Current Sustainable/Renewable Energy Reports manuscript No.
(will be inserted by the editor)

Data-Driven Modeling of Operating Characteristics of Hydroelectric Generating Units

Rachel Hunter-Rinderle · **Ramteen Sioshansi**

Received: 6 July, 2021 / Accepted: 19 August, 2021

Abstract
Purpose of Review Hydroelectric generation is a potential flexible electricity source that can ease the transition to a decarbonized energy economy. As such, using scarce hydroelectric generating resources efficiently is important. We examine approaches to represent the operating characteristics of hydroelectric resources.
Recent Findings Many hydroelectric-plant owners use water tables or generic unit characteristics for operational planning. Such practice may be inefficient, as it does not account for unit-specific operating-characteristic changes or time-related impacts, *e.g.*, plant degradation. We demonstrate a data-driven approach to modeling plant operations that is unit-specific and depends solely on observable and controllable variables.
Summary Numerical results using historical data for four hydroelectric units illustrate the proposed methodology.

Keywords Hydroelectric generators · dispatching · head level · wicket gate

1 Introduction

Given its operational flexibility, hydroelectric generation may have an outsize role to play in decarbonizing electricity systems. Thus, there is increased onus to use scarce water resources as efficiently as possible, which depends on a number of important factors. First, is having forecasts of water-availability conditions. Second, is having

R. Hunter-Rinderle · R. Sioshansi
Department of Integrated Systems Engineering, The Ohio State University, Baker Systems Engineering Building, 1971 Neil Avenue, Columbus, OH 43210, USA
E-mail: hunter-rinderle.2@buckeyemail.osu.edu
E-mail: sioshansi.1@osu.edu (Corresponding author)

R. Sioshansi
Department of Electrical and Computer Engineering, The Ohio State University, Dreese Laboratories, 2015 Neil Avenue, Columbus, OH 43210, USA

optimization or market models that can trade-off between current and future water use. A third important factor is having an accurate representation of the relationship between water flow through and power output from a hydroelectric unit.

This paper surveys approaches to operational modeling of hydroelectric generators, focusing on the aforementioned challenges in efficient water-resource use. We demonstrate, using historical data from four hydroelectric generators, a data-driven approach to capturing the relationship between observable and controllable variables and plant operations. Such an approach to modeling plant operations can yield more efficient use of hydroelectric plants.

The remainder of this paper is organized as follows. First, we provide our literature survey. Next, we outline our proposed approach to modeling hydroelectric-unit operations, data that are used to calibrate and validate the model, and model results. This is followed by concluding remarks.

2 Literature Survey

Using hydroelectric generation efficiently is of particular importance given the competing (*e.g.*, agricultural, habitat-maintenance, or direct-consumption) uses of water, which is scarce in many regions [1, 2]. One factor to efficient use of hydroelectricity is having forecasts of water-availability conditions, which can help to avoid undesirable water spillage. García-Morales and Dubus [3] propose the use of ensemble forecasts to develop more skillful seasonal precipitation forecasts. Boucher and Ramos [4] survey the application of ensemble techniques to forecasting streamflows for hydroelectric generators.

A second important factor in using hydroelectricity efficiently is having optimization models that can trade-off between current and future uses. Such models can be used directly by the owner of a hydroelectric plant, *e.g.*, if its manages the electricity supply chain as a vertically integrated firm, or by a third party, *e.g.*, a market operator. There are numerous works developing such optimization techniques, with much of the focus being on reservoir as opposed to run-of-river hydroelectric plants [5]. Gfrerer [6] models hydroelectric-energy production using optimal-control models, which are transformed into equivalent nonlinear optimization problems. Bauer *et al*. [7] expand this work by using dynamic and nonlinear optimization to solve numerical case studies. Catalão *et al*. [8] develop a deterministic optimization model that focuses on non-linearities that are caused by head-sensitive output. Pousinho *et al*. [9] employ a stepped market-price function to model a hydroelectric unit that accounts for the impacts of its production decisions on market prices. Philpott *et al*. [10] develop a cost-minimization model that accounts for unit-commitment decisions and random demand. They linearize the non-linearities that are caused by head-dependent generation, which yields a mixed-integer linear optimization that they solve efficiently using a rounding heuristic. Pérez-Díaz *et al*. [11] survey operational and optimization challenges that are raised by pumped hydroelectric energy storage plants. Seguin *et al*. [12] focus on representing uncertainty in optimizing hydroelectric operations. They employ two-stage stochastic optimization, with scenarios representing uncertainty (water inflows in their case). Marchand *et al*. [13] develop a two-stage stochastic approach to optimizing hydroelectric operations that employs tabu search as an efficient solution algorithm. Taktak and D'Ambrosio [14] focus on capturing hydroelectric units in unit-commitment models, which could be used to co-optimize the operation of hydroelectric units with other power-system resources.

