.81 – the value of normal acceleration of gravity [m/s^2].

Table 1. Turbine performance characteristics

Q [m^3/s]	η [-]
24.0	0.76
22.0	0.79
20.0	0.85
18.0	0.81
16.0	0.77
14.0	0.71
12.0	0.62
10.0	0.48
8.0	0.27

The 24-h production of energy was determined from the formula:

$$E = 24 \cdot P \quad (2)$$

On the basis of the equation (1) and 2), the curve $E = f(Q, H, \eta)$ was determined. Moreover, using the dependences 1 and 2, the theoretical 1–14-day production of electricity was calculated.

3 Long-term electricity forecast

This publication presents a hybrid statistical model for forecasting operation of a hydropower plant, including production of electricity. The obtained results of the simulation were compared with the forecasts from typical regression models. Due to the fact that the

analysed power plant has the concept nature and no long term measurement series of the above parameter were available, the dependence between flow intensity and the parameters of the initially selected turbines was used to determine the theoretical quantity of energy. The possibility of forecasting periods (1–14 days) was included in the developed model, during which the power plant is not functioning, using logistic regression. To forecast discrete values of energy, the selected data mining methods were used (support vector, the k nearest neighbour, random forests). The logit model constitutes a simple and clear regression dependence and is one of classification models commonly used in economics and medicine and is often implemented in statistical software packages (R, STATISTICA, SPSS, etc.).

The paper calls for two methods of calculation of electricity produced by the hydropower plant. In the first case, on the basis of the measurement results of the flow rate $Q(t-i)$ and the determined theoretical energy production, long-term forecasts were provided for theoretical production of energy $E(t = 1-14$ days). In the second case, the hybrid model was used to forecast electricity. In the first stage of the analyses, a classification type of the model was planned, which will allow to identify the periods where the planned power plant will not be operated – the application of logistic regression was considered for this purpose. When calculations provided with the logistic regression model showed that the turbines will produce energy in the consecutive days ($t = 1-14$), long-term forecasts were provided with regression models. Prediction of electricity $E(t) > 0$ was executed on the basis of the value of the flow from the preceding days $Q(t-i)$.

In the logistic regression model, the probability (p) of occurrence of an event that would consist in production of electricity with the turbines or its lack may be expressed in general with the dependence:

$$p = P(Y|X_1 = x_1, X_2 = x_2, \dots, X_j = x_j) = \frac{\exp(\beta_0 + \sum_{i=1}^j \beta_i x_i)}{1 + \exp(\beta_0 + \sum_{i=1}^j \beta_i x_i)} \quad (3)$$

where: β_0 – absolute term, $\beta_1, \beta_2, \dots, \beta_j$ – regression coefficients determined with the maximum likelihood method, X_j - dependent variables, which include here: the daily inflow to the power plant in the previous days, i.e. $Q(t-i)$.

Assessment of the prediction capacity of the logistic regression model (accuracy of forecasts) was provided on the basis of sensitivity (SENS), specificity (SPEC) and calculation error (R_z^2). To assess prediction capacity of the model, the calculated values of McFadden's R^2 and Cox-Snell R^2 coefficient and the Akaike information criterion were also used.

The obtained result of the logistic regression model is the probability of operation of the water power plant (p). Thus, in the case when the value p defined with the formula (3) for the adopted independent variables is smaller than $p = 0.5$, production of electricity is 0. In the case when the value of probability p determined with the formula (3) is larger than 0.5, the value $E > 0$ and then the regression model is used.

To calculate flow rate, three methods of data mining were used: support vector, the k nearest neighbours and random forests.

To make the training process appropriate, and then to properly assess the performance of a statistical model, the data were partitioned into the training set (75%), and the validating and testing set (25%). Prior to the start of the construction of mathematical models, input and

$$J(x_i, x_j) = \begin{cases} 1, & \text{when } x_i \text{ is one of } K \text{ of nearest neighbours of } x_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

In the computations shown above, Euclidean (employed here) and Mahalanobis distances are most frequently used. The number of neighbours (K) was established by trial and error, seeking such a value of K , for which the model devised would show the best predictive abilities.

