First Principles Advisory

Integrating Machine Learning with PLEXOS to Enhance Long-Term Day-Ahead Price Forecasts in CAISO

Jim Himelic

jhimelic@firstprinciples.run

April 2024

First Principles Advisory

Table of CONTENTS

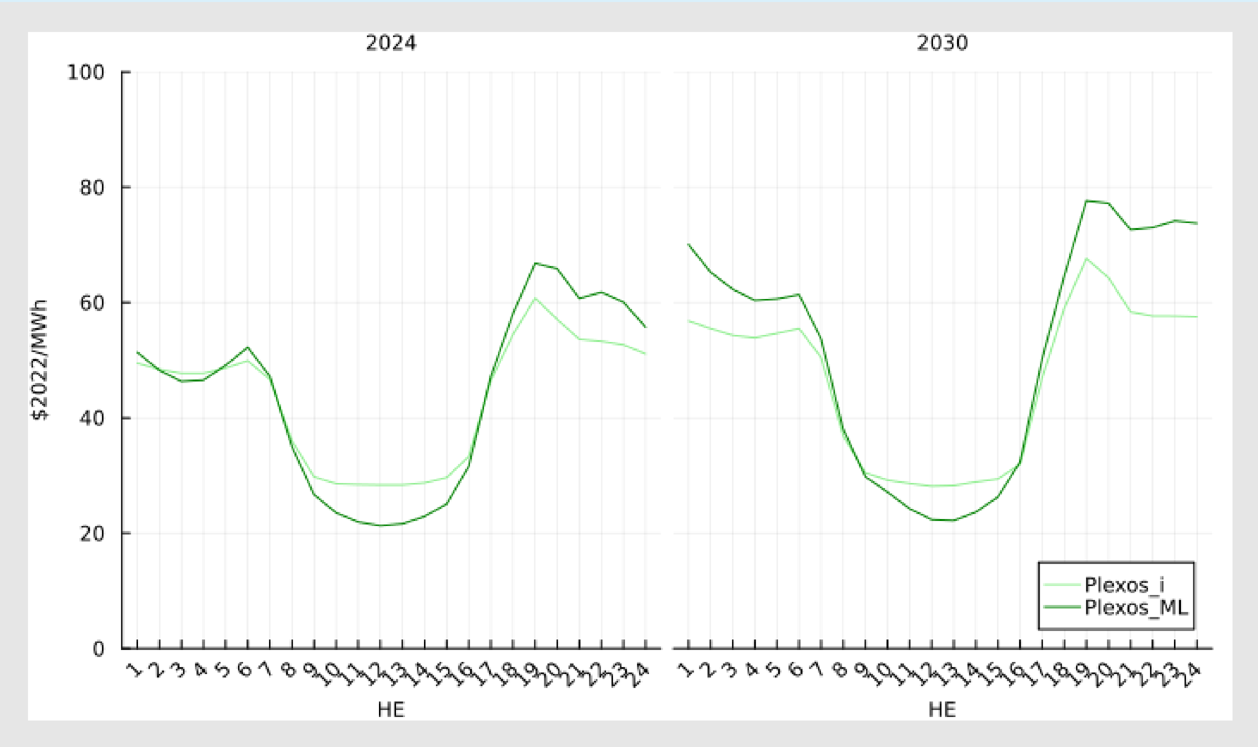
Executive Summary

Price forecasts suffer from the 'muddy middle' – a tendency to project middling prices while the real market experiences greater price extremes. As a response, this study introduces a novel approach that integrates Energy Exemplar's PLEXOS, an industry-standard fundamental production cost model (PCM), with advanced machine learning (ML) techniques to improve the accuracy of long-term day-ahead price forecasts. This method addresses PCMs' tendency to produce over-optimized results, especially during more challenging market conditions that do not align with actual market dynamics. Specifically, these models tend to underestimate prices during periods of market scarcity, when the supply of excess generation capacity is limited, and overestimate prices during periods of generation oversupply, when excess renewable energy is curtailed to maintain balanced grid operations. Enhancements to the long-term forecasts benefit organizations by providing them with more realistic future pricing environments that improve decisions related to planning and procurement responsibilities.

