



# Day-ahead energy production in small hydropower plants: uncertainty-aware forecasts through effective coupling of knowledge and data

Korina-Konstantina Drakaki, Georgia-Konstantina Sakki, Ioannis Tsoukalas, Panagiotis Kossieris, and Andreas Efstratiadis

Department of Water Resources & Environmental Engineering, National Technical University of Athens, Greece

**Correspondence:** Korina Konstantina Drakaki (korina.k.drakaki@gmail.com)

Received: 25 June 2021 – Revised: 3 December 2021 – Accepted: 6 December 2021 – Published: 14 January 2022

**Abstract.** Motivated by the challenges induced by the so-called Target Model and the associated changes to the current structure of the energy market, we revisit the problem of day-ahead prediction of power production from Small Hydropower Plants (SHPPs) without storage capacity. Using as an example a typical run-of-river SHPP in Western Greece, we test alternative forecasting schemes (from regression-based to machine learning) that take advantage of different levels of information. In this respect, we investigate whether it is preferable to use as predictor the known energy production of previous days, or to predict the day-ahead inflows and next estimate the resulting energy production via simulation. Our analyses indicate that the second approach becomes clearly more advantageous when the expert's knowledge about the hydrological regime and the technical characteristics of the SHPP is incorporated within the model training procedure. Beyond these, we also focus on the predictive uncertainty that characterize such forecasts, with overarching objective to move beyond the standard, yet risky, point forecasting methods, providing a single expected value of power production. Finally, we discuss the use of the proposed forecasting procedure under uncertainty in the real-world electricity market.

velocity, solar radiation, streamflow), that make particularly difficult to ensure a credible power supply scheduling. At the same time, the new legal framework in energy market called “Target Model”, introduce further complexities to day-ahead trades, thus making short-term forecasting not only a challenging technical problem but also an emerging need in the imminent new era of decentralized electricity, dominated by renewable energy sources.

The associated research and operational applications so far mostly span over two main directions. The first refers to the short-term energy production forecasting by solar and wind power systems, typically on the basis of Numerical Weather Prediction (NWP) models, providing deterministic point forecasts. The second field of interest deals with the long-term energy production by large hydropower reservoirs, based on projections of their inflows (e.g., Cassagnole et al., 2021). Nowadays, emphasis is given to data-driven approaches (e.g., machine learning), also combined with stochastic-probabilistic schemes for representing uncertainties that are ignored by NWP models (Felder et al., 2018; Talari et al., 2018; Croonenbroeck and Stadtmann, 2019).

Small hydroelectric works are classified as one of the most cost-effective technologies, establishing them as one of the most widespread form of renewable energy. This concerns hydropower systems up to a specific capacity value (e.g., 15 MW in Greece), commonly of negligible storage capacity, where the energy production is a direct conversion of the streamflow arriving at the intake. In contrast to other renewables, the short-term energy forecasting problem, in the field of SHPPs without storage, has not gained the necessary attention from the research community (Yildiz and Açıkgöz, 2021). The typical input information used in fore-

## 1 Introduction

Short-term forecasting of energy production is of high importance for power systems of all scales. This task becomes even more crucial for the renewable sources, which are governed by stochastic drivers, namely weather processes (e.g., wind

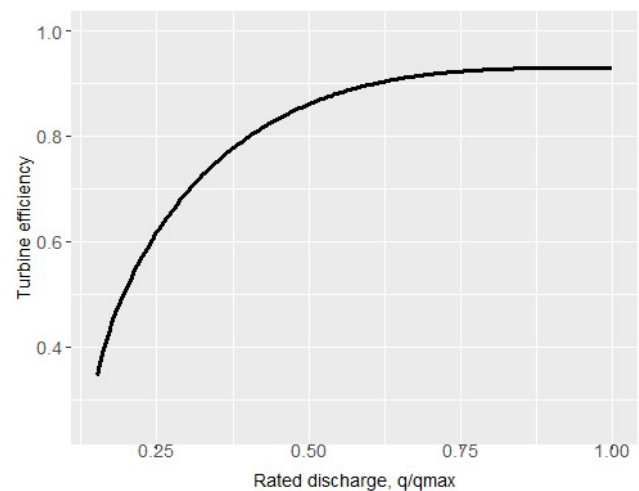
casting schemes appears to be the observed energy production and rainfall (e.g., Li et al., 2015), while some researchers also use forecasted precipitation, provided by NWP models (Monteiro et al., 2013). However, we should highlight that the accuracy of NWPs with respect to rainfall forecasting is still questionable, particularly in complex mountainous reliefs (Ólafsson and Ágústsson, 2021).

Surprisingly, streamflow forecasting procedures, followed by turbine operation models employing flow-energy conversions, seem to be missing. A plausible explanation is the scarcity of streamflow observations, since most of SHPPs are located in small remote catchments, lacking of hydrometric infrastructures. On the other hand, given that the technical and operational characteristics of the SHPP are known (e.g., turbine scheduling and efficiency curves), the past inflows can be retrieved with quite satisfactory accuracy, on the basis of observed power production data, through reverse engineering (Sakki et al., 2022).