An added complication that arises in optimizing some hydroelectric resources is having multiple units in a single catchment area, which can couple the operation of the individual units. Piekutowski *et al*. [15] develop a model, which includes coupling constraints across the catchment, to optimize the use of water resources across a catchment area. Blom *et al*. [16] propose the use a linearized equivalent model for different parts of a coupled hydroelectric system. They demonstrate that this approach reduces computation time significantly (*e.g*., from nine hours to less than a minute in their examples) with some loss of model fidelity (9%–15% average relative errors in modeled power production). Other works expand upon these by developing efficient modeling of connections between elements in a catchment area [17], which may be owned by independent agents [18], and non-linearities that are created by water-spillage conditions [19].

A third important factor in the use of hydroelectricity is having an accurate representation of the relationship between water flow and power output. Oftentimes, hydroelectric operators model this relationship using specifications that are provided by the equipment manufacturer. Diniz *et al*. [20] employ a function, which is fit to the performance hill chart that is provided by an equipment manufacturer, that relates the efficiency factor for a hydroelectric unit to its head level and water flow. Thus, their methodology interpolates performance data that the manufacturer provides. An issue with relying on manufacturer specifications it that it does not account for degradation, which occurs naturally over time, component replacements or upgrades, or other time-related changes in the operational characteristics of a unit. Hidalgo *et al*. [21] address this shortcoming by using measured historical power-output, head-level, and water-flow data to estimate unit efficiency. Kong *et al*. [22] propose the use of piecewise-linear approximations of the power-consumption function of a variable-speed pump. They demonstrate the use of this technique to capture head variations that stem from reservoir water levels and head loss due to penstock friction.

3 Proposed Data-Driven Approach to Operational Modeling

The literature on power-output modeling of hydroelectric units has a gap, in that it does not relate control variables (*e.g*., wicket-gate position) to water flow and power output. Rather, existing techniques assume that water flow is known in estimating unit efficiency and power output [20, 21]. Thus, these models may rely on generic relationships between wicket-gate position, water flow, and power output that are not unit-specific and do not capture time effects on generator efficiency.

We demonstrate, using historical hydroelectric-operations data, a two-step data-driven approach to modeling unit efficiency. First, we fit a model that relates water flow to head level and wicket-gate position. Then, we fit a second model that relates power output to water flow and head level. The latter model uses a standard production function [23]. We know of no such standard function in the literature for the former relationship. Thus, we develop a model structure that is based on known physical relationships, which allows us to employ parametric regression. Combining the two proposed models, one can relate power output directly to wicket-gate position and head level, which are controllable and observable variables, respectively.

3.1 Water-Flow Model

Flow-rate equations define water flow, q, which is given in $\mathrm{m}^3 \cdot \mathrm{s}^{-1}$, as the following product of water velocity, v, which is given in $\mathrm{m} \cdot \mathrm{s}^{-1}$, and the effective cross-sectional area of the pipe through which the water flows, A_e, which is given in m^2:

$$q = vA_e. \tag{1}$$

From Bernoulli's principle [24], water velocity is the following function of head level, h, which is given in m, and the gravitational constant, g, which is given in $\mathrm{m} \cdot \mathrm{s}^{-2}$:

$$v = \sqrt{2gh}. \tag{2}$$

A_e is defined as the following product of the actual cross-sectional area of the pipe, A_a, which is given in m^2, and the per-unit wicket-gate position, w:

$$A_e = wA_a. \tag{3}$$

Combining (1)–(3), and assuming a per-unit water-flow-efficiency factor, σ, gives:

$$q = wA_a\sigma\sqrt{2gh}, \tag{4}$$

as a model that relates water flow to head level and wicket-gate position.