This analysis focuses on the California Independent System Operator (CAISO) grid. The PLEXOS model used for this study applies the results from the California Public Utilities Commission's (CPUC) 2022 Integrated Resource Planning (IRP) proceeding to define the buildout of CAISO's projected landscape of generators and storage facilities in both 2024 and 2030. The anticipated large increases in solar, wind, and battery storage are poised to significantly influence CAISO grid operations due to a transformational shift in the system's adjusted net load profile. This shift will diminish the role of natural gas facilities in setting prices, as storage becomes the marginal resource with increasing frequency. With this transfer of roles, notable changes in energy pricing are likely, as bi-modal pricing profiles become more common. However, the abundance of storage and its accompanying operational flexibility benefits is likely to amplify PCMs' propensity to over-optimize the system because actual grid operations can realize only a fraction of the projected benefits due to the inherent uncertainty associated with reliably operating the bulk electric grid. This growing divergence highlights the need for post-processing adjustments to align fundamental price forecasts more closely with actual market dynamics.

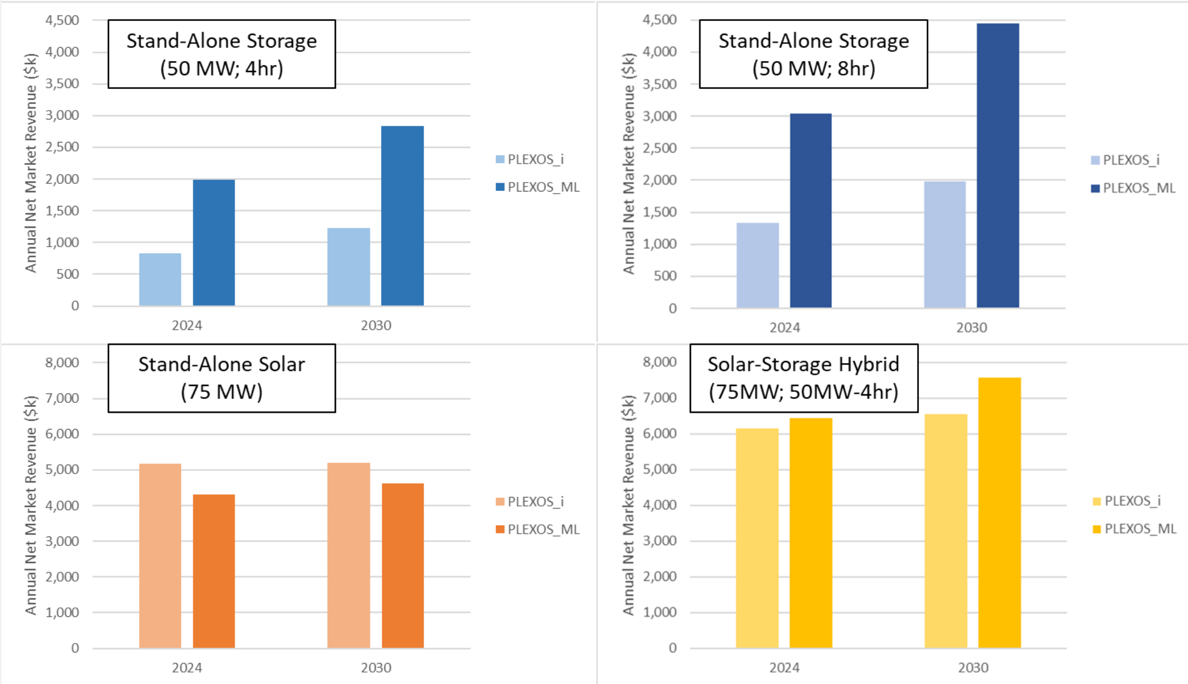
The electric power industry has always grappled with uncertainty when conducting fundamental modeling exercises, and it continues to beleaguer analysts today. While modelers have traditionally used statistical techniques to mitigate the effects of uncertainty, this study pioneers the application of machine learning to enhance the post-processing of price forecasts. Uncertainty can come in one of two forms: parametric and structural. Parametric uncertainty arises due to an inability to obtain perfect information for all input parameters such as commodity fuel prices, load forecasts, and technology costs. This differs from structural uncertainty, which is the unavoidable byproduct of having to use simplifying assumptions in the model's abstract representation of the complex real-world system it aims to emulate. This analysis focuses on the influence of structural uncertainty on PCM models when generating long-term day-ahead price forecasts.