Taking as an example a run-off-river SHPP, in the upper course of river Achelous, Western Greece, we investigate different day-ahead power forecasting approaches, driven by alternative data sources. Since the limited scale of the SHPP industry makes difficult to support highly sophisticated operational forecasting systems, we seek for establishing simple and parsimonious regression-type approaches, instead of more complex schemes, e.g., from the domain of machine learning (ML) (cf. Papacharalampous et al., 2019), that yet require significant expertise to be properly used and often demanding computational infrastructures. This fact is probably associated with the growing interest in explainability of such techniques (cf. discussion by Ribeiro et al., 2016). Key objective, and at the same time novelty of this research, is the maximization of information gathered from the available data, by taking advantage of the hydrological expertise and knowledge about the system's properties (i.e., turbine capacity, operational flow range, and efficiency). Our research also highlights the training and evaluation procedure of each forecasting approach, as well as the representation of uncertainty and its practical interpretation. In this vein, our overall objective is to move beyond the standard, yet risky, point forecasting methods, providing a single expected value of hydropower production, thus quantifying the overall predictive uncertainty of each method, and use it as a guidance for modelling energy market behaviors and support decision-making.

## 2 Small hydroelectric plants: An overview

To define a hydroelectric plant as “small”, the installed power capacity of the turbines must be under a certain limit, determined by the national legislation. This limit varies considerably globally, but the most common values are varying from 10 to 30 MW. While the above definition can include hydropower stations as additions to reservoirs, the common type of small hydropower plants (SHPPs) refers to systems



**Figure 1.** Efficiency curve for the Francis turbine applied in the hydropower system under study (analytically-derived relationship provided by Sakki et al., 2021, based on nomographs by Papantonis, 2008, p. 231).

without storage capacity, taking advantage of different combinations of discharge and head. These are actually diversion systems that exploit a specific range of the arriving streamflow, where the maximum flow value depends on the head and installed capacity of turbines, while the minimum one also depends on the turbine type.

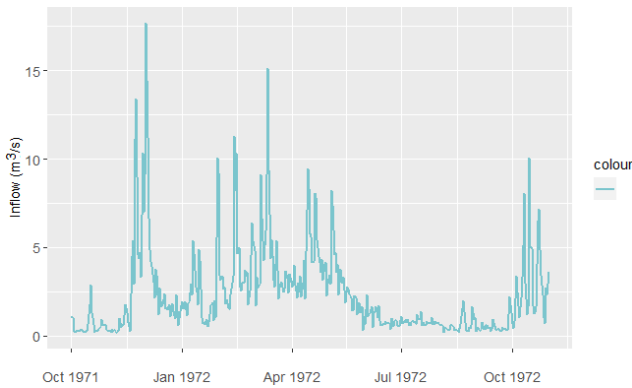
The conversion of the turbine flow to hydropower is made via the well-known formula:

$$P = \gamma \eta(q_T) q_T h_n(q_T) \quad (1)$$

where  $\gamma$  is the specific weight of water ( $9.81 \text{ kN/m}^3$ ),  $q_T$  is the flow passing through the turbines,  $h_n$  is the net head, i.e. the gross head, after subtracting hydraulic losses, which are function of discharge and the penstock characteristics, and  $\eta(q_T)$  is the total efficiency of the system, which is a function of discharge and depends on the turbine type. Figure 1 demonstrates a typical efficiency curve for a Francis-type turbines, which is also applied in our case study. In most cases, a mixing of two turbines of different power capacity is considered, in order to maximize the range of exploitation of the highly varying streamflow. Often, this implies establishing one large and one small turbine, and operate them under a specific hierarchy, to ensure the optimal overall efficiency.

## 3 Study area and data

In the context of our analysis, we consider a run-of-river plant under study, in the upper course of river Achelous, Western Greece. The available hydrological information comprises spatially-averaged daily precipitation data from five representative meteorological stations, and daily streamflow data at the intake. The latter input is extracted by ad-



**Figure 2.** Streamflow time series at the intake for hydrological year 1971–1972.

justing the observed inflows to a downstream site, i.e., Kremasta reservoir (Efstratiadis et al., 2014), by accounting for the ratio of the corresponding drainage areas (about 1 : 40). The common period of the two records extends over 39 years (May 1969 to December 2008). Figure 2 illustrates the adjusted flow time series, the mean annual value of which is  $2.15 \text{ m}^3/\text{s}$ .