One limitation in (4) is that it does not capture degradation or other time-related impacts. Thus, we use the following model:

$$q_t = b_0 + b_1 w_t \sqrt{2gh_t} + b_2 t + \varepsilon_t^q, \tag{5}$$

to represent the relationship between water flow, head level, and wicket-gate position, where b_0, b_1, and b_2 are parameters. Water flow, wicket-gate position, and head level are indexed by time in (5). That is, q_t, w_t, and h_t are the time-t water flow, wicket-gate position, and net-head level, respectively, and retain the same units as q, w, and h, respectively. The time index, t, reflects our goal of estimating b_0, b_1, and b_2 by fitting (5) to historical time-series data using ordinary-least-squares (OLS) estimation [25]. ε_t^q is the error term.

3.2 Power-Output Model

Breton *et al*. [23] propose the physical model:

$$p = \eta g h q,$$

to relate power output, p, which is given in MW, to head level, water flow, and a per-unit generator-efficiency factor, η. As with (4), this relationship does not capture time-related impacts. Thus, we use the following model:

$$p_t = c_0 + c_1 g h_t q_t + c_2 t + \varepsilon_t^p, \tag{6}$$

to represent the relationship between power output, head level, and water flow, where c_0, c_1, and c_2 are parameters and ε_t^p is the time-t error term. p_t is time-t power output (given in MW) in the historical data that are used for OLS estimation of c_0, c_1, and c_2.

We use historical water-flow data for fitting (5) and (6). However, once the values of b_0, b_1, b_2, c_0, c_1, and c_2 are estimated, power output can be modeled in decision-support and market models by using the fitted form of (5) to estimate water flow based on head level and wicket-gate position and inputting those estimates into (6) to estimate power output. In doing so, power output can modeled using observable and controllable variables only.

4 Case Study

4.1 Case-Study Data

We apply our proposed model to historical data for four hydroelectric units, concentrating this discussion on one unit, as the results and model performance are similar for the other units. All data processing, model fitting, and out-of-sample validation are conducted using `R` version 3.5.1.

Our case study uses hourly historical measurements of head level, power output, wicket-gate position, and water flow for the 15-year period spanning 2004–2019. Some portion of the measurements for the period spanning 2004–2006 are inaccurate, due to known instrument failures. These observations are removed from our case-study data. Measurements when a unit is non-operational and has zero power output are removed as well. These observations are removed because the aim of our modeling approach is to estimate water flow and power output when a unit is operational (water flow and power output can be estimated trivially as zero when a plant is non-operational). Finally, we remove observations with two other types of obvious measurement errors. The first are measurements with negative power output. The second are measurements for which the head level is negative or greater than 45.72 m in magnitude, which are outliers based on our analysis of the data. Figure 1 summarizes the time series of the head level, without observations that are removed due to known instrument failure or negative power output having been recorded. Figure 1 shows that there are five outliers that reflect measurement errors.

