Systems & Control Transactions

Research Article
Foundations of Computer Aided Process Design (FOCAPD 2024)
Breckenridge, Colorado, USA. July 14-18, 2024
Peer Reviewed Conference Proceeding

PSE PRESS

Conceptual Design of Integrated Energy Systems with Market Interaction Surrogate Models

Xinhe Chen[a], Radhakrishna Tumbalam-Gooty[b,c], Darice Guittet[d], Bernard Knueven[d], John D. Siirola[e], and Alexander W. Dowling[a]*

[a] University of Notre Dame, Department of Chemical and Biomolecular Engineering, South Bend, IN 46556, United States
[b] National Energy Technology Laboratory (NETL), Pittsburgh, PA 15236, United States
[c] NETL Support Contractor, Pittsburgh, PA 15236, United States
[d] National Renewable Energy Laboratory, Golden, CO 80401, United States
[e] Sandia National Laboratories, Albuquerque, NM 87185, United States
* Corresponding Author: adowling@nd.edu

ABSTRACT

Most integrated energy system (IES) optimization frameworks employ the price-taker approximation, which ignores important interactions with the market and can result in overestimated economic values. In this work, we propose a machine learning surrogate-assisted optimization framework to quantify IES/market interactions and thus go beyond price-taker. We use time series clustering to generate representative IES operation profiles for the optimization problem and use machine learning surrogate models to predict the IES/market interaction. We quantify the accuracy of the time series clustering and surrogate models in a case study to optimally retrofit a nuclear power plant with a polymer electrolyte membrane electrolyzer to co-produce electricity and hydrogen.

Keywords: Integrated Energy System, Surrogate Models, Machine Learning, Optimization, Time Series Clustering

INTRODUCTION

Integrated energy systems (IES) exploit synergies between different technologies and energy carriers [1] such as fossil, nuclear, renewable (e.g., solar, wind), and storage to produce multiple products such as electricity, heat, and chemicals while increasing energy efficiency and supporting greater renewable integration into the grid. Moreover, IES can provide more flexibility to the grid, which is critical for increasing the integration of non-dispatchable renewable energy sources and meeting decarbonization goals.

The price-taker approximation is a widely used approach to incorporate dynamic market signals into the optimization of IES. The price-taker approximation treats the electricity grid as an “infinite bus” that can receive any amount of electricity produced by each generator or IES at any time without affecting the location marginal price (LMP) of electricity. Price-taker allows for co-optimizing IES design and operating decisions using historical or forecasted time-series LMPs. Lakey et al. [2] proposed a framework that allows users to generate a multi-period, price-taker model instance and automatically generate common operational constraints for their model in the IDAES-PSE platform [3].

Recent papers [2-5] highlight the inaccuracies of the price-taker approximation compared to more rigorous production cost models (PCM) for analyzing IES. Specifically, Martinek et al. [4] compare the revenue, dispatch, and scheduling of concentrating solar power plants using the PCM and price-taker models. They show the price-taker approximation can over-aggressively respond to short-duration peaks in LMP. Likewise, Frew et al. [5] show the price-taker approximation can overestimate the value of IES that co-produces electricity and hydrogen. Sousa et al. [6] evaluate the integration of wind energy and pump storage systems and find that in the PCM, storage operation will decrease the electricity price and dispatch less wind power than the price-taker. Finally, Xian et al. [7] propose a multiscale simulation framework that combines rigorous process (IES) and grid (PCM) models to quantify IES-grid interactions across hourly to annual timescales. They find that optimizing the design, operation, and control of a single IES impacts the prices and dispatch of generators across the electric grid.

We recently proposed a machine learning surrogate-assisted optimization framework [8] to integrate the process-centric model (price-taker) and grid-centric model (PCM) into the IES conceptual design optimization. Specifically, we contemplate how to size a replacement generator while considering market impacts. We compare algebraic and neural network surrogate models to predict market revenue, annual capacity factor distribution, and the number of shutdowns as a function of the characteristics of the replacement generator. We embed these surrogates into a nonlinear optimization problem and show more accurate results versus the price-taker approximation compared to PCM simulations (ground truth). While promising, this approach only considers steady-state models and surrogates.

