Forecasting the Nysa Kłodzka flow rate in order to predict the available flow for a run-off-river (ROR) power plant

Jakub Jurasz^{1,*}, Marcin Wdowikowski²

¹AGH University, Department of Management Engineering, 30-067, Gramatyka 10 St., Cracow, Poland

²Institute of Meteorology and Water Management - NRI, 01-673 Warsaw, Podleśna 61, Poland

Abstract. Hydroelectricity is generally perceived as a stable and predictable power source. However ROR power plant without reservoir energy output is mainly driven by changing flow rate. This study applies artificial neural networks to create flow rate forecasts with one hour lead time. Forecasting models were built for Nysa Kłodzka catchment which possesses significant potential for new hydropower plants development as well as leads to frequent floods. The best of the obtained model gives satisfactory results both in terms of root mean square error (0.6379 m³/s) as well as Nash-Sutcliffe performance indicator (0.9978). Obtained results were compared with currently used forecasting models and were proven to be superior.

1 Introduction

Development of the so-called renewable energy sources (RES) in Poland is driven mainly by the European Union regulations and directives. To a much lesser degree an increasing installed capacity in RESs can be attributed to their superior economic performance and social awareness of the environmental protection. However in future it is the economics that may play a crucial role due to dwindling non-renewable resources and/or their picking up prices. Additionally foreseeable is a reduction in RES costs as well as their increasing efficiency which will undoubtedly lead to their greater role in national energy markets. Unfortunately three the most important representatives of RESs namely: solar, wind and hydropower in some circumstances (e.g. hydroelectricity dependent on current flow rate, photovoltaics and wind energy without energy storage facilities) introduce undesirable variability to the operation of energy system (by some they are denoted as VRE – Variable Renewable Energy). Their non-dispatchable nature results from dependence on naturally varying phenomena such as: wind speed, irradiation and flow rate. Therefore there is not only a need to overcome fluctuating demand side of energy market but also its supply side. Hence the increasing research interest in areas such as: energy yield forecasting from solar [1,2] wind [3] power sources; temporal and spatial complementarity assessment of certain RES

* Corresponding author: jakubkamiljurasz@gmail.com

[4,5]; energy storage [6]; hybrid RES power sources [7]; RES integration to the national power systems [8].

The Directive 2009/28/EC of the European Parliament on the promotion of the use of energy from renewable sources required from Poland to 2020 an achievement of 15% share of renewables in the total of primary generated energy. The national authorities concerns indicate that the goals of the Directive rather won't be fulfilled. These fears were confirmed by the computation of renewable energy (RE) share in the national electricity consumption for the last several years. In 2009 the share of RE was 5.7% while in 2011 it rose to 8% which it is still not enough [9,10].

According to the Polish Energy Regulatory Office (<http://www.ure.gov.pl/>) as on 30.06.2016 the installed capacity in RES amounted to almost 8.25 GW (excluding co-firing biomass in commercial power plants). This was however unevenly distributed among: wind power (68.7%), biomass power plants (15.4%), hydroelectricity (12%), biogas power plants (2.8%) and slightly above 1% in photovoltaics. As can be seen, hydroelectricity amounts to almost 1 GW of the installed capacity in Polish RES. The biggest contribution to this capacity is made by hydro-pumped storage located on rivers (382.7 MW) and run-off-river power plants with capacity greater than 10 MW (311 MW). However, Polish hydroelectricity is dominated by relatively a small run-off-river units. From total of 752 hydropower plants 98.7% has nominal power smaller than 10 MW, from which 76.7% is not greater than 300 kW. According to the RESTOR HYDRO (<http://www.restor-hydro.eu/pl/>) project there is an opportunity to develop over 8 500 pico, micro and even bigger hydro installations in Poland. In order to minimize the environmental impact and financial investments those hydro generators may be built on existing weirs, dams and other lateral structures and become a run-off-the-river power plants. Therefore in future not only the role of small scale hydroelectricity may increase in Poland but in some cases it will play crucial role in smoothing and balancing inherently variable energy generation coming from photovoltaics and wind turbines [11]. Hence it will be a necessity for a power grid operator to possess knowledge on future VRE energy generation capabilities.