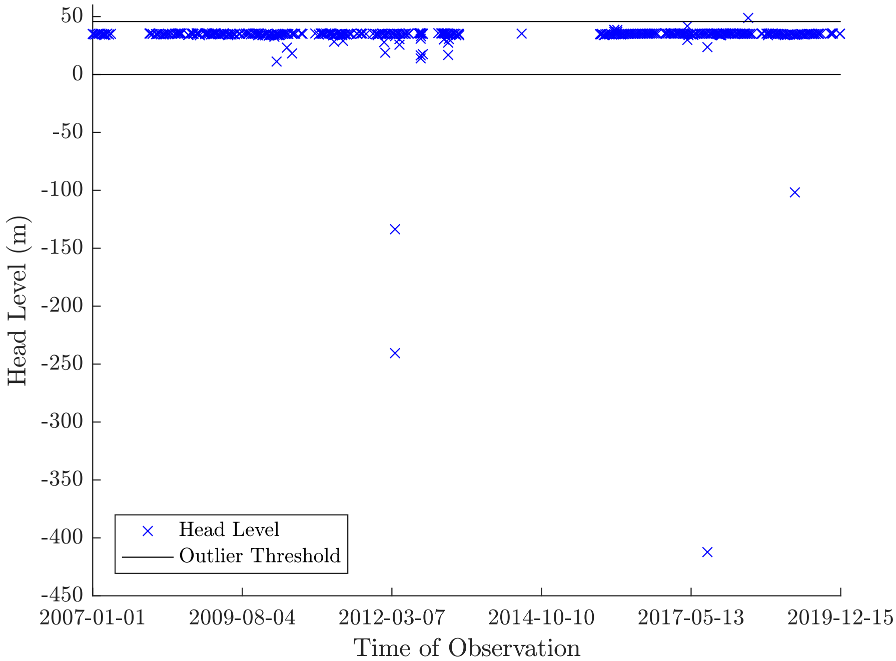

Fig. 1 Time series of head level

In total, we retain roughly 40 000 hourly observations for each unit once these outliers and erroneous observations are removed. To allow for out-of-sample validation, 80% of the retained observations are selected randomly for fitting models (5) and (6). The other 20% of the retained observations are used for model validation.

4.2 Case-Study Results

4.2.1 Water-Flow Model

Table 1 summarizes the OLS estimates of the coefficients, b_0, b_1, and b_2, in (5). As indicated in the table, all of estimates are statistically significantly non-zero at the 1% confidence level. b_2 having a negative value indicates that there is slight (but statistically significant) degradation in water flow over time (*e.g.*, due to sedimentation in the penstock). The model R^2 is 0.9138, meaning that 91.38% of variation in water flow can be described by the model. Figure 2 shows the residuals (errors) of the OLS estimate of (5), which have relatively constant variance and no noticeable autocorrelation. Thus, OLS estimates should be unbiased and efficient [25].

Figure 3 provides a scatter plot that demonstrates model performance. Specifically, it shows modeled water flow using the OLS estimate of (5) and head-level and wicket-gate-position observations for the 20% of observations that are retained for out-of-sample validation. These modeled water flows are plotted against mea-

Table 1 OLS estimates of coefficients in (5)†

Coefficient	Value
b_0	19.513^{**}
b_1	$2.440 \times 10^{-2**}$
b_2	$-8.379 \times 10^{-9**}$

†: Values with a single asterisk are statistically significantly non-zero at a 5% confidence level. Values with a double asterisk are statistically significantly non-zero at a 1% confidence level.

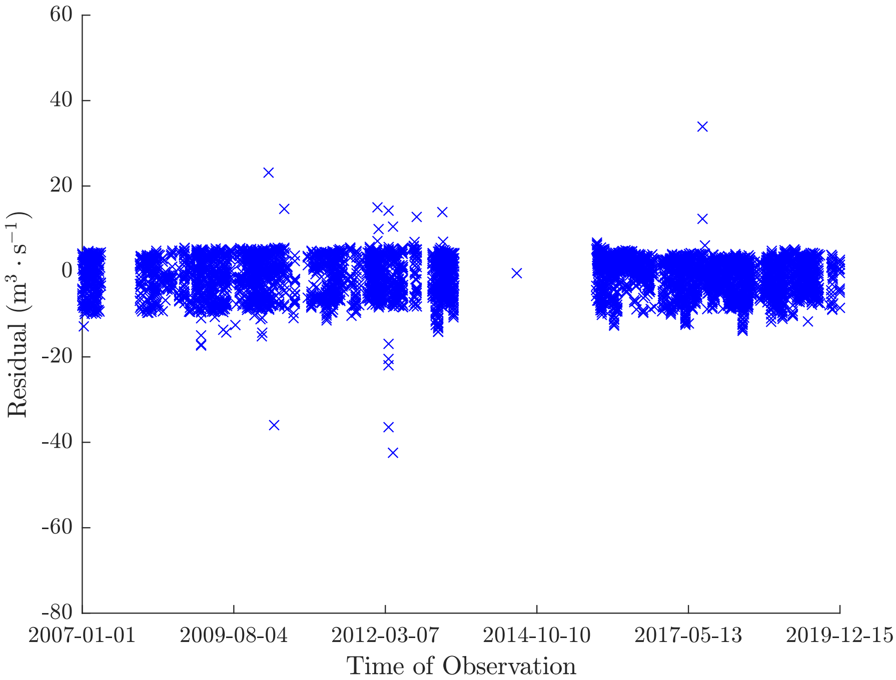

Fig. 2 Time series of residuals in OLS estimate of (5)

sured water flows for those 20% of observations. The figure shows that the model has significant predictive power, being able to model water flows well. The root-mean-square error (RMSE) for the 20% of observations that are retained for out-of-sample validation is 144.9 $m^3 \cdot s^{-1}$ as opposed to an RMSE of 142.9 $m^3 \cdot s^{-1}$ for the OLS estimate of the model. Moreover, the R^2 for the observations that are retained for out-of-sample validation is 0.9124. These results show that (5) has similar predictive capability when applied to the observations that are retained for out-of-sample validation as it does for the 80% of observations that are used for model fitting.