In this short conference paper, we extend our prior work [8] by using time series clustering to generate time series representative dispatch scenarios. Then, we train a frequency surrogate model that predicts the frequency of each scenario according to the different IES designs. Together, these surrogate models provide dynamic operation profiles that may be embedded in multiscale IES optimization problems. We develop the surrogate models and clustering methods using a baseload nuclear power plant (NPP) with a polymer electrolyte membrane (PEM) electrolyzer that produces hydrogen (H_2) as a case study.

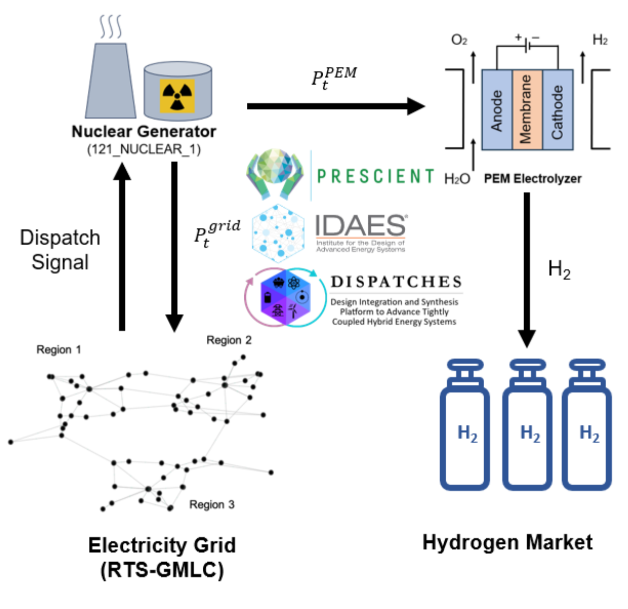

Figure 1. The nuclear + PEM IES can divert the power to co-produce hydrogen when the LMP is low to increase its profitability and flexibility.

METHODS

Nuclear and H_2 IES

We consider retrofitting an existing nuclear power plant with a polymer electrolyte membrane (PEM) electrolyzer to co-produce electricity and hydrogen, as shown in Figure 1. The NPP has a nameplate capacity of 400MW and operates strictly as a baseload generator. After retrofitting, the IES (i.e., NPP + PEM) operates as a hybrid baseload and peaking generator when power from the NPP can be diverted to the PEM to produce hydrogen. Thus, from the perspective of the electric grid, the PEM provides flexibility. Hydrogen is an important resource for the future energy industry and is essential to the decarbonization of many industries [9]. The retrofit thus increases and diversifies the NPP's revenue.

Production Cost Models

A PCM simulates the operation of an electricity system by scheduling generators and clearing the market (i.e., setting time-varying prices). In summary, each generator communicates its (time-varying) production costs and operational constraints (e.g., minimum up and down times, startup costs) to the market. This includes a bid curve (see Figure 2 for an example), which communicates the cost of producing electricity as a function of the market price. Using this information, the market operator solves unit commitment and economic dispatch optimization problems, which seek to minimize the total system-wide generation costs subject to the costs and constraints of each generator and forecasted demand and renewables production. Most regions schedule most electricity generation in the day-ahead market (DAM). The real-time market (RTM) operates (sub)hourly to correct for forecasting errors and unplanned events. The PCM simulates the rolling horizon operation of the DAM and RTM. See [10] for details.

In this work, all PCM simulations were performed using the open-source production cost model Prescient [11] and the RTS-GMLC [12] dataset. RTS-GMLC is an open-source test network that approximates the characteristics of the southwest United States [7]. We consider retrofitting the 121_NUCLAER-1 nuclear generator in RTS-GMLC region 1 with a PEM.

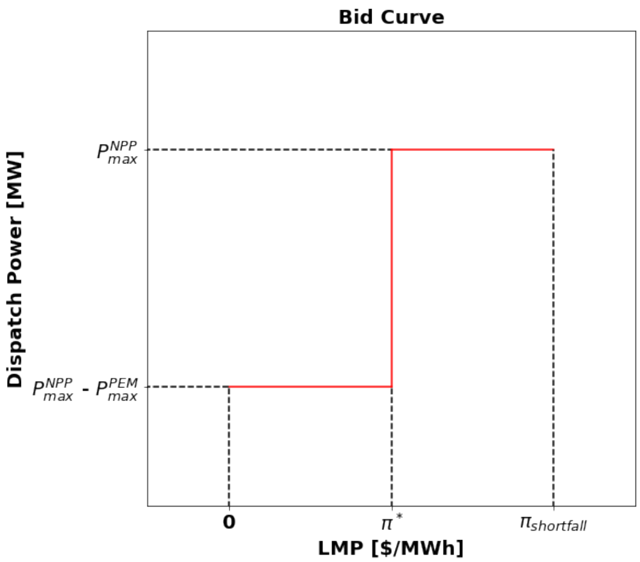

Figure 2. Bid curve of the NPP-PEM IES. The IES will dispatch all power to the grid when the electricity price exceeds π^*. Otherwise, the IES will divert $P_{max}^{NPP} - P_{max}^{PEM}$ power to co-produce H_2.