The random forests algorithm was proposed by [20] and it is a development of the bootstrap method. In the first stage, k-time sampling of the n-element training set is done, allowing repetitions, and then regression trees are created based on the obtained sets. The process of their construction in reference to the classic algorithm was modified so as to make the best breaking down in each node of the tree not on the basis of all, but random attributes (the explaining variables). In this way, k regression trees are obtained that make up the forest on the basis of which the forecast is determined that consists in calculation of arithmetical mean of individual forecasts of single trees as a result of the entire model.

To assess prediction capacity of the developed regression models, the values were used: mean absolute error (MAE), mean absolute percentage error (MAPE) and correlation coefficient (R).

output data normalization was performed by means of normal form transformation.

Support Vector Machines (SVM) cover a group of methods developed by Vapnik [19] first exclusively for classification purposes, which expanded over time to include regression issues (SVR). For that reason, the dependence between the model output and input variables can be non-linear. As a result, in this method a non-linear transformation of N – dimensional space to K – dimensional feature space of much larger size is applied. In this study, the support vector regression method with the radial kernel function was applied to predict electric energy.

The k - nearest neighbour method (k-NN) is one of the simplest non-parametric methods, and like those already mentioned can be applied to classification and regression problems [20, 21]. In this case, the dependent variable prediction is expressed by formula:

$$\hat{y} = \frac{1}{K} \cdot \sum_{i=1}^K y_i \cdot J(x_i, x_j) \quad (4)$$

where: x_i is one of K number of nearest neighbours of x_j when the distance $d(x_i, x_j)$ belongs the smallest distances between observations from set $Z_N = \{(x_1, y_1), \dots, (x_n, y_n)\} \in R^{m+1}$ where: $x_i = (x_{1,i}, \dots, x_{p,i})$ is the i -th vector of independent variables with m number of coordinates, y_i is the i -th dependent variable, N – number of observations, $J(x_i, x_j)$ – function of the form:

4 Results

On the basis of historical measurement series of flows, one may conclude that the flow rate in the analysed water-level section varied from 1.25 m³/s to 1019 m³/s, with the average value equal to 21.66 m³/s.

The analysis of data showed that over a major period of time, i.e. 104 days a year, the analysed hydropower plant will not produce energy due to too low flows run in the river (below the so-called turbine start) or flood flows that force lowering of impoundage and turning off the power plant.

With the historical data and the knowledge that the power plant is operating in the range of the flows of $Q = <8; 140 \text{ m}^3/\text{s}>$, the input data $Q(t-i)$ were identified, where: $i = 1, 2, 3, 4, 5 \dots n, n = 14$, and the output data for the logit model. When the determined value of the daily production of energy for of the period of 1–7 days was larger than 0, the output value of the logistic regression model equal to 1 (the power plant running), or 0 was assigned. The calculations proved (Table 2) that the logit models developed on the basis of $Q(t-1), Q(t-2), Q(t-3)$, and $Q(t-4)$ in case of $E(t = 1-14 \text{ days})$ feature satisfactory predictive capacity with statistically significant parameters at the adopted confidence limit $p = 0.05$.