Above described situation in Polish renewable energy market and current problems in area of RES enable formulation of a following goal of this study: "Assessment of the possibility of applying artificial neural networks (ANN) to create a flow rate forecast with a one hour lead time for Nysa Kłodzka river and juxtapose obtained results with those coming from current models used by HydroProg [12].

2 Artificial neural networks

The concept of artificial neural networks has been inspired by biological neural networks which operate in human and animal brains to estimate approximate functions based on usually large sets of often unknown input variables. Typical ANN consists of three layers, namely: input, hidden and output each of which performs its idiosyncratic tasks. Additionally the process of ANN model creation can be divided into three phases, namely: teaching, validation and testing each of which is performed on previously distinguished subsets. Thanks to the ANN structure and applicability for non-linear systems they have been widely applied in many areas of research such as: medicine [13]; energy demand [14]; waterworks operation/water demand [15]; drought forecasting [16]. Comprehensive description of ANN and their performance as forecasting tools can be found for example in [17].

Due to the rapid development of computing capabilities of modern computers the process of ANN creation has been in large portion automated and the analysis of various ANN architectures even when the data set is very large tend not to be a big problem. This study applied the Statistica software developed by StatSoft for creating multilayer perceptron (MLP) ANNs. Statistica has Broyden-Fletcher-Goldfarb-Shanno teaching algorithm built in

along with tool dividing data into teaching, validation and testing subsets which correspond to accordingly 70%, 15% and 15% of the initial set. However the area which has not been automated so far is the selection of appropriate explanatory variables - this problem is usually encountered in area of building forecasting models. According to [18] following approaches to selecting most appropriate explanatory variables subsets may be implemented in area water resources modelling:

- based on expertise and former knowledge on investigated system,
- correlation analysis between dependent and independent variables,
- heuristic approaches based on creating various ANN models and adding or subtracting subsequent explanatory variables,
- extracting data from newly created ANN and using it to perform sensitivity analysis,
- combination of tools used in various methods.

On the basis of the previous experiences with ANN models creation [14, 15, 19]] it has been assumed that in this study the explanatory variables subset will be selected based on correlation analysis combined with expertise and reports from previous studies on Nysa Kłodzka catchment properties. Obtained results were then compared with those available at HydroProg hydrological forecasting system (free access: <http://www.klodzko.hydroprog.uni.wroc.pl/>) which is a well-known and established forecasts source for the Nysa Kłodzka catchment [12, 20, 21]. Because the forecasting performance of the models available there is assessed based on two criteria, namely: root-mean squared error (RMSE) and Nash–Sutcliffe [25] model efficiency coefficient (E) they were also applied in this study. They were calculated based on formulas (1) and (2). In case of the RMSE criterion the smaller its value the better is the forecasting model – therefore it is a useful tool to compare various models tested on the same data sets. The second criterion values can range from $-\infty$ to 1 where $E = 1$ indicates the perfect match between modelled and observed values, whereas $E = 0$ informs that the model accuracy is as good as using a mean of the set as a forecast. The Nash-Sutcliffe values less than 0 show that created model can be easily outperformed by simply using the mean of the set to predict system behaviour.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (1)$$

where: n – number of observations in testing subset, e – forecast error

$$E = 1 - \frac{\sum_{t=1}^T (Q_o^t - Q_m^t)^2}{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2} \quad (2)$$

where: $\bar{Q}_o$ - mean value of observed discharges, Q_m^t - modelled discharge at time t , Q_o^t - is observed discharge at time t .

3 Case study – Nysa Kłodzka catchment

The Nysa Kłodzka River is a left tributary of the Odra River, into which it flows at km 181.3 in the Opolskie voivodeship – Southern West Poland. The river's total length is 181.7 km, while the catchment area reaches 4565.7 km². The Nysa Kłodzka is a mountain river, located in Kłodzko Valley. It has an average gradient of 9.05‰, and starts in the Lower Silesia voivodeship, through which it flows nearly 50% of its total length – 89.4 km. The highest elevation within the basin is equal to 1425 m a.s.l. The river supplies several important for the flood protection purposes reservoirs: the Topola, Kozielno, Otmuchow and