Surrogate-Assisted Optimization Workflow

Figure 3 illustrates the four-step surrogate-assisted optimization workflow.

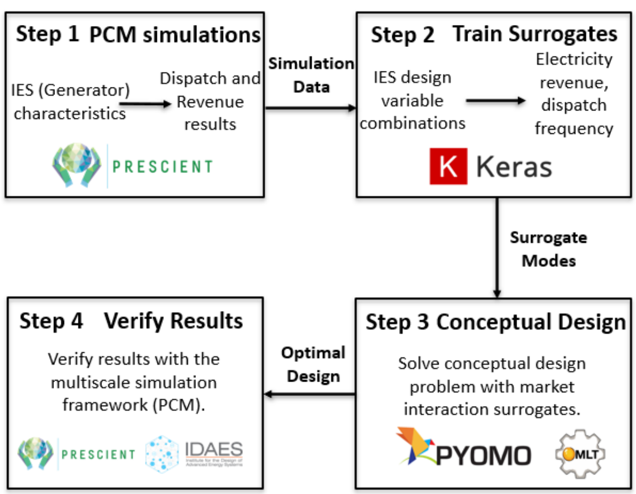

Figure 3. Surrogate-assisted IES conceptual design workflow, adapted from [6].

Step 1: Perform PCM simulations to generate training data for surrogate models. We consider two IES design variables: (1) the ratio of maximum PEM power (P_{max}^{PEM}) to the maximum power plant power (P_{max}^{NPP}), and (2) the threshold LMP price (π^*) that the IES diverts from the electricity production mode to the hydrogen co-production mode. In PCM simulations, we use different bidding curves [13] (production-cost pairs, see Figure 2) to communicate the IES's flexibility and economics to the market. We discretize the continuous IES design space and simulate all combinations in the PCM. We performed 192 annual IES PCM simulations in the NPP-PEM IES case study. PCM simulation results give detailed hourly LMP and dispatch profiles within the simulation horizon (366 days).

Step 2: Train IES-market surrogate models. Using the LMP and dispatch profile, we calculate the total electricity revenue of each IES design combination. Then, we train a neural network revenue surrogate model (2 hidden layers, 25 nodes each layer and activated by hyperbolic tangent) to predict the annual electricity revenue according to the IES design. We subdivide the annual hourly time series IES dispatch profile into 366 daily dispatch profiles. Next, we use time series clustering [14] to identify 24-hour representative dispatch profiles for the IES conceptual design optimization problem. A neural network dispatch frequency surrogate model (3 hidden layers, 75 nodes each layer and activated by sigmoid) is trained to predict the occurrence probability of each representative day. Inputs of surrogate models are two design variables mentioned in Step 1, plus PCM simulation reserve factor and load shed price. Both surrogate models are trained by Keras [15], and we use 80% of the data for training and 20% for validation.

Step 3: Solve conceptual design optimization of IES with surrogate models. We then formulate and solve a stochastic optimization problem, Equations (1) to (5), with IES-market surrogate models to obtain the optimal IES design and operation decisions:

$$\max \varphi[R(x) + \textstyle\sum_s \omega_s(x) \sum_t R_{H2}(x, u_{s,t}, \delta_{s,t}) - C(x, u_{s,t}, \delta_{s,t}))] - CAPEX(x) \quad (1)$$

$$\text{s.t.} \quad h(x, u_{s,t}, \delta_{s,t}) \leq 0, s \in S, t \in T \quad (2)$$

$$g(x, u_{s,t}, \delta_{s,t}) = 0, s \in S, t \in T \quad (3)$$

$$R(x) = f_{rev}(x) \quad (4)$$

$$\omega_s(x) = f_{dis}(x), s \in S \quad (5)$$

The objective function is the 30-year NPV value of the IES retrofit. The two-stage stochastic optimization problem aims to optimize the IES design and operation together under the operation uncertainty. Table 1 summarizes the nomenclature.