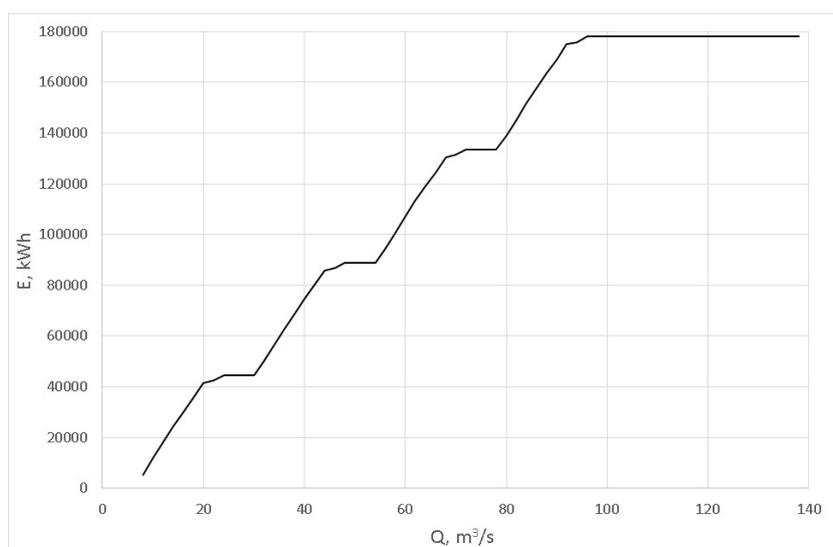


Fig. 3. The curve of daily production of electricity (theoretical and approximated) depending on the flow rate.

The values of the parameters SENS, SPEC and R_z^2 in these models change in the ranges 92.18–95.15%, 82.95–92.47% and 86.24–92.14% respectively, which indicates satisfactory matching of the results of the calculations to the measurements.

In the next stage of the calculations, based on discrete values of production of 1- to 14-day production of electricity and the input data $Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-4)$, the mathematical models were developed for prediction of E with support vector machines, k-nearest neighbour and random forest: the results of the obtained calculations are presented in Table 3. Table 4 presents the results of the calculations of the theoretical production of energy by a hydropower plant obtained based on the traditional support vector, random forests and k nearest neighbours methods. These calculations included also values in a time series, when production of energy during the analysed period was zero and the power plant did not produce electricity. It was found from the analyses that error in the forecast of electricity

generated by the power plant is reduced along with extension of the period for which the forecast is provided. Tables 3 and 6 show that the lowest values of errors in the forecast of energy was received with the support vector methods, both in the classic and hybrid models. In turn, the largest values of errors in the forecast of energy were obtained with the models in which the k nearest neighbours method was used. It follows from the conducted analyses (Tables 3 and 4) that smaller values of errors in the forecast of electricity were obtained with the hybrid model than with the classic models. These differences result from the fact that the days / periods in the developed mathematical model in which the power plant is not functioning are forecast with satisfactory accuracy with logistic regression (Table 2). In the regression models in which only the SVM, RF or k-NN method was used, the errors in the forecast of production of energy during the period when it is not generated are considerably larger than those obtained in the hybrid model.

Table 2. The summary of the values of parameters (β) and the corresponding test probabilities (p) in logistic regression models and parameters of matching the results of calculations for the measurements.

variables	β_0	Q(t-4)	Q(t-3)	Q(t-2)	Q(t-1)	SENS, %	SPEC, %	R_z^2 , %	$R^2_{McFadden}$		AIC
1 day	-1.689	0.039	0.021	0.096	-0.023	95.18	92.47	92.14	0.496	0.497	5175
2 days	-18.733	0.057	-0.008	-0.586	3.226	94.36	89.14	91.76	0.545	0.543	4873
3 days	-4.649	0.016	-0.005	-0.173	0.75	95.24	88.02	91.19	0.588	0.563	4288
4 days	-4.744	0.004	-0.001	-0.146	0.774	93.91	86.51	89.52	0.640	0.597	3817
5 days	-4.507	0.014	-0.009	-0.13	0.743	93.04	85.03	88.14	0.640	0.62	3826
6 days	-4.297	0.019	-0.008	-0.134	0.725	92.69	83.88	87.15	0.635	0.603	3931
7 days	-4.130	0.023	-0.009	-0.112	0.688	92.18	82.95	86.24	0.624	0.601	4097
8 days	-9.488	0.114	0.183	0.154	0.212	96.10	93.86	95.33	0.839	0.660	1539
9 days	-8.685	-0.029	0.348	-0.064	0.651	93.63	95.93	95.17	0.818	0.646	1715
10 days	-8.753	0.231	0.186	0.016	0.576	96.10	93.62	95.31	0.818	0.641	1690
11 days	-7.988	-0.032	0.311	-0.074	0.723	95.47	93.18	94.77	0.795	0.626	1876
12 days	-7.679	-0.037	0.309	-0.085	0.737	95.22	93.10	94.59	0.784	0.616	1955
13 days	-7.398	-0.026	0.298	-0.085	0.738	95.11	92.48	94.34	0.773	0.606	2030
14 days	-7.205	-0.018	0.304	-0.091	0.746	94.98	91.54	94.01	0.764	0.597	2084