Nysa. Its key tributaries are: the Bystrzyca Kłodzka, Biała Łądecka, Bystrzyca Dusznicka, Ścinawka and Buszówka. The Nysa Kłodzka catchment was developed intensively by hydropower facilities in recent years. There are 24 hydroelectric power plants (of either run-of-the-river (ROR) or conventional and dam (CON) type) with capacities from 0,002 to 5 MW (detailed table [9]). The Otmuchow and Nysa-Glebinow power plants stands as the highest capacities (with 4.8 MW and 3.28 MW respectively), while the lowest capacities (several kW each) goes to Nowa Morawa on the Biała Łądecka, Kłodzko on the Bystrzyca Dusznicka and Rynarce on the Ścinawka Niemodlińska. Most of the functioning power plants belongs to the local or national authorities, but there are also commercial companies like (among others) Hydroelektrownie Dolnego Śląsk, TAURON Ekoenergia or Hydroenergia as well as private investors. Total capacity of all actually existing hydropower plants on the Nysa Kłodzka River covers nearly 17.6 MW. [9, 12, 20, 21]

Hydrometeorological data were derived from the national hydrological and meteorological measurement and observational network held by Institute of Meteorology and Water Management – National Research Institute (IMWM-NRI). In the Nysa Kłodzka river catchment there are monitoring structures there are 12 water level gauges including hydrological station in Bardo, and about 12 meteorological stations including synoptic station in Kłodzko (Fig. 1). Water level is measured on all mentioned stations and the flow rate is computed from the relationships based on flow measurements made by IMWM-NRI. All stations in the Nysa Kłodzka river basin are automatically gathered. Data series used in this study becomes from standard measurement system, that can be seen from (<http://www.pogodynka.pl>). Basic time period for water level measurement is 10 minutes.

Because of the very frequent flooding occurrence in the Kłodzko Valley, local authorities decided to build early warning system based on the 22 automatic hydrologic gauging stations where water level is measured every 15 minutes, and 18 automatic meteorological stations (with various sets of measured parameters) recording observations every 15 minutes. That hydrometeorological system is owned by Crisis Management Center in Kłodzko and it is called LSOP (Lokalny System Osłony Przeciwpowodziowej in polish, free access at <http://lsop.powiat.klodzko.pl>). Besides measuring intervals the main difference between LSOP and IMWM-NRI monitoring is that LSOP gives only water level in case of the hydrological data.

Using LSOP hydrological data the HydroProg forecasting system was applied in the recent years [12, 20, 21].

Fig. 1. The Nysa Kłodzka catchment along with location of weather stations and water level gauges from which data were obtained for the purpose of this study.

Over the considered time period (2010-2013) the hourly values of Nysa Kłodzka flow rate exhibited following statistical properties: mean – 20.87 m³/s; standard deviation – 19.75 m³/s; coefficient of variation – 94.65 %; maximal value – 240.78 m³/s; minimal value – 5.22 m³/s. Example of flow rate variation over twelve month period is given in Fig. (2) notice rapidly occurring flooding flow rate peaks and how much time it takes for them to smooth out. Random swells of waters may not only lead to loss of human life, destruction of infrastructure but also have impact on the operation of hydroelectricity.

Fig. 2. Hourly values of Nysa Kłodzka flow rate over the year 2013 measured in Bardo.

4 Forecasting models - results

The ANN forecasting models have been created to predict the flow rate in Bardo (which is located in the upper part of the scheme depicted in Fig. (1). Based on described in the second section approach eight sites with water gauges were selected to use collected there data sets as explanatory variables. Additionally historical hourly flow rates measured in Bardo were also considered as input variables. Expertise indicate that in the case of Nysa Kłodzka catchment the lag time ranging from 1 to 12 hours is sufficient (denoted in this paper as e.g. $t-1$). It has been assumed that forecast will be made with one hour lead time. Table 1 contains calculated correlation coefficients, all of which were significant with $p - \text{value} < 0.05$, therefore all those subsets were used as explanatory variables.

Table 1. Matrix of correlation coefficients between set of explanatory variables and a dependent variable (flow rate in Bardo in time t).