For brevity, we do not report optimization results for (1) – (5) in this short conference paper. We recommend using OMLT [16] to embed neural network surrogate models in Pyomo [17]. Because the neural network uses nonlinear activation functions, we recommend using IPOPT [18] with HSL linear algebra [19] distributed as part of the IDAES platform [3].

Step 4: Verify the optimal IES design with PCM. Perform PCM simulation of the optimal design IES and compare the result obtained from Step 3. If the verification result varies dramatically from the optimization results, the training data can be expanded and Steps 2 to 4 repeated.

This conference paper focuses on the time-series clustering in Steps 1 and 2.

Table 1. Conceptual design model nomenclature.

Symbol	Meaning
s, S	Scenario (Set)
t, T	Time step (Set)
$R()$	Revenue surrogate model
$\omega_s()$	Frequency surrogate model
$C()$	Operation cost
$CAPEX()$	Capital cost
$h()$	Inequality constraints
$g()$	Equality constraints
x	Surrogate input variables
u	Operation variables
δ	Representative dispatch profiles
φ	NPV multiplier

Time Series Clustering of PCM Data

Time series clustering is employed to generate representative IES operation profiles. In Step 1, we perform PCM simulation of different IES designs and obtain an annual IES operation profile. We slice the dispatch profile into 24-hr daily time series and use the K-means clustering algorithm to minimize the within-cluster sum of squares Euclidean distance to get the representative days.

In the NPP-PEM model, due to the different PEM max capacity, the minimum power of the IES will be different. To increase the clustering accuracy and avoid the potential infeasibility in the conceptual design model, we scale the capacity factor of NE/PEM IES between 0 and 1, given in Equation (6). The 0 corresponds to the minimum power output to the grid, i.e., when the PEM is operating at the maximum power.

$$cf_t^{NPP} = \frac{P_t^{NPP} - P_{min}^{NPP}}{P_{max}^{NPP} - P_{min}^{NPP}} \quad (6)$$

Before applying K-means clustering, we filter the time series with capacity factors equal to 0 (minimum dispatch at $P_{max}^{NPP} - P_{max}^{PEM}$) or 1 (max dispatch at P_{max}^{NPP}) at every time step. These time series are placed into two clusters labeled "min" and "max" in Figure 4. We find this filtering improves the accuracy of K-means clustering and reduces the size of the clustering dataset by 43.5%.

RESULTS AND DISCUSSION

Time Series Clustering Centers Reliably Represent the Entire Dispatch Dataset

In Figure 4, we use a box plot to sort the daily capacity factor of all clusters from "min" to "max" calculated via:

$$cf_{daily}^{NPP} = \frac{\sum_{t=1}^{T} cf_t^{NPP}}{T} \quad (7)$$

Figure 4 shows that cluster centers accurately represent the data in the cluster, and each cluster center has a substantial amount of data. As we can see there, except for the "min" and "max," all other clusters hold the time