4.2.2 Power-Output Model

Table 2 summarizes the OLS estimates of the coefficients, c_0, c_1, and c_2, in (6). The coefficients, c_0 and c_1, are statistically significantly non-zero, whereas c_2 is not. This

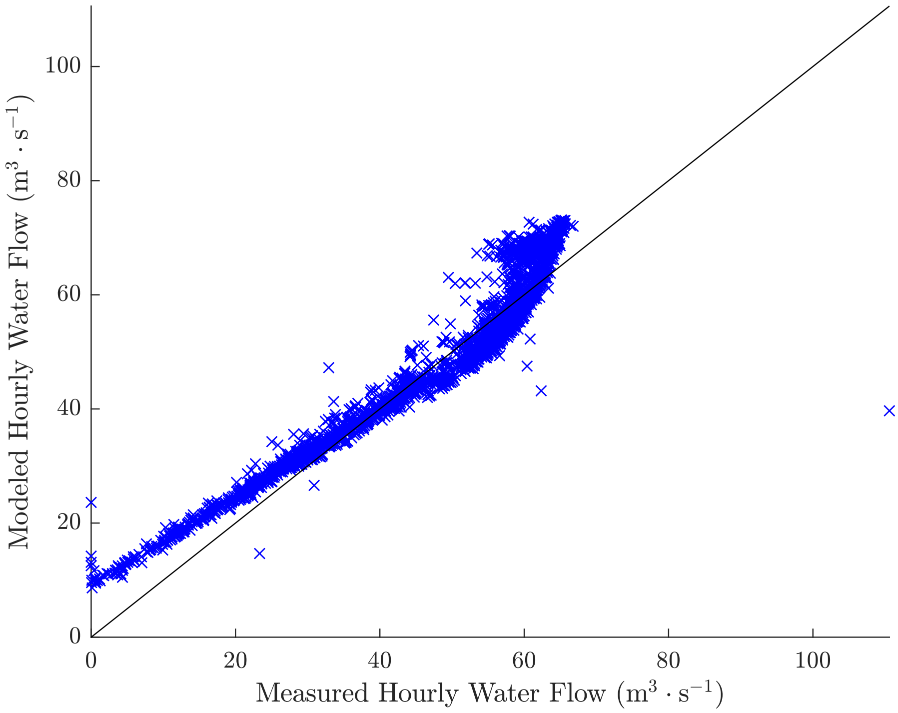

Fig. 3 Scatter plot of modeled water flow using (5) and actual water flow using 20% of historical data

Table 2 OLS estimates of coefficients in (6)†

Coefficient	Value
c_0	$-5.695 \times 10^{-1**}$
c_1	$7.904 \times 10^{-6**}$
c_2	1.018×10^{-10}

†: Values with a single asterisk are statistically significantly non-zero at a 5% confidence level. Values with a double asterisk are statistically significantly non-zero at a 1% confidence level.

indicates that there is a slight but statistically insignificant improvement in the efficiency of the turbine over time. The model R^2 is 0.8901. Figure 4 shows the residuals of the OLS estimate of (6), which have relatively constant variance and no noticeable autocorrelation. Thus, OLS estimates should be unbiased and efficient.

Figure 5 provides a scatter plot that demonstrates the performance of our proposed two-step power-output-modeling technique. Modeled power output is computed by using (5) to estimate water flow (based on wicket-gate position and head level), which is input to (6) with head level to estimate power output. The scatter plot considers the 20% of observations that are retained for out-of-sample validation. Figure 5 shows that the two-step modeling technique has good predictive power, being able to model power output. The RMSE for the 20% of observations that are retained for out-of-sample validation is 1.543 MW as opposed to an RMSE of 1.435 MW