Table 3. The parameters for matching measurement data for the results of the calculations obtained with artificial neural networks for forecasting production of electricity during the period of 1–14 days.

E, kWh	SVM			RF			k-NN		
	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE
t=1	0.82	10549	34.4	0.81	11791	36.67	0.72	13301	41.36
t=2	0.89	15196	29.6	0.81	17907	32.52	0.79	18366	33.17
t=3	0.91	17889	25.3	0.84	19216	28.51	0.78	20726	30.63
t=4	0.95	18641	19.6	0.89	19233	20.36	0.78	22076	23.37
t=5	0.96	18818	18.1	0.88	19932	19.75	0.81	21791	21.48
t=6	0.97	18997	16.9	0.92	21143	19.66	0.79	24514	22.65
t=7	0.98	18521	15.7	0.93	20915	18.50	0.84	23319	20.43
t=8	0.99	19263	15.02	0.94	21949	17.29	0.85	24413	19.23
t=9	0.98	19083	12.13	0.95	25839	15.27	0.85	28900	17.08
t=10	0.99	22151	10.64	0.93	23802	14.83	0.92	24011	14.98
t=11	0.99	21044	9.16	0.95	22032	12.46	0.91	22900	12.97
t=12	0.98	22663	8.26	0.96	23979	11.83	0.91	25271	12.50
t=13	0.99	22039	8.09	0.95	24685	11.14	0.92	25530	11.53
t=14	0.99	22743	8.07	0.94	23783	10.04	0.89	23639	10.60

Table 4. The parameters for matching measurement data for the results of the calculations obtained with artificial neural networks for forecasting production of electricity during the period of 1–14 days.

E, kWh	SVM			RF			k-NN		
	R	MAE	MAPE	R	MAE	MAPE	R	MAE	MAPE
t=1	0.74	11789	36.21	0.61	15133	43.06	0.51	18738	49.21
t=2	0.76	18146	35.45	0.51	21832	36.65	0.58	25516	43.15
t=3	0.77	18575	27.85	0.62	21649	30.12	0.57	27887	39.28
t=4	0.81	22412	22.55	0.62	21914	26.20	0.56	31447	33.29
t=5	0.82	19390	18.33	0.59	22968	26.76	0.54	32287	28.69
t=6	0.82	21756	19.24	0.72	24548	27.82	0.50	38743	33.15
t=7	0.83	20719	15.98	0.69	27071	26.95	0.52	37322	30.70
t=8	0.84	23314	15.90	0.72	27256	25.47	0.59	34976	24.26
t=9	0.83	21349	12.64	0.60	29933	19.69	0.55	45307	25.12
t=10	0.82	23609	12.93	0.72	27075	18.42	0.63	32551	20.85
t=11	0.83	20549	10.74	0.61	28356	20.79	0.62	31707	16.89
t=12	0.83	21136	10.66	0.63	29336	17.51	0.59	39288	17.98
t=13	0.84	21830	10.81	0.76	28716	13.87	0.64	36348	15.87
t=14	0.84	27320	10.99	0.70	32847	17.73	0.59	35842	15.96

5 Conclusions

It follows from the provided analyses that a hybrid model presented in the publication may be useful in modelling production of electricity in a hydropower plant operated on the run-of-river principle, i.e. in a non-continuous way, resulting from periods of low water and freshet of flood nature. The obtained values of errors in the forecast of electricity produced by the hydropower plant with the support vector, random forests and k nearest neighbours method are higher than with the proposed mathematical model. Within the analysed methods (classical and the proposed hybrid model), the lowest values of errors in the forecast of electricity were received with the support vector method. On the basis of the calculations, the logistic regression was found out to be applicable in forecasting periods during which a power plant is not operated. The possibility of

forecasting the length of idle periods in operation of a hydropower plant is very important because it offers the possibility of planning inspection and maintenance work for the individual components of a hydropower plant, which is included in the models developed so far in a limited degree.