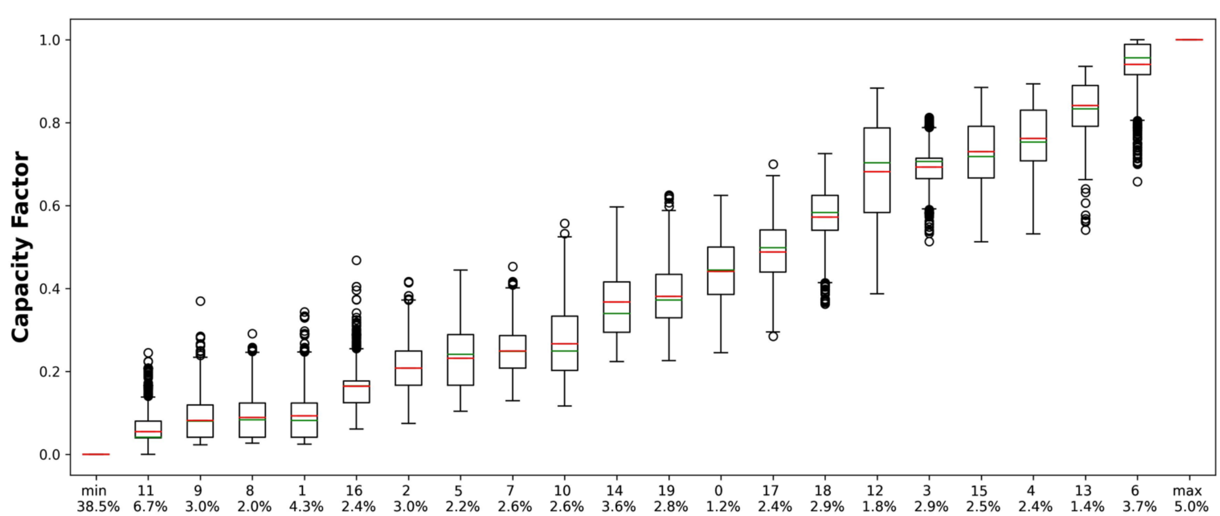

Figure 4. Daily capacity factors of clustering centers. Green lines are median capacity factors, and red lines are representative capacity factors. The clusters are sorted by the capacity factor from low to high. In the horizontal, we show the index of each cluster and the amount of data that falls in the cluster.

series from 1.2% to 6.7%, and there is no single cluster with a significant large or small amount of data. None of the clusters have a significant number of outliers.

As shown in Figure 2, the NPP-PEM IES offers power between $P_{max}^{NPP} - P_{max}^{PEM}$ (min) and P_{max}^{NPP} (max). Moreover, because the PEM model is linear (constant efficiency assumption), the bid curve is piecewise constant with two segments. This means the NPP-PEM is almost always dispatched at min or max (unless it is the marginal generator). However, the K-means clustering algorithm gives the average of the time series data within the cluster, whose values may deviate from the true NPP-PEM IES operation. To avoid this, we can choose the time series data with the median capacity factor in each cluster to replace the representative time series generated by the K-means algorithm. Figure 4 shows that the median and mean capacity factors are very similar.

Figure 5 shows the training data and results for cluster 13. The median and representative time series have similar patterns and capacity factors. However, all values in the median time series are either 0 or 1, which mimics the NPP-PEM IES dispatch from the PCM simulations.

Figure 6 compares the cumulative density of the NPP capacity for the entire dataset and representative days. Using a large number of representative days (N = 22) helps ensure the capacity factors of the representative days match the entire dataset, as all red dots are distributed closely around the blue line.

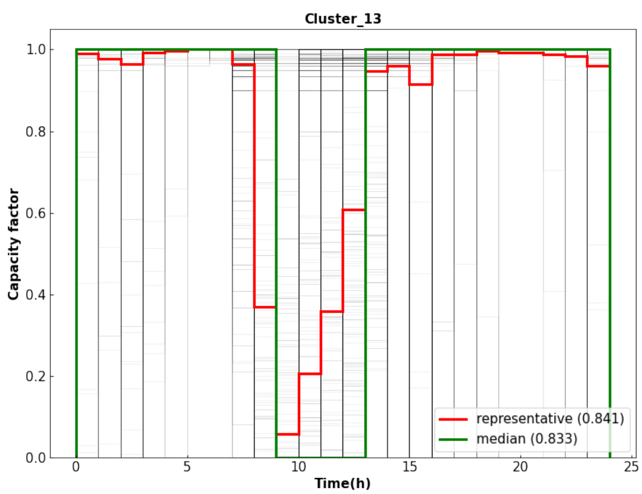

Figure 5. Clustering results of cluster 13. Black lines are the time series data in this cluster. The red and green lines are the representative time series calculated by K-means and the time series data from cluster 13 with the median capacity factor, respectively.