Before proceeding with the forecasting problem, it is essential to specify the technical characteristics of the project. First, we estimate the environmental flow to be released downstream of the intake, in order to sustain the riverine ecosystems (Efstratiadis et al., 2014). Following the Greek legislation for SHPPs, this is defined as the 30 % of mean discharge of September, which here equals to  $0.25 \text{ m}^3/\text{s}$ . The flow arriving at the intake is diverted through an open channel to a forebay and next conveyed to the power station through a penstock, thus creating an elevation difference of 150 m. In next computations, we consider this value as a constant net head, for simplicity, taking advantage of the quite large diameter of the penstock (1500 mm), causing minimal only hydraulic losses even at the maximum discharge capacity. We also apply a mixing of two Francis-type turbines, which characteristics are summarized in Table 1. The two capacities have been estimated through optimization, by maximizing the net annual profit of the system, estimated as the difference between the anticipated revenues from energy production and the depreciated costs of the electromechanical equipment (Sakki et al., 2021).

The operation policy of the SHPP is demonstrated in Fig. 3. This has been obtained by seeking for the optimal hierarchy of the two turbines, in order to maximize the power production across different discharge ranges. In particular, from  $0.12$  to  $0.85 \text{ m}^3/\text{s}$ , i.e., the minimum flow of the small and large turbine, respectively, only the small turbine operates, while next, and up to the maximum discharge of the large turbine, i.e.,  $5.69 \text{ m}^3/\text{s}$ , it stops operating, since all diverted flow passes through the large turbine. For larger discharges up to the total system's discharge, the small turbine

**Table 1.** Design characteristics of the two Francis-type turbines.

	Turbine 1	Turbine 2	Total
Power capacity, $P$ (MW)	7.40	1.00	8.40
Maximum discharge, $q_{\max}$ ( $\text{m}^3/\text{s}$ )	5.69	0.77	6.46
Minimum discharge, $q_{\min}$ ( $\text{m}^3/\text{s}$ )	0.85	0.12	0.12

operates in its full capacity, which ensures a maximized overall efficiency. Under these assumptions, the project is expected to produce 47.5 MWh, on mean daily basis, thus ensuring a capacity factor of about 24 %.

## 4 Application of alternative forecasting schemes

### 4.1 Generic framework

The consecutive conversions across SHPPs allows for establishing two alternative routes to the power forecasting problem, here employed on a day-ahead basis. The direct route aims at predicting the next-day energy production via regression models that use as explanatory variables past observations, in terms of power production and the past rainfall, as the sole source of hydrological data. On the other hand, the indirect route initially aims at predicting the day-ahead discharge, given that such data exist. The forecasted flows are next introduced to the operation model of the system, for extracting the forecasted energy. For each approach, we assess alternative forecasting schemes, in terms of model structure and data. In order to calibrate the free parameters of each model and evaluate their predictive capacity, we introduce a quite strict skill score in terms of the generic efficiency formula:

$$F = 1 - \frac{\sum_{t=1}^n (E_{t,\text{obs}} - E_{t,\text{forecast}})^2}{\sum_{t=1}^n (E_{t,\text{obs}} - E_{t,\text{benchmark}})^2} \quad (2)$$

where  $E_{t,\text{obs}}$  is the “observed” energy at day  $t$ , which is known from the simulation model,  $E_{t,\text{forecast}}$  is the forecasted value, which is estimated on the basis of past data ( $x_{t-1}, x_{t-2}, \dots$ ), and  $E_{t,\text{benchmark}}$  is a reference prediction, provided by a benchmark model. In the classical definition of efficiency, this coincides with the mean observation (thus the daily average energy production), yet here we also apply a stricter benchmark prediction, i.e., the so-called naïve forecasting model  $E_t = E_{t-1}$  (hereafter referred to as modified efficiency). The aforementioned expression ensures an efficiency up to 75.1 %, for the entire period of historical data (1969–2008; see Table 2). For each model, we also compute the marginal statistical characteristics of residuals,  $w_t = E_{t,\text{obs}} - E_{t,\text{forecast}}$  (mean, standard deviation, coefficient of skewness) and the lag-1 autocorrelation, which is measure of dependence.

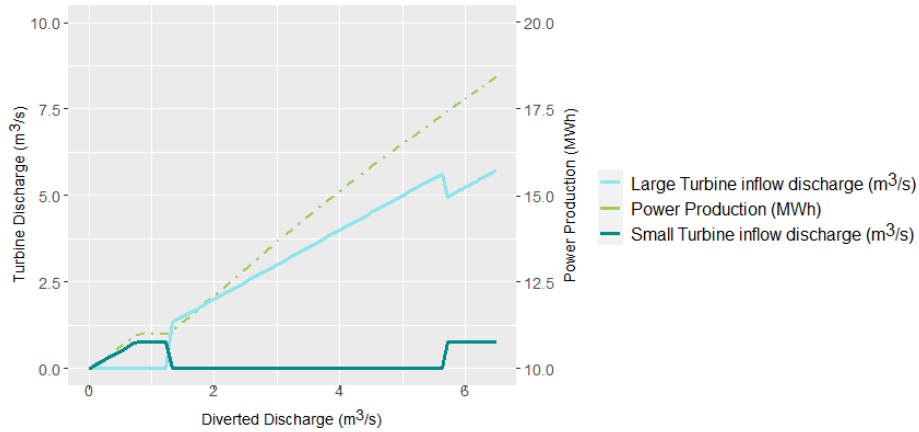


Figure 3. Optimal sharing of discharge conveyed to the two turbines and associated power production.