To isolate the effects of structural uncertainty, the author first conducted a comprehensive backcast study in PLEXOS to simulate historical day-ahead market operations in CAISO for 2021 and 2022, thereby partially normalizing for the effects of parametric uncertainty. The author then partnered with Max Kanter at GridStatus.io, an energy market intelligence company that provides system operations data, to implement a machine learning model based on an industry-standard forecasting algorithm. The ML model was trained on the results from the backcast study and only fundamental outputs from PLEXOS – such as adjusted net load, effective market heat rates, and net imports – were considered to maximize the robustness of this post-processing technique. Using 2022 as the in-sample testing period, the ML model was able to improve the accuracy of the original 2021 PLEXOS price forecast by 13%. Once fully calibrated, the ML model was then used to adjust the results from the 2024 and 2030 PLEXOS runs by using the projected future values of the same variables that were used to train the model. The resulting differences in hourly price profiles between the original PLEXOS forecast, PLEXOS_i, and the ML-adjusted forecast, Plexos_ML, are illustrated in ES Figure 1. Key adjustments included lowering midday prices and raising early evening prices, thereby partially offsetting a PCM's tendency to produce unrealistic results when compared to the degree of complexity and uncertainty faced by grid operators during actual operations.



ES Figure 1: Comparison of 2024 and 2030 Annual Average Hourly Prices: Original PLEXOS Forecast (PLEXOS_i) vs. PLEXOS Forecast with Machine-Learning Adjustments (PLEXOS_ML)

The revised price forecasts had a significant impact on the financial performance of solar, battery, and solar-battery hybrid projects. ES Figure 2 illustrates the difference in annual net market revenue (i.e., includes any storage charging costs) between the original and ML-adjusted PLEXOS price forecasts. Storage facilities see a substantial revenue increase (125-138%). This contrasts with stand-alone solar facilities, which experience a moderate decline in revenue (11-14%). The hybrid project demonstrates a more nuanced impact, with a 5% increase in 2024 that grows to 16% by 2030, reflecting the evolving nature of the grid as more renewables and storage facilities come online.



ES Figure 2: Economic Impact of Forecast Adjustments Applied by Machine Learning Model on Candidate Projects

Although this initial study demonstrates promising results, additional work is required to realize the full potential of a hybrid forecasting platform that integrates fundamental modeling with machine learning. Key areas for improvement include updating the PLEXOS backcast model to cover 2023, exploring a variety of machine learning algorithms for optimal performance, and refining the PLEXOS forward model to reflect the latest trends in REC pricing and interregional transmission projects. Moreover, the platform's forecasting capabilities can be broadened to include both short- and mid-term applications by merging ML-generated forecasts with PLEXOS outputs according to user-specified weightings. In conducting this additional work, a hybrid forecasting platform will be well positioned to provide the industry with deeper market insights and support more informed decision-making.

This study illustrates the synergistic benefits that are available when leveraging machine learning techniques to refine PLEXOS' long-term day-ahead price forecasts, especially during market scarcity and oversupply grid conditions. The continued adoption of renewable energy and storage facilities will further highlight the complementary strengths of both modeling methodologies, aiding the industry in addressing market dynamics. This innovative post-processing technique paves the way for more sophisticated forecasting platforms in the future and will be instrumental in helping organizations navigate the complexities associated with the grid's transition to a low-carbon future.

Introduction

In the electric power industry, understanding how changes to the bulk power system will affect energy market prices is of immense value, especially as the grid proceeds with its decarbonization transition. To forecast future market prices for electricity, analysts commonly use two types of models: fundamental and statistical. A fundamental model simulates grid operations at hourly or intra-hourly intervals based on the technical characteristics of the generator fleet and transmission system that is tasked with serving the defined load obligation. While all fundamental models incorporate the physical constraints of the grid, the degree to which this is accomplished will vary significantly depending on the spatiotemporal settings and how the problem is formulated. However, irrespective of its level of operational detail, all fundamental models calculate energy prices based on the shadow price of the energy balance constraint for each location and simulation timestep. In contrast, statistical methods, in essence, look at the present and past to forecast future conditions. They accomplish this by utilizing sophisticated time-series analysis tools to generate price forecasts by analyzing historical datasets. Recently, the industry has expanded this to include machine learning by implementing a methodology known as 'supervised learning.' As opposed to requiring explicit programming instructions, this form of machine learning involves a self-learning algorithm that creates abstract representations from historical datasets and receives user feedback on its performance. With respect to forecasting assistance, the application of machine learning, heretofore, has primarily been focused on trading and other short-term horizons.

This study explores enhancing long-term day-ahead energy price forecasts by integrating a fundamental production cost model that is rooted in mathematical optimization with a machine learning model that is based on statistical analysis principles. In combining these two forecasting modalities, the author looks to consider how the forecast can be adjusted to better reflect actual operating conditions, thereby providing additional assistance with long-term planning and asset valuation exercises.