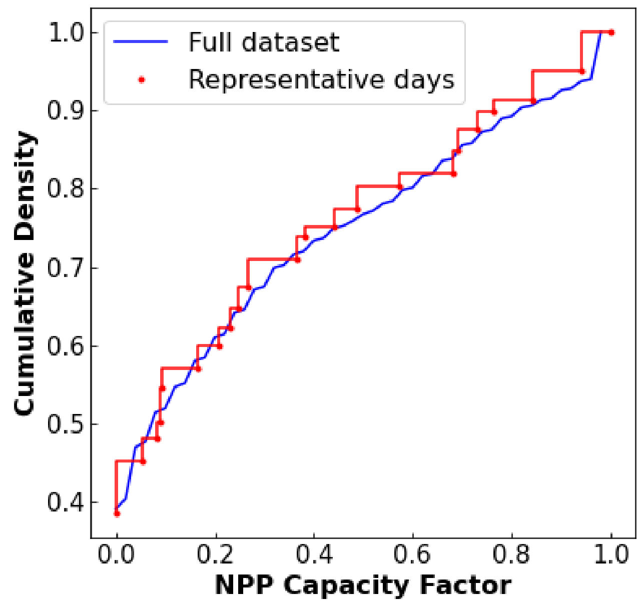

Figure 6. Comparison of the cumulative density of the capacity factor. The blue line is the full dataset used as input to the time series clustering algorithm, and the red line and dots are the representative days.

Surrogate Models are Accurate

Figure 7 compares contour plots of the electricity revenue from PCM simulations and surrogate model predictions. These plots are strikingly similar, with the root-mean-squared error (RMSE) of 0.71 M$ and R^2 of 0.995, emphasizing the surrogate's accuracy. Moreover, in both plots, the electricity revenue does not monotonically decrease with the increasing power of the PEM, which highlights the non-trivial IES-market interactions. Diverting power to co-produce hydrogen reduces the supply of electricity in the market, which increases the LMP; the average day-ahead LMP of all PCM simulations (22.43 $/MWh) is 1.5% higher than the price-taker signal (22.09 $/MWh).

The output of the frequency surrogate model predicts the weight of each representative day regarding the given IES design. For the different input x, the frequency surrogate model predicts the frequency of each representative day within a year. Table 2 summarizes the R^2 value of all 22 representative days.

Table 2: R^2 of dispatch frequency surrogate models

	Value
Mean R^2	0.9362
Standard Deviation	0.0747
Max R^2	0.9988
Min R^2	0.7631

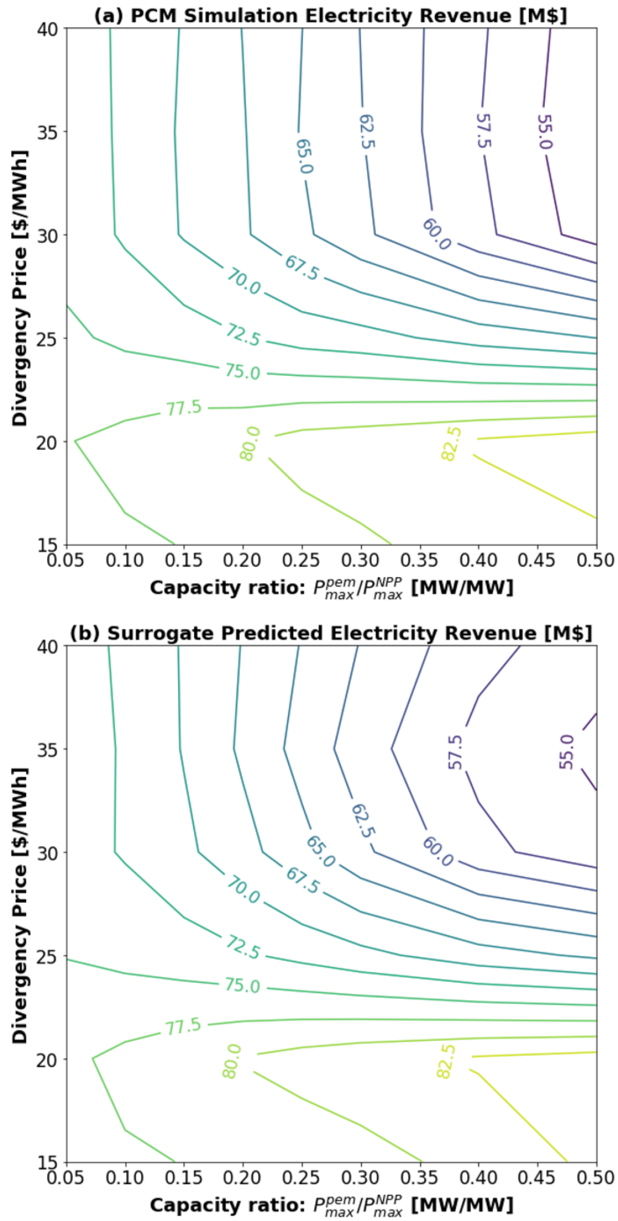

Figure 7. Contour plots of electricity revenue from (a) PCM Simulation and (b) revenue market surrogate model. In (a), electricity revenue is calculated from 48 PCM simulations with different divergency price and capacity ratio combinations. In (b), we use the surrogate model to predict the electricity revenue using divergency price and capacity ratio combinations as inputs.