### 4.2 Direct (energy-based) approaches

In the direct approaches we use as independent variables (predictors) for the energy generated at time step (day)  $t + 1$ , the past energy production,  $E$ , as well as the available hydrological data, by means of rainfall,  $p$ , from a representative meteorological station or the spatially-aggregated rainfall from a set of stations. After investigations, we concluded to a branched expression, that takes into consideration the appearance of a rainfall event at day  $t$ , which is expected to influence the generation of streamflow due to the increase of soil moisture over the basin:

$$E_{t+1} = \begin{cases} 3.88(E_t)^{0.54}(E_{t-1})^{0.12}(p_t)^{0.16}, & p_t > 0.1 \text{ mm} \\ 1.51(E_t)^{0.63}(E_{t-1})^{0.25}, & p_t \leq 0.1 \text{ mm} \end{cases} \quad (3)$$

Following the typical split-sample approach, we estimated the seven model parameters by calibrating against the actual energy production values in the half of observations, and validating its predictive capacity in the other half. The skill scores in calibration and validation, expressed in terms of classical and modified efficiency, as well as the statistical characteristics of the model error are summarized in Table 2.

### 4.3 Indirect (flow-based) approaches

In the indirect approaches we aim to provide day-ahead forecasts of the discharge to feed the flow-energy conversion model, which is summarized in the diagram of Fig. 3. In this respect we use as predictors of streamflow at day  $t + 1$  the past streamflow and the rainfall. In order to account for the baseflow component of streamflow, we also extract the minimum value of last five days,  $q_{\min 5}$ , and the mean monthly value of the full data sample; these predictors were determined by following an input variable selection procedure (Galelli and Castelletti, 2013). We remark that the baseflow component may incorporate several slow-flow elements, associated with groundwater runoff, snow melting (which is quite significant, during the spring period) as well as the falling limb of floods.

After investigations, we conclude to the parametric expression:

$$q_{t+1} = \begin{cases} a_1(q_{\min 5}) + \beta_1(q_t) + \gamma_1(q_{\text{mean}t}), & p_t < 0.1 \text{ mm} \\ a_2(q_{\min 5}) + \beta_2(q_t) + \gamma_2(q_{\text{mean}t}) + \delta(p_t), & p_t \geq 0.1 \text{ mm} \end{cases} \quad (4)$$

As shown in Table 2, by employing a typical calibration on the basis of maximizing the efficiency of the simulated against the observed streamflows, in terms of day-ahead energy prediction we obtain a small only improvement with respect to the direct modelling approach, namely from 79.9 % to 80.7 % (for the full data). On the other hand, the model error characteristics are less satisfactory, since the forecasting model underestimates the energy production by about  $-3.8$  MWh, on average, while with the direct approach the bias is negligible. This is due to the attempt of the calibration procedure to predict streamflows outside of the operational range of turbines, and particularly the peak flows, which result to large errors. Yet, these errors are beyond our concerns, since during these periods the system operates continuously in its nominal capacity.

In order to remedy the above shortcomings, we adjust the fitting metric, i.e., efficiency, to the turbine operation range ( $q_{\min, \text{tot}}, q_{\max, \text{tot}}$ ), in order to ignore the errors that are produced from the flow forecasting model, if this correctly predicts that the flows being outside this range. Moreover, if the forecasted flow is inside the range, whereas the observed is outside, we only account from the distance from the two flow limits. Under this premise, the error is calculated as follows:

$$e = \begin{cases} q_{\max, \text{tot}} - q_{\text{forecast}}, & q_{\text{obs}} > q_{\max, \text{tot}} \text{ and } q_{\text{forecast}} < q_{\max, \text{tot}} \\ q_{\text{forecast}} - q_{\min, \text{tot}}, & q_{\text{obs}} < q_{\min, \text{tot}} \text{ and } q_{\text{forecast}} > q_{\min, \text{tot}} \\ q_{\text{obs}} - q_{\text{forecast}}, & q_{\text{obs}} > q_{\min, \text{tot}} \text{ and } q_{\text{obs}} < q_{\max, \text{tot}} \end{cases} \quad (5)$$

The above error expression incorporates within calibration, apart from the hydrological data, the expert’s knowledge about the technical properties of the system that affect the flow-energy conversions. The knowledge-based calibration approach, herein called Indirect Model B, ensures a clearly better skill score in terms of modified efficiency than the typical calibration approach (Indirect Model A), and good er-