		Location of hydrological water level gauges							
		Bardo	Kłodzko	Szalejów	Międzylesie	Bystrzyca Kłodzka	Wilkanów	Żelazno	Gorzuchów
Time lag in a reference to Bardo	t-1	0.998	0.937	0.851	0.744	0.832	0.750	0.849	0.828
	t-2	0.993	0.940	0.854	0.754	0.835	0.758	0.849	0.829
	t-3	0.987	0.940	0.855	0.764	0.836	0.766	0.849	0.827
	t-4	0.979	0.937	0.854	0.772	0.836	0.772	0.846	0.823
	t-5	0.971	0.932	0.851	0.780	0.833	0.776	0.841	0.817
	t-6	0.962	0.925	0.846	0.785	0.830	0.779	0.835	0.811
	t-7	0.953	0.917	0.840	0.789	0.824	0.779	0.829	0.803
	t-8	0.943	0.907	0.833	0.790	0.817	0.777	0.821	0.795
	t-9	0.934	0.898	0.824	0.790	0.809	0.772	0.813	0.787
	t-10	0.924	0.888	0.815	0.788	0.800	0.767	0.805	0.779
	t-11	0.914	0.878	0.806	0.785	0.790	0.761	0.796	0.771
	t-12	0.904	0.867	0.796	0.780	0.780	0.754	0.787	0.762

According to the literature [22, 23] using meteorological parameters such as precipitation, temperature or humidity in hydrological forecasting models, based on flow hydrodynamic transformation between neighbouring gauges does not lead to an improvement in accuracy.

The so called “Automatic network search” in-built functionality in Statistica was used to construct ANN models. Although it initially suggests to investigate only 20 various ANN structures it has been calculated that exactly 500 different architectures of MLP ANN should be tested. This number results from the following assumptions:

- the minimal and maximal number of neurons in so called hidden layer is a problematic issue which is often very controversial. Therefore an approach presented in [24] has been applied. It states that the minimal number of neurons in hidden layer should be equal to the square root of the product of the number of neurons in input and output layers. Whereas the maximal number should not be greater than three times the minimal number of neurons in the hidden layer plus one,
- five various (namely: *sine*, *linear*, *logistic*, *hyperbolic tangent* and *exponential*) activation functions have been considered in the hidden and output layers.

Therefore assuming 20 various numbers of neurons in hidden layer, 5 different activation functions in both hidden and output layers it is possible to create as much as 500 unique ANN architectures. Created models have been assessed based on criteria previously mentioned (Eq. 1 and 2) but only with a reference to the testing subset which does not take part in the process of ANN creation. Table 2 summarizes the most important characteristics of selected five ANN which outperformed the reaming ones in terms of forecasting accuracy measured as a correlation between observed and models flow rate values within the testing subset. Those five models were further investigated in the next section.

Table 2. Architecture of five ANN models selected for a further investigation (BFGS – stands for Broyden-Fletcher-Goldfarb-Shanno teaching algorithm).

No.	Number of neurons in the hidden layer	Teching alogrithm	Activation function	
			Hidden layer	Output layer
ANN1	23	BFGS 59	Expotential	Sine
ANN2	9	BFGS 97	Hyperbolic tangent	Linear
ANN3	14	BFGS 58	Expotential	Sine
ANN4	22	BFGS 80	Logistic	Linear
ANN5	28	BFGS 124	Hyperbolic tangent	Linear

The original dataset was a 26294 hours long time series. However some missing values in flow rate due to the equipment failure or malfunctions were observed. Therefore this set has been reduced to 15856 hours long subset in order to make sure that each forecast will be created based on a complete set of explanatory variables. Obtained complete set was further divided into three smaller ones, namely:

- teaching subset 11000 cases,
- validation subset 2378 cases,
- testing subset 2378 cases.

Classification of the cases to the individual subsets was performed based on an in-built in the Statistica software algorithm. But in order to avoid type III errors [26] a k-fold cross-validation with k = 10 was conducted. It basically means that the procedure of dividing the original set into three subsets was done ten times and the obtained performance indicators values (RMSE and E) were averaged.