To further test the accuracy of the frequency model, we calculated the annual NPP-PEM IES capacity factor predicted by the frequency surrogate model. We compared then compare these predictions with the PCM simulation data. The annual capacity factor is calculated by Equation (8):

$$cf_{annual}^{NPP} = \sum_s N_{day} \cdot \omega_s \cdot \sum_t^T \frac{cf_t^{NPP}}{T} \qquad (8)$$

ω_s is the weight predicted by the frequency surrogate model. cf_t^{NPP} is the capacity factor of the IES at time t in each representative day, N_{day} is the number of days in a year (equal to 366).

Figure 9 shows the parity plot of the predicted versus actual annual IES capacity factors. The predicted annual capacity factors closely agree with the PCM simulation results, with the root-mean-squared error (RMSE) of 0.006.

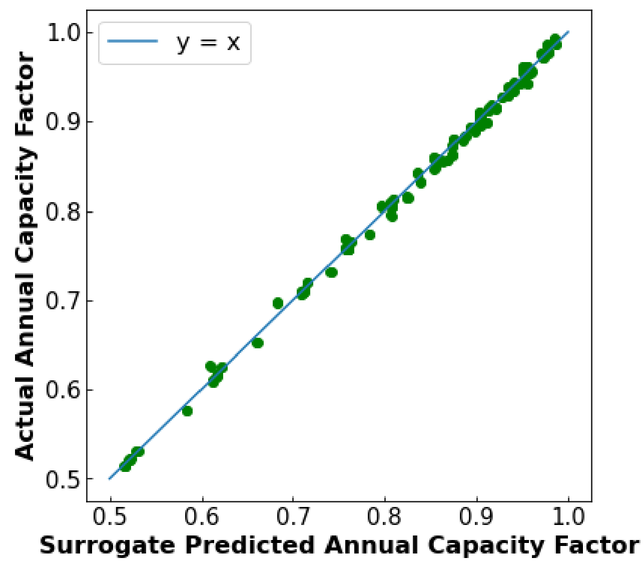

Figure 9. Annual capacity factor parity plot. The green dots correspond to the 192 annual simulations considered in the training dataset. The parity line is shown in blue.

CONCLUSIONS AND FUTURE WORK

This work develops time-series clustering methods for surrogate-assisted IES design and operation optimization considering IES/market interactions. The result shows that time series clustering identifies accurate representative days. The representative days and frequency surrogate models accurately predict the annual capacity factor. Finally, the neural network surrogate models for revenue accurately capture complex IES/market interactions as revealed by the PCM simulations (ground truth).

In future work, these surrogate models should be compared within the context of the IES optimization problem (Step 3) and compared to PCM verification simulations (Step 4). Moreover, the modeling strategy should be extended to consider energy storage systems [19], uncertainty from renewables, and additional revenues from ancillary services [20]. We also plan to extend the approach to consider price fluctuations in co-products, which is especially important as hydrogen and other renewable-energy-derived chemical markets continue to mature.

ACKNOWLEDGEMENTS

This work was conducted as part of the Design Integration and Synthesis Platform to Advance Tightly Coupled Hybrid Energy Systems (DISPATCHES) project through the Grid Modernization Lab Consortium with funding from the U.S. Department of Energy's Office of Fossil Energy and Carbon Management, Office of Nuclear Energy, and Hydrogen and Fuel Cell Technologies Office. Additional work was conducted as part of the Institute for the Design of Advanced Energy Systems (IDAES) with support through the Simulation-Based Engineering within the U.S. Department of Energy's Office of Fossil Energy and Carbon Management. Additional work was conducted in part by appointments for A.W.D. and X.C. to the U.S. Department of Energy (DOE) Postgraduate and Faculty Research Programs at the National Energy Technology Laboratory administered by the Oak Ridge Institute for Science and Education (ORISE). This work was authored in part by authors at the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCE

1. Arent, D. J., Bragg-Sitton, S. M., Miller, D. C., Tarka, T. J., Engel-Cox, J. A., Boardman, R. D., ... and Garfield, D. J. Multi-input, multi-output hybrid energy systems. Joule, 5(1), 47-58 (2021).
2. Laky, D.J., Tumbalam-Gooty, R., Jaffe, T., Holly, M., Atia, A., Chen, X., & Dowling, A.W. IDAES-PSE Software Tools for Optimizing Energy Systems and Market Interactions. FOCAPD, 2024.
3. Lee, A., Ghouse, J.H., Eslick, J.C., Laird, C.D., Siirola, J.D., Zamarripa, M.A., Gunter, D., Shinn, J.H., Dowling, A.W., Bhattacharyya, D. and Biegler, L.T. The IDAES process modeling framework and model library—Flexibility for process simulation and optimization. Journal of Advanced Manufacturing and Processing, 3(3), p.e10095 (2021). Sorourifar, F., Zavala, V.M. and Dowling, A.W.
4. Martinek, J., Jorgenson, J., Mehos, M., and Denholm, P. A comparison of price-taker and production cost models for determining system value, revenue, and scheduling of concentrating solar power plants. Applied energy, 231, 854-865 (2018).
5. Frew, B., Levie, D., Richards, J., Desai, J. and Ruth, M. Analysis of multi-output hybrid energy systems interacting with the grid: Application of improved price-taker and price-maker approaches to nuclear-hydrogen systems. Applied Energy, 329, p.120184 (2023).
6. Sousa, J.A., Teixeira, F. and Faias, S. Impact of a price-maker pumped storage hydro unit on the integration of wind energy in power systems. Energy, 69, pp.3-11 (2014).
7. Gao, X., Knueven, B., Siirola, J.D., Miller, D.C. and Dowling, A.W. Multiscale simulation of integrated energy system and electricity market interactions. Applied Energy, 316, p.119017 (2022).
8. Jalving, J., Ghouse, J., Cortes, N., Gao, X., Knueven, B., Agi, D., Martin, S., Chen, X., Guittet, D., Tumbalam-Gooty, R., Bianchi, L., Beattie, K., Gunter, D., Siirola, J.D., Miller, D.C. and Dowling, A.W. Beyond price taker: Conceptual design and optimization of integrated energy systems using machine learning market surrogates. Applied Energy, 351, p.121767 (2023).
9. Mekhilef, S., Saidur, R. and Safari, A. Comparative study of different fuel cell technologies. Renewable

and Sustainable Energy Reviews, 16(1), pp.981-989 (2012).
10. FER Commission. Energy Primer A Handbook for Energy market Basics (2020).
11. RTS-GMLC, Reliability test system-grid modernization lab consortium. https://github.com/GridMod/RTS-GMLC. [Accessed 16 November 2023].
12. Prescient production cost modeling platform. https://www.osti.gov/doecode/biblio/47057. [Accessed 16 November 2023].
13. Gao, X. and Dowling, A.W. Making money in energy markets: Probabilistic forecasting and stochastic programming paradigms. In 2020 American Control Conference (ACC) (pp. 168-173). IEEE (2020).
14. Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K. and Woods, E. Tslearn, a machine learning toolkit for time series data. The Journal of Machine Learning Research, 21(1), pp.4686-4691 (2020).
15. Keras, a Python interface for artificial neural networks. https://keras.io. [Accessed 16 November 2023].
16. Ceccon, F., Jalving, J., Haddad, J., Thebelt, A., Tsay, C., Laird, C.D. and Misener, R. OMLT: Optimization & machine learning toolkit. The Journal of Machine Learning Research, 23(1), pp.15829-15836 (2022).
17. Hart, W.E., Watson, J.P. and Woodruff, D.L. Pyomo: modeling and solving mathematical programs in Python. Mathematical Programming Computation, 3, pp.219-260 (2011).
18. Wächter, A. and Biegler, L.T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical programming, 106, pp.25-57 (2006).
19. HSL, A collection of Fortran codes for large scale scientific computation. http://hsl.rl.ac.uk/. [Accessed 16 November 2023].
20. Integrated multiscale design, market participation, and replacement strategies for battery energy storage systems. IEEE Transactions on Sustainable Energy, 11(1), pp.84-92 (2018).
21. Dowling, A.W., Kumar, R. and Zavala, V.M. A multi-scale optimization framework for electricity market participation. Applied Energy, 190, pp.147-164 (2017).