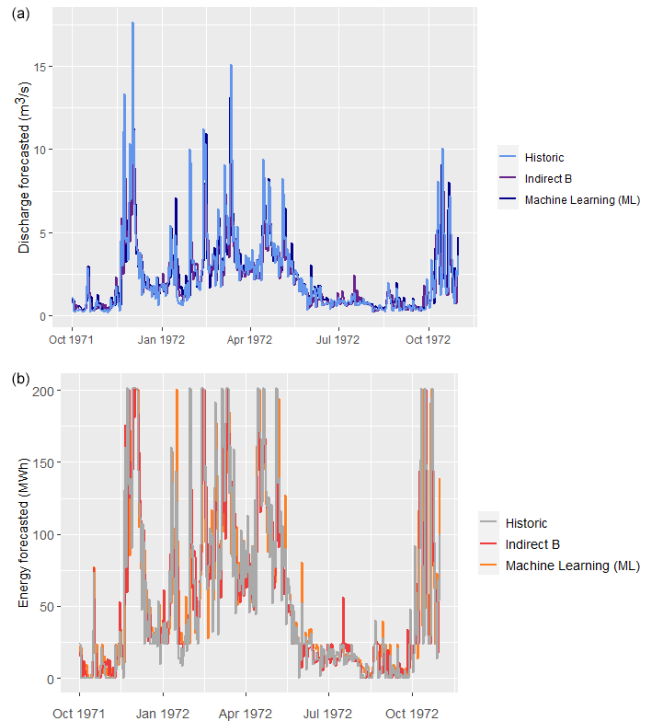
ror properties as well (practically zero mean and autocorrelation).

An interesting question that arises is whether just a better day-ahead flow forecasting model that does not account for the operational characteristics of the system, would outperform the optimal model so far. In this respect, we apply a more complex approach from the Machine Learning (ML) family, namely a Deep Feedforward Neural Network (DNN). The DNN model is composed by three hidden layers with 128, 64 and 64 neurons, respectively, while the Rectified Linear Unit (ReLU) activation function is adopted for all neurons. As inputs, we use the streamflow of past 5 d and the rainfall of past two days. The model is fitted on the basis of Mean Square Error (MSE), for a number of 100 epochs, by using a batch size of 64.

In Fig. 4 we compare the actual and forecasted flow and energy values provided by the Indirect Model B and the ML approach, for hydrological year 1971–1972. Surprisingly, while the ML model ensures a much better fitting to the observed flows than the simple regression expression (4), (84 % vs. 63 %), the conversion to energy is rather disappointing. In particular, the classical efficiency metric is only 50.7 %, while the modified efficiency is strongly negative. Furthermore, the derived error properties are clearly non satisfactory (underestimation of the average energy up to 1 MW, quite large standard deviation, and, significant autocorrelation). The poor predictive capacity of the data-driven approach is attributed to the training procedure, in which we ignored the range of operation of the small hydroelectric plant, which is key feature of its management.

## 5 Forecasting and uncertainty: reconciliation in practice

Predictive uncertainty is defined as the probability of occurrence of a predictand's value (in the particular case, energy production) conditional upon prior observations and knowledge, as well as on all the information we have obtained on that specific value from model forecasts (Coccia and Todini, 2011). A typical means to quantify the predictive uncertainty of a deterministic simulation model, is to add a random component (noise),  $w_t$ , to its output,  $y_t$ , where the random process  $w_t$  should be consistent with the statistical and stochastic regime of the associated residuals (Efstratiadis et al., 2015). In the generic case of autocorrelated errors, the process  $w_t$  can be obtained by a stochastic generator (e.g., Kossieris et al., 2019; Tsoukalas et al., 2018, 2020), or a statistical distribution model, provided that the errors do not exhibit significant dependencies in space and time. By generating a large enough set of random variables  $y_{i,t} = y_t + w_{i,t}$ , where  $i = 1, \dots, n$ , we obtain an ensemble of  $n$  model realizations at each time step  $t$ , which allows for the detection of empirically-derived probabilistic quantities. This procedure is widely known as Monte Carlo simulation.



**Figure 4.** Comparison of historical vs. predicted time series obtained by two forecasting models (indirect B and ML) for (a) in-flows and (b) energy production (hydrological year 1972–1973).

In our case, we use the more robust forecasting scheme (Indirect Model B) and provide  $n = 100$  realizations of the day-ahead energy at each time step (day), from which we get the median and the 10th and 90th largest values (quantiles) of forecasted energy production, as estimators of the 80 % empirical confidence intervals. Given that the observed residuals are uncorrelated, the errors are considered as a stationary process that follows a three-parameter gamma distribution (i.e., Pearson type III), which reproduces the mean value,  $\mu_e$  the standard deviation,  $\sigma_e$ , and the coefficient of skewness,  $\gamma_e$ , of the entire sample of residuals (Table 2). We also implement the same analysis by applying seasonally-varying (cyclostationary) generation models, which account for the individual statistical characteristics per month (Tsoukalas et al., 2018). As shown in Table 3, there is a considerable difference of the statistical behavior of the error across different seasons, that is also reflected in the uncertainty of energy predictions. In Fig. 5 we compare the uncertainty bounds obtained from the two methods, for hydrological year 1971–1972. We observe that during the low-flow period, these bounds are substantially reduced, by accounting for the issue of seasonality. This indicates that the more detailed analysis, where the prediction error is represented as a cyclostationary process, is much more realistic.