Table 3 summarises the RMSE and E for each of the five ANN models. Observed values are quite similar and differentiable at two decimal places. In terms of Nash-Sutcliffe criterion all models exhibited equal performance and the values of E suggests that created forecasting models explain almost completely the behaviour of modelled phenomena. In case of forecasts generated by models using one hour prediction time period e.g used by HydroProg, exhibited values of E equal to 0.93 what means that all five proposed in this paper models came out to be superior. However it is clear that model ANN2 which had only nine neurons in the hidden layer and used hyperbolic tangent and liner functions as activations ones respectively in hidden and output layer outperformed the remaining ones in terms of RMSE criterion. Although the greatest observed absolute error in case of the ANN2 model was 12.48 m³/s, the mean of all absolute errors was 0.2836 m³/s. Majority (95%) of absolute errors was smaller than 1 m³/s and almost 86% of them was not greater than 0.5 m³/s. Figure (3) depicts the 30 observed and forecasted by the ANN2 model flow rates for 30 subsequent observations.

Table 3. RMSE and Nash-Sutcliffe forecasting metrics values for five ANN models.

	Artificial neural network model				
	ANN1	ANN2	ANN3	ANN4	ANN5
RMSE	0.6498	0.6397	0.6616	0.6456	0.6454
E	0.9987	0.9987	0.9986	0.9987	0.9987

Fig. 3. Relation between modelled (ANN2) and observed flow rate for 30 subsequent observations for The Nysa Kłodzka dataset were obtained for the purpose of this study.

5 Conclusions

The results and observations coming from this study may be applied for the future analysis of the operation of run-off-river power plants with or without reservoir. Information about the flow rate which will be available an hour ahead may give the operator/owner of the ROR type hydropower plant an opportunity to perform better (financially) on the energy market but also to adjust energy yield to the current market needs. It is important to note that flow rate forecasts are currently of a great importance for civilian safety services but in future

the variable renewable energy may start such a huge role in energy system that it will be necessary to put in service all possible measures which ease their inherent stochastic nature.

There seems to emerge several interesting directions for future studies, which should focus on:

- investigating the impact of smaller explanatory variables subsets and longer lag times on the forecast accuracy;
- building hybrid forecasting models based on other machine learning tools and data pre-processing techniques such as wavelet transform;
- scrutinizing the impact of forecast accuracy on the real ROR hydropower energy output;
- creating a simulation and optimization model (which will incorporate data from flow rate forecasts) of a ROR power plant with a reservoir operating on the energy market.

References

1. A. G. R. Vaz, B. Elsinga, W. G. J. H. M. van Sark, M. C. Brito, *An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands* Renewable Energy, **85**, 631-641 (2016)
2. S. Kim, Y. Seo, V. P. Singh, *Estimating Global Solar Irradiance for Optimal Photovoltaic System* Procedia Engineering, **154**, 1237-1242 (2016)
3. S. S. Soman, H. Zareipour, O. Malik, P. Mandal, *A review of wind power and wind speed forecasting methods with different time horizons* In North American Power Symposium (NAPS), 2010 (pp. 1-8). IEEE. (2010)
4. J. Jurasz, A. Piasecki, *Evaluation of the Complementarity of Wind Energy Resources, Solar Radiation and Flowing Water—a Case Study of Pila* Acta Energetica **2** (2016).
5. C. E. Hoicka, I. H. Rowlands *Solar and wind resource complementarity: Advancing options for renewable electricity integration in Ontario, Canada* Renewable Energy, **36(1)**, 97-107 (2011)
6. X. Luo, J. Wang, M. Dooner, J. Clarke, *Overview of current development in electrical energy storage technologies and the application potential in power system operation* Applied Energy, **137**, 511-536, (2015)
7. M. A. Ramli, A. Hiendro, Y. A. Al-Turki, *Techno-economic energy analysis of wind/solar hybrid system: Case study for western coastal area of Saudi Arabia* Renewable Energy, **91**, 374-385, (2016)
8. S. Weitemeyer, D. Kleinhans, T. Vogt, C. Agert, *Integration of Renewable Energy Sources in future power systems: The role of storage* Renewable Energy, **75**, pp. 14-20 (2015)
9. R. Kasperek, M. Wiatkowski: *Hydropower generation on the Nysa Kłodzka river* Ecological Chemistry and Engineering, **21(2)**, pp. 327-336 (2014)
10. Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC.
11. I. Graabak, M. Korpås, *Balancing of variable wind and solar production in Continental Europe with Nordic hydropower—A review of simulation studies* Energy Procedia, **87**, 91-99 (2016)
12. T. Niedzielski, B. Miziński, M. Kryza, P. Netzel, M. Wiczorek, M. Kasprzak, P. Migoń, M. Szymanowski, J. Jeziorska, M. Witek, *HydroProg: a system for hydraulic forecasting in real time, based on the multimodelling approach*, Meteorology Hydrology and Water Management vol. **2** issue 2. IMGW-PIB, Warszawa, pp. 65-72. (2014)