Due to the fact that the small hydropower plant analysed in this paper was a concept design, and the used data and their further analysis were aimed at showing the possibility of application of the method presented in the paper for forecasting electricity and the length of the period in which the power plant is not functioning, which is why verification of the developed methodology is recommended based on the data in the functioning facility.

Acknowledgement

Publication supported by the Polish Ministry of Science and Higher Education as a part of the program of activities disseminating science from the project „Organization of the First International Science Conference – Ecological and Environmental Engineering”, 26-29 June 2018, Kraków.

References

1. Tsireme AI, Nikolaou EI, Georgantzis N, Tsagarakis KP. *Clean Technol Environ Policy* **14(5)**:953–964 (2012)
2. Kulczycka J, Smol M. *Clean Technol Environ Policy* **18**:829–842 (2016)
3. European Parliament and Council Directive 2009/28/EC of 23 April 2009 on the promotion of the use of energy from renewable sources
4. Operacz A., Tomaszewska B. B. *Environ Sci Pollut Res* **23**, 18953. (2016)
5. Abolhosseini S., Heshmati A., Altmann J. A Review of Renewable Energy Supply and Energy Efficiency Technologies. IZA Discussion Paper 8145. (2014)
6. Dmitrieva K. Forecasting of a hydropower plant energy production. Master's Thesis in Computer Science, Ostfold University College, Halden, Norway. (2015)
7. Operacz A. *Renew and Sustain Energy Reviews* **75 (9)** 1453-1463. (2017)
8. Kao S.-C., Sale M.J., Ashfaq M., Martinez R.U., Kaiser D.P., Wei Y., Diffenbaugh N.S. *Energy* **80**, 239. (2015)
9. Karakoyun Y., Yumurtaci Z., Donmez A.H.. *Clean Technol. Environ. Policy*, **18**: 583-591. (2016)
10. Wałęga A., Młyński D., Kokoszka R., Miernik W. *Pol. J. Environ. Studies* **24(6)**, 2663-2676 (2015)
11. Catalao J.P.S., Mariano S.J.P.S., Mendes V.M.F., Ferreira L.A.F.M. *Euro. Trans. Electr. Power* **20**, 172. (2010).
12. Abdulkadir T.S., Salami A.W., Anwar A.R., Kareem A.G. *Ethiopian Journal of Env Stud and Manag* **6**, 310. (2013)
13. Estoperez N, Nagasaka N, *Clean Technol Environ Policy* **8**:123–130. (2006)
14. Grieu S., Thiéry F., Traoré A., Nguyen T.P., Barreau M., Polit M. *Chemical Engineering Journal*. **116**, 1. (2006)
15. Gang L., Yongjun S., Yong H., Xiufeng L., Qiyu T. *Mathematical Problems in Engineering*. Volume **2014**, Article ID 381387 (2014)
16. Gang L., Bao-Jian L., Xu-Guang Y., Chun-Tian Ch. *Energies* **8**, 1228. (2015)
17. Wałęga A., Bergel T. *Infrastruktura i Ekologia Terenów Wiejskich* **5/2009**, 183-195 (2009)
18. USGS Water Data – National Water Information System:
http://waterdata.usgs.gov/nwis/inventory?agency_code=USGS&site_no=03147500 (02.02.2018).
19. Vapnik V. *Statistical Learning Theory*. John Wiley and Sons. New York. (1998)
20. Liaw A., Wiener M. *Classification and regression by random forests*. *The R News* **2/3**, 18. (2002)
21. Yingying L. *Water Science at Technology* **70**, 1488. (2014)