In order to take advantage of the concept of uncertainty in practice, as would be made in a real-world energy market, we can determine alternative market policies in terms

**Table 2.** Comparison of different forecasting schemes.

Model type	Efficiency			Modified efficiency			Error statistics			
	Calibr.	Valid.	Full data	Calibr.	Valid.	Full data	Mean (MWh)	SD (MWh)	Coeff. of skewness	Lag-1 auto-correlation
Naïve	n/a	n/a	0.751	n/a	n/a	n/a	0.00	27.76	1.21	−0.119
Direct	0.806	0.790	0.799	0.181	0.203	0.194	−0.48	24.92	1.47	0.059
Indirect A	0.817	0.796	0.807	0.228	0.226	0.227	−3.83	24.11	−0.37	0.079
Indirect B	0.848	0.819	0.833	0.356	0.314	0.331	0.16	22.70	1.25	0.068
DNN	n/a	n/a	0.507	n/a	n/a	−0.977	−0.99	39.04	−0.42	0.540

n/a: not applicable

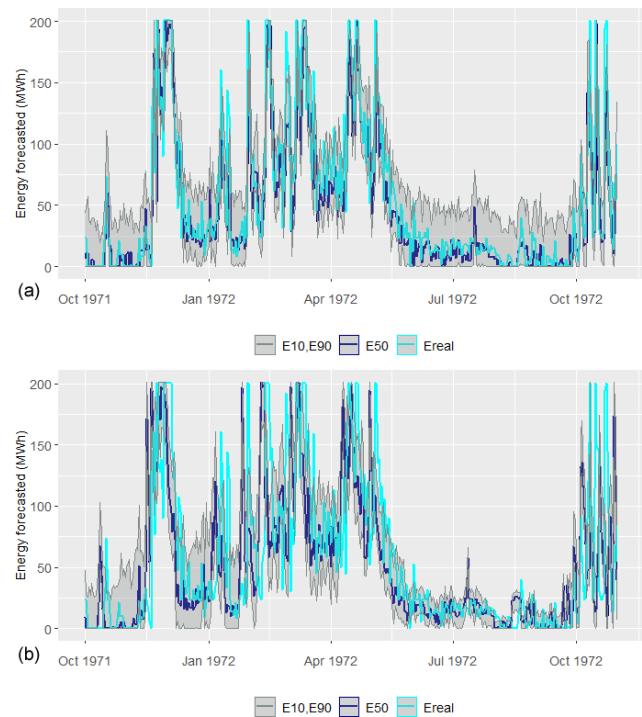
**Table 3.** Monthly statistical characteristics of residuals derived from the application of Indirect Model B.

Month	Mean (MWh)	Standard deviation (MWh)	Skewness	Lag-1 correlation
JAN	0.69	28.57	1.23	0.069
FEB	1.48	28.98	0.87	0.138
MAR	4.79	27.99	0.95	0.066
APR	1.87	26.48	0.48	0.059
MAY	1.45	19.06	1.37	0.089
JUN	−0.29	8.42	0.99	−0.062
JUL	0.12	7.72	−0.28	−0.014
AUG	−0.69	7.11	−1.89	0.056
SEP	−1.90	11.06	−1.40	0.044
OCT	−2.56	20.71	1.75	−0.050
NOV	−0.57	31.43	1.33	0.043
DEC	−2.92	30.53	0.84	0.045

of quantiles. In particular, we can apply the upper, middle and low quantiles as representatives of a risky, mild and conservative forecast of the day-ahead energy, and evaluate them in economic terms, by assigning a unit profit value for delivering the energy produced up to the forecasted value, and a unit penalty for the deviations (i.e., deficits with respect to the forecasted value). For instance, we account for the 90 %, 50 % and 10 % quantiles and apply a fixed profit of EUR 60/MWh and a penalty value of EUR 50/MWh; the aforementioned values are representative of the recent system marginal price and price of deviations, respectively, of the Hellenic Electricity Market. Under this premise, the mild policy ensures a mean annual profit of EUR 0.86 million, the conservative EUR 0.81 million, and the risky EUR 0.43 million. This quick pseudo-financial analysis allows for comparing the different interpretations of a forecasting approach under uncertainty.