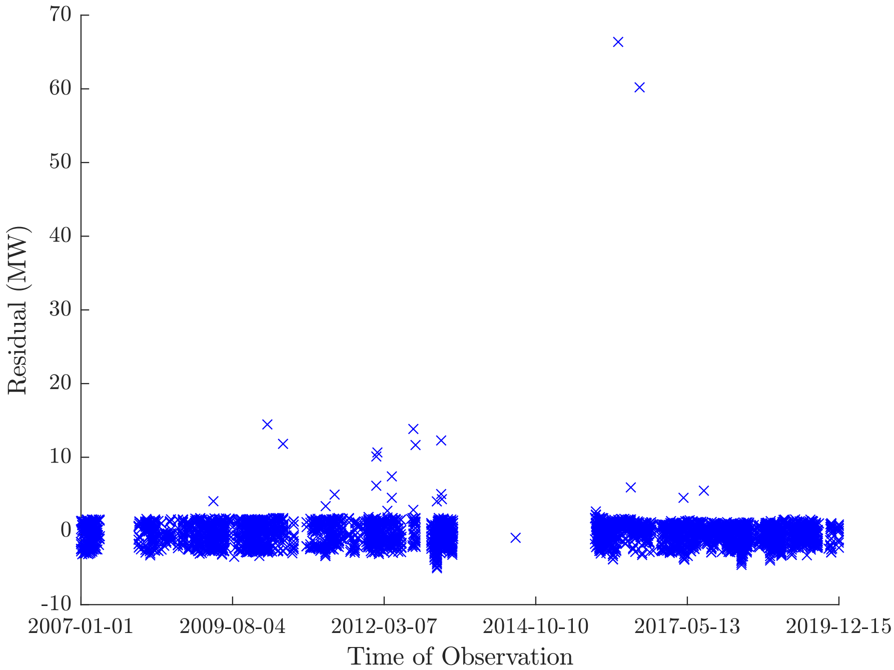

Fig. 4 Time series of residuals in OLS estimate of (6)

Table 3 Out-of-sample R^2 of OLS estimates of (5) and (6) for other units

Unit	Model (5)	Model(6)
1	0.9375	0.9327
2	0.9305	0.9258
3	0.9874	0.7704

for the OLS estimate of (6). The R^2 for the observations that are retained for out-of-sample validation is 0.8754. These results show that our model has similar predictive power when applied to the observations that are retained for out-of-sample validation as it does for the observation that are used for model fitting. This is despite our two-step-modeling approach introducing error in both water-flow and power-output estimation.

4.2.3 Results for Other Generating Units

Table 3 summarizes the goodness of fit of the proposed models for the other three units. Specifically, the table reports R^2 of the estimated models when they are applied to the 20% of observations that are retained for out-of-sample validation. The table shows that water-flow model (5) has very similar goodness of fit across all four units. Power-output model (6) has performance that is more mixed for these three units relative to the one on which we focus above.

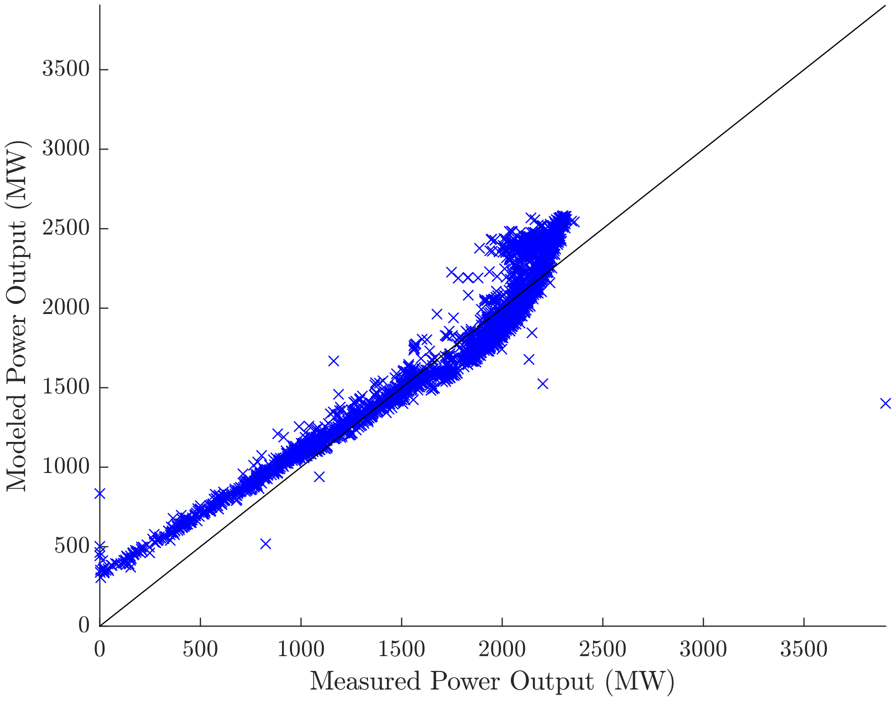

Fig. 5 Scatter plot of modeled power output using (6), with (5) used to model water flow, and actual power output using 20% of historical data