13. U. Smoczyńska, J. Smoczyńska, & R. Tadeusiewicz. *Neural modelling of growth hormone therapy for the prediction of therapy results* Bio-Algorithms and Med-Systems. **11(1)**: 33-45. doi: 10.1515/bams-2014-0021 (2015)
14. J. Jurasz, J. Mikulik, *Day ahead electric power load forecasting by WT-ANN*, Przegląd Elektrotechniczny, **4**, pp. 152-154, (2016)
15. A. Piasecki, J. Jurasz, W. Marszelewski. *Wykorzystanie wielowarstwowych sztucznych sieci neuronowych do średnioterminowego prognozowania poboru wody-studium przypadku*. Ochrona Środowiska **38.2** (2016): 17.
16. J. Adamowski, H. F. Chan *A wavelet neural network conjunction model for groundwater level forecasting*. Journal of Hydrology, **407(1-4)**, 28-40. doi:10.1016/j.jhydrol.2011.06.013 (2011)
17. D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press (2003)
18. G. J. Bowden, G. C. Dandy, H. R. Maier: *Input determination for neural network models in water resources applications. Part 1 – background and methodology*. Journal of Hydrology, Vol. **301**, No. 1–4, pp. 75–92 (2005)
19. A. Piasecki, J. Jurasz, and Skowron, R. *Application of artificial neural networks (ANN) in Lake Drwęckie water level modelling*. Limnological Review, **15(1)**, 21-30. (2015)
20. T. Niedzielski, B. Miziński, D. Yu, *Hydrological forecasting in real time: an experimental integrated approach*, et. al.: Jasiewicz, J., Zwoliński, Z., Mitsova, H., Hengl, T. (Eds.), *Geomorphometry for Geosciences*, Bogucki Wydawnictwo Naukowe, Adam Mickiewicz University in Poznań – Institute of Geoecology and Geoinformation, Poznań, pp. 97–101 (2015)
21. T. Niedzielski, B. Miziński *Real-time hydrograph modelling in the upper Nysa Kłodzka river basin (SW Poland): a two-model hydrologic ensemble prediction approach*. Stochastic Environmental Research and Risk Assessment, Online first: 28 April 2016 DOI: 10.1007/s00477-016-1251-5 (2016)
22. O. Bozorg-Haddad, M. Zarezadeh-Mehrizi, M. Abdi-Dehkordi, H. A. Loáiciga, M. A. Mariño, *A self-tuning ANN model for simulation and forecasting of surface flows*, Water Resources Management, **30(9)**, pp 2907–2929. DOI:10.1007/s11269-016-1301-2 (2016)
23. B. Kang, Y.H. Ku, Y.D Kim, *A case study for ANN-based rainfall–runoff model considering antecedent soil moisture conditions in Imha Dam watershed, Korea*. Environmental Earth Sciences, **74(2)**, pp 1261–1272, DOI:10.1007/s12665-015-4117-0 (2015)
24. K. Shibata, Y. Ikeda, *Effect of number of hidden neurons on learning in large-scale layered neural networks*. Proc. of the ICROS-SICE International Joint Conference, pp. 5008–5013, (2009)
25. J. Nash, and J. V. Sutcliffe, *River flow forecasting through conceptual models part I—A discussion of principles*, Journal of hydrology **10.3** (1970): 282-290
26. F. Mosteller, *A k-sample slippage test for an extreme population*. In: Selected Papers of Frederick Mosteller, Springer, New York pp. 101–109, (2006)