## 6 Conclusions

This research aims to revisit the problem of day-ahead power forecasting in the case of small hydropower plants without

**Figure 5.** Comparison of actual energy production for hydrological year 1972–1973 with three characteristic prediction quantiles (10 %, 50 % and 90 %), by considering the error process as: (a) stationary and (b) cyclostationary.

storage capacity, which has received little attention so far. Taking as an example a typical project of this category, and by using simple yet effective modeling schemes, we attempt to revisit several issues that may have been well-addressed in the generic context of hydrological forecasting, but not in the specific case of SHPPs, namely: (a) the essential information as input to hydropower forecasting; (b) the advantages of the indirect forecasting approach, involving the use of a stream-flow forecasting model, against the direct one, that does not account for the inflow input, but relies solely on the energy production data; (c) the importance of past precipitation data as exogenous predictor, providing macroscopic information

about the catchment state (e.g. antecedent soil moisture conditions); (d) the training procedure and the skill score to be applied; and (e) the representation of the predictive uncertainty around the point forecast of day-ahead energy; (f) and the use of uncertainty-aware forecasts from the practitioners' point-of-view (investors, power engineers, stakeholders).

Our investigations indicated that the proposed flow-based approach is more flexible and physically consistent, since it provides forecasts of the hydropower system's driver, i.e., the inflow arriving at the intake. We also revealed that apart from the inflow data per se, additional information should be introduced within prediction schemes in order to better reflect our hydrological knowledge, in terms of statistical characteristics. In the particular example, these were the mean monthly inflows and the past five-day average values, as representative of the long and short-term regime of the upstream catchment, respectively. However, it is worth mentioning that even a very good prediction of inflows (as quantified in terms of efficiency), does not guarantee an equally good performance in energy prediction (see Fig. 4). Equivalently important is the training procedure and the associated performance measure, where the system's characteristics, i.e., the range of operation of turbines, are embedded as inputs to calibration.

Key outcome of this research was also the quantification of uncertainty, by means of empirical quantiles, which were estimated through a Monte Carlo approach, after fitting a suitable probability distribution to the model residuals. This task, although proved to be simple and effective in its implementation, requires more careful examination, including analysis of the error properties and their seasonal variability, as well as could be benefited from more advanced concepts and tools, such as copulas and conditional non-Gaussian distributions (cf. Tsoukalas, 2018, for a development of this kind). Nevertheless, the interpretation of uncertainty is essential as a guidance for modelling energy market behaviors and providing decision support in the Target model era.

*Code availability.* The simulation and forecasting models have been developed in the R environment and they are available at <https://doi.org/10.5281/zenodo.5843449> (Drakaki, 2022).

*Data availability.* The daily inflow and rainfall time series are available at <https://doi.org/10.5281/zenodo.5841828> (Efstratiadis, 2022), and they are delivered under the Creative Commons Attribution 4.0 International. The time series have been created by Andreas Efstratiadis, on the basis of raw hydrometeorological data retrieved by the National Data Bank of Hydrological and Meteorological Information of Greece (<http://www.hydroscope.gr>, last access: 12 January 2022).

*Author contributions.* The methodology was developed in close collaboration of all authors, and all authors contributed to the editing of the paper (mainly KKD, GKS, and AE). KKD employed the

literature review, the forecasting analyses, and the development of the source code. GKS prepared the essential data for the case study (design of SHPP, simulation model). IT and PK employed the generation of synthetic data. PK also conducted the analysis with the Deep Feedforward Neural Network. AE supervised the overall research and the preparation of the paper.

*Competing interests.* The contact author has declared that neither they nor their co-authors have any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Special issue statement.* This article is part of the special issue "European Geosciences Union General Assembly 2021, EGU Division Energy, Resources & Environment (ERE)". It is a result of the EGU General Assembly 2021, 19–30 April 2021.

*Acknowledgements.* We are grateful to the Topical Editor, Gregor Giebel, who coordinated the review procedure, and the two anonymous reviewers for their useful comments, which helped substantially in improving the presentation of our work.

*Review statement.* This paper was edited by Gregor Giebel and reviewed by two anonymous referees.

## References

- Cassagnole, M., Ramos, M.-H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., and Ouillon, T.: Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs – a conceptual approach, *Hydrol. Earth Syst. Sci.*, 25, 1033–1052, <https://doi.org/10.5194/hess-25-1033-2021>, 2021.
- Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, *Hydrol. Earth Syst. Sci.*, 15, 3253–3274, <https://doi.org/10.5194/hess-15-3253-2011>, 2011.
- Croonenbroeck, C. and Stadtmann, G.: Renewable generation forecast studies – Review and good practice guidance, *Renew. Sust. Energy. Rev.*, 108, 312–322, <https://doi.org/10.1016/j.rser.2019.03.029>, 2019.
- Drakaki, K. K.: corinadrakaki/Day-ahead-energy-production-in-small-hydropower-plants: (v1.0.1), Zenodo [code], <https://doi.org/10.5281/zenodo.5843449>, 2022.
- Efstratiadis, A.: Daily inflow and rainfall data, Zenodo [data set], <https://doi.org/10.5281/zenodo.5841828>, 2022.
- Efstratiadis, A., Tegos, A., Varveris, A., and Koutsoyiannis, D.: Assessment of environmental flows under limited data availability – Case study of the Acheloos River, Greece, *Hydrol. Sci. J.*, 59, 731–750, <https://doi.org/10.1080/02626667.2013.804625>, 2014.