5 Discussion and Conclusions

Efficient use of hydroelectric resources is becoming more important as the pressure to decarbonize electricity systems increases. We survey the literature related to water-resource forecasting, optimizing the operation of hydroelectric plants, and estimating the physical operating characteristics of hydroelectric units. With respect to the third focus area, we observe a gap in models capturing changes in the operating characteristics of plants *vis-à-vis* performance hill charts and related characteristics that a plant manufacturer may provide.

We demonstrate a potential two-step data-driven methodology to model power output and efficiency of a hydroelectric unit with three key characteristics. First, the model uses directly observable and controllable variables (*e.g.*, head level and wicket-gate position) as inputs only. Second, the model can capture time-related changes in operating characteristics. Third, unit-specific models can be estimated, which would capture unit-specific time-related changes in operational characteristics.

Performance of our proposed methodology could be improved by considering other dynamics. For instance, we do not allow for unit efficiency to vary with operating characteristics [20]. Adding such considerations to the model may give more predictive power, at the cost of making the model more complex or nonlinear. The benefit of our assumed model structures is that when (5) is substituted into (6), we obtain the power-output relationship:

$$p_t = c_0 + c_1 g h_t \cdot \left(b_0 + b_1 w_t \sqrt{2 g h_t} + b_2 t \right) + c_2 t,$$

which is linear in the control variable, w_t. Having a linear relationship is beneficial in operational and market modeling of a hydroelectric unit [10].

Acknowledgements This work was supported by National Science Foundation grant 1808169. The authors thank the editors, reviewer, and Armin Sorooshian (University of Arizona) for helpful comments and discussions and staff of American Electric Power Company, Inc. for providing historical hydroelectric-unit operational data. Any opinions and conclusions that are expressed in this paper are solely those of the authors.

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Andy Burrow, Alexandra M. Newman, and Morgan Bazilian. Reservoir Design and Operation for the Food-Energy-Water Nexus. *Current Sustainable/Renewable Energy Reports*, 6:71–89, September 2019.
2. Charalampos Avraam, Ying Zhang, Sriram Sankaranarayanan, Benjamin Zaitchik, Emma Moynihan, Prathibha Juturu, Roni Neff, and Sauleh Siddiqui. Optimization-Based Systems Modeling for The Food-Energy-Water Nexus. *Current Sustainable/Renewable Energy Reports*, 8:4–16, March 2021.
3. Marta Benito García-Morales and Laurent Dubus. Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts. *International Journal of Climatology*, 27:1691–1705, October 2007.
4. Marie-Amélie Boucher and Maria-Helena Ramos. Ensemble Streamflow Forecasts for Hydropower Systems. In Qingyun Duan, Florian Pappenberger, Andy Wood, Hannah L. Cloke, and John C. Schaake, editors, *Handbook of Hydrometeorological Ensemble Forecasting*, pages 1289–1306. Springer, Berlin, Germany, 5 January 2019.
5. Vineet Kumar Singh and Sunil Kumar Singal. Operation of hydro power plants-a review. *Renewable and Sustainable Energy Reviews*, 69:610–619, March 2017.
6. Helmut Gfrerer. Optimization of Hydro Energy Storage Plant Problems by Variational Methods. *Zeitschrift für Operations Research*, 28:B87–B101, June 1984.