- Efstratiadis, A., Nalbantis, I., and Koutsoyiannis, D.: Hydrological modelling of temporally-varying catchments: Facets of change and the value of information, *Hydrol. Sci. J.*, 60, 1438–1461, <https://doi.org/10.1080/02626667.2014.982123>, 2015.
- Felder, M., Sehnke, F., Ohnmeiß, K., Schröder, L., Junk, C., and Kaifel, A.: Probabilistic short term wind power forecasts using deep neural networks with discrete target classes, *Adv. Geosci.*, 45, 13–17, <https://doi.org/10.5194/adgeo-45-13-2018>, 2018.
- Galelli, S. and Castelletti, A.: Tree-based iterative input variable selection for hydrological modelling, *Water Resour. Res.*, 49, 4295–4310, <https://doi.org/10.1002/wrcr.20339>, 2013.
- Kossieris, P., Tsoukalas, I., Makropoulos, C., and Savic, D.: Simulating marginal and dependence behaviour of water demand processes at any fine time scale, *Water*, 11, 885, <https://doi.org/10.3390/w11050885>, 2019.
- Li, G., Li, B.-J., Yu, X.-G., and Cheng, C.-T.: Echo state network with Bayesian regularization for forecasting short-term power production of small hydropower plants, *Energies*, 8, 12228–12241, <https://doi.org/10.3390/en81012228>, 2015.
- Monteiro, C., Ramirez-Rosado, I. J., and Fernandez-Jimenez, L. A.: Short-term forecasting model for electric power production of small-hydro power plants, *Renew. Energ.*, 50, 387–394, <https://doi.org/10.1016/j.renene.2012.06.061>, 2013.
- Ólafsson, H. and Ágústsson, H.: Mesoscale orographic flows, in: Ólafsson, H., and Bao, J.-W., *Uncertainties in Numerical Weather Prediction*, Chapter 11, Elsevier, 297–308, <https://doi.org/10.1016/B978-0-12-815491-5.00011-2>, 2021.
- Papacharalampous, G., Tyralis, H., Langousis, A., Jayawardena, A. W., Sivakumar, B., Mamassis, N., Montanari, A., and Koutsoyiannis, D.: Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms, *Water*, 11, 2126, <https://doi.org/10.3390/w11102126>, 2019.
- Papantonis, D. E.: *Small Hydroelectric Works*, 2nd edition, Symeon Editions, Athens, 2008 (in Greek).
- Ribeiro, M. T., Singh, S., and Guestin, C.: Why should I trust you? Explaining the predictions of any classifier, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144, <https://doi.org/10.1145/2939672.2939778>, 2016.
- Sakki, G.-K., Tsoukalas, I., Kossieris, P., and Efstratiadis, A.: A dilemma of small hydropower plants: Design with uncertainty or uncertainty within design?, *EGU General Assembly 2021*, online, 19–30 Apr 2021, EGU21-2398, <https://doi.org/10.5194/egusphere-egu21-2398>, 2021.
- Sakki, G.-K., Tsoukalas, I., and Efstratiadis, A.: A reverse engineering approach across small hydropower plants: a hidden treasure of hydrological data?, *Hydrol. Sci. J.*, accepted, 67, <https://doi.org/10.1080/02626667.2021.2000992>, 2022.
- Talari, S., Shafie-Khah, M., Osório, G. J., Aghaei, J., and Catalão, J. P. S.: Stochastic modelling of renewable energy sources from operators' point-of-view: A survey, *Renew. Sust. Energ. Rev.*, 81, 1953–1965, <https://doi.org/10.1016/j.rser.2017.06.006>, 2018.
- Tsoukalas, I.: Modelling and simulation of non-Gaussian stochastic processes for optimization of water-systems under uncertainty, PhD thesis, Dept. of Civil Engineering, National Technical University of Athens, available at: <https://www.itia.ntua.gr/1933/> (last access: 6 January 2022), 2018.
- Tsoukalas, I., Efstratiadis, A., and Makropoulos, C.: Stochastic periodic autoregressive to anything (SPARTA): Modelling and simulation of cyclostationary processes with arbitrary marginal distributions, *Water Resour. Res.*, 54, WRCR23047, <https://doi.org/10.1002/2017WR021394>, 2018.
- Tsoukalas, I., Kossieris, P., and Makropoulos, C.: Simulation of non-Gaussian correlated random variables, stochastic processes and random fields: Introducing the anySim R-Package for environmental applications and beyond, *Water*, 12, 1645, <https://doi.org/10.3390/w12061645>, 2020.
- Yildiz, C. and Açikgöz, H.: Forecasting diversion type hydropower plant generations using an artificial bee colony based extreme learning machine method, *Energ. Source. Part B*, 16, 216–234, <https://doi.org/10.1080/15567249.2021.1872119>, 2021.