7. Wolfgang Bauer, Helmut Gfrerer, and Hansjörg Wacker. Optimization Strategies for Hydro Energy Storage Plants. *Zeitschrift für Operations Research*, 28:B103–B131, June 1984.
8. João P. S. Catalão, Sílvio J. P. S. Mariano, Victor M. F. Mendes, and Luís A. F. M. Ferreira. Scheduling of Head-Sensitive Cascaded Hydro Systems: A Nonlinear Approach. *IEEE Transactions on Power Systems*, 24:337–346, February 2009.
9. Hugo M. I. Pousinho, Javier Contreras, and João P. S. Catalão. Short-term optimal scheduling of a price-maker hydro producer in a pool-based day-ahead market. *IET Generation, Transmission & Distribution*, 6:1243–1251, December 2012.
10. Andrew B. Philpott, Mark Craddock, and Hamish Waterer. Hydro-electric unit commitment subject to uncertain demand. *European Journal of Operational Research*, 125:410–424, September 2000.
11. Juan I. Pérez-Díaz, Manuel Chazarra, Javier García-González, Giovanna Cavazzini, and Anna Stoppato. Trends and challenges in the operation of pumped-storage hydropower plants. *Renewable and Sustainable Energy Reviews*, 44: 767–784, April 2015.
12. Sara Séguin, Stein-Erik Fleten, Pascal Côté, Alois Pichler, and Charles Audet. Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259:1156–1168, 16 June 2017.
13. Alexia Marchand, Michel Gendreau, Marko Blais, and Jonathan Guidi. Optimized operating rules for short-term hydropower planning in a stochastic environment. *Computational Management Science*, 16:501–519, July 2019.
14. Raouia Taktak and Claudia D'Ambrosio. An overview on mathematical programming approaches for the deterministic unit commitment problem in hydro valleys. *Energy Systems*, 8:57–79, February 2017.
15. Marian R. Piekutowski, Tudeusz Litwinowicz, and Roderick J. Frowd. Optimal Short-Term Scheduling for a Large-scale Cascaded Hydro System. *IEEE Transactions on Power Systems*, 9:805–811, May 1994.
16. Evelin Blom, Lennart Söder, and Daniel Risberg. Performance of multi-scenario equivalent hydropower models. *Electric Power Systems Research*, 187:106486, October 2020.
17. Ziad K. Shawwash, Thomas K. Siu, and Samuel O. Denis Russell. The B.C. Hydro Short Term Hydro Scheduling Optimization Model. *IEEE Transactions on Power Systems*, 15:1125–1131, August 2000.
18. Ivan Rajšl, Perica Ilak, Marko Delimar, and Slavko Krajcar. Dispatch Method for Independently Owned Hydropower Plants in the Same River Flow. *Energies*, 5:3674–3690, September 2012.
19. Yoshiro Ikura and George Gross. Efficient Large-Scale Hydro System Scheduling with Forced Spill Conditions. *IEEE Transaction on Power Apparatus and Systems*, PAS-103:3502–3520, December 1984.
20. André Luiz Diniz, Pedro Paulo Ielo Esteves, and Claudia Alejandra Sagastizábal. A Mathematical Model for the Efficiency Curves of Hydroelectric units. In *2007 IEEE Power Engineering Society General Meeting*, Tampa, Florida, 24-28 June 2007. Institute of Electrical and Electronics Engineers.

21. Ieda G. Hidalgo, Darrell G. Fontane, João E. G. Lopes, José G. P. Andrade, and André F. de Angelis. Efficiency Curves for Hydroelectric Generating Units. *Journal of Water Resources Planning and Management*, 140:86–91, January 2014.
22. Jiehong Kong, Hans Ivar Skjelbred, and Hubert Abgottspon. Short-term hydro scheduling of a variable speed pumped storage hydropower plant considering head loss in a shared penstock. *IOP Conference Series: Earth and Environmental Science*, 240:082002, March 2019.
23. Michèle Breton, Saeb Hachem, and Abdelghani Hammadia. Accounting for Losses in the Optimization of Production of Hydroplants. *IEEE Transactions on Energy Conversion*, 19:346–351, June 2004.
24. Victor L. Streeter and Evan Benjamin Wylie. *Fluid Mechanics*. The McGraw-Hill Companies, Inc., New York, New York, seventh edition, 1979.
25. Paul Arthur Ruud. *An Introduction to Classical Econometric Theory*. Oxford University Press, New York, New York, 2000.

Bulleted Annotated References

Very Important References

- Hidalgo *et al*. [21]. Uses historical measured power, head level, and water flow to fit efficiency curves for hydroelectric generating units to determine performance characteristics.

Important References

- Taktak and D'Ambrosio [14]. Considers various assumptions that are used for simplifications of unit-commitment models with hydroelectric generators, including linearizations of penstock loss.