Fundamental Model Description

Overview

Fundamental models come in various forms, with both commercial and open-source options available, and provide analysts with a large number of input variables and constraint types to capture a detailed formulation of the bulk electric power grid. Afforded this high degree of flexibility in model design, modelers face a critical tradeoff between the degree of operational detail they incorporate and the required runtimes to successfully complete a model run. This tradeoff depends on a suite of factors such as user application, data availability, CPU processing power, and runtime constraints.

First Principles Advisory (FPA) utilizes Energy Exemplar's PLEXOS software program to generate fundamental energy price forecasts for the CAISO system and the broader Western Electricity Coordinating Council (WECC) region. FPA maintains its own customized WECC zonal database, which is cross-referenced with the PLEXOS databases publicly shared by the CAISO and the California Energy Commission (CEC). In addition, the database integrates information from the California Public Utilities Commission's (CPUC) Integrated Resource Planning (IRP) cycle to maintain alignment with the latest assumptions under that proceeding.[1]

Figure 1 below provides a simplified geographic representation of FPA's zonal model. California is represented by seven distinct zones that include both CAISO and non-CAISO regions. The CAISO region is defined by PGE, SCE, and SDGE. The non-ISO zones of California include LDWP, BANC, TID, IID, and IV-NG and are individually represented in the model along with the corresponding path ratings for key intrastate transmission corridors. Similarly, the Pacific Northwest (PacNW) and the Desert Southwest (DSW) regions are divided into approximately 10 subregions each. Interstate transmission across the entire WECC footprint is modeled as aggregate paths, incorporating transfer ratings from multiple sources, including WECC's 2022 Path Rating Catalog. When training the machine learning model, as discussed in greater detail further below, FPA aggregates the individual subzones external to CAISO into broader regional zones (e.g., CA-NonISO, PacNW, and DSW).

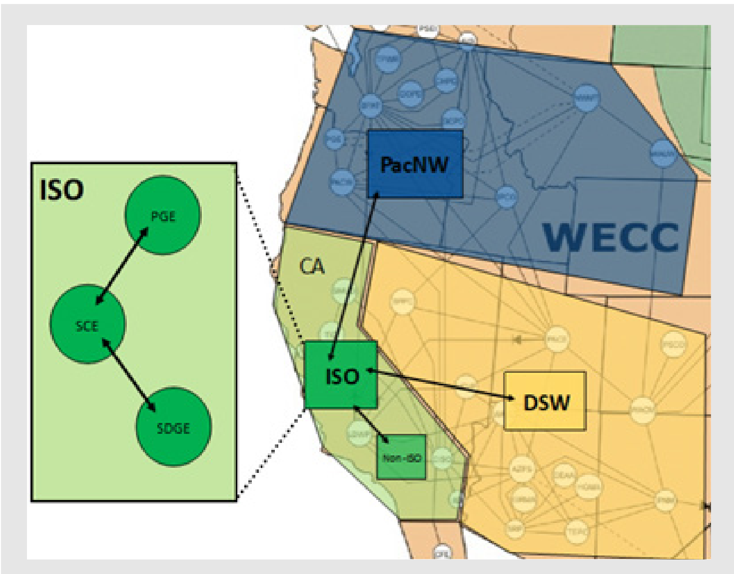


Figure 1: Simplified Representation of First Principles Advisory's WECC Zonal Model

[1] https://www.cpuc.ca.gov/industries-and-topics/electrical-energy/electric-power-procurement/long-term-procurement-planning/2022-irp-cycle-events-and-materials.

CAISO Planned Capacity

This analysis assumes the future planned capacity of CAISO's system based on the 25 MMT buildout scenario in the CPUC 2022 IRP cycle. This scenario reflects a strategic shift towards renewable generation and storage that comports with the state's long-term decarbonization goals. The left side of Figure 2 illustrates the changes in total system capacity broken down by resource type for the years 2022, 2024, and 2030. Conversely, the right side of the figure presents the same information but arranges it by resource type to emphasize the expected changes from 2022 to 2030 for each technology. Solar, wind, and storage all exhibit significant growth rates, while the system concurrently retires approximately 5 GWs of natural gas-fired generation capacity. Additionally, it's worth noting that the author assumes the postponement of the retirement of Diablo Canyon Nuclear Power Plant's two units until 2029 and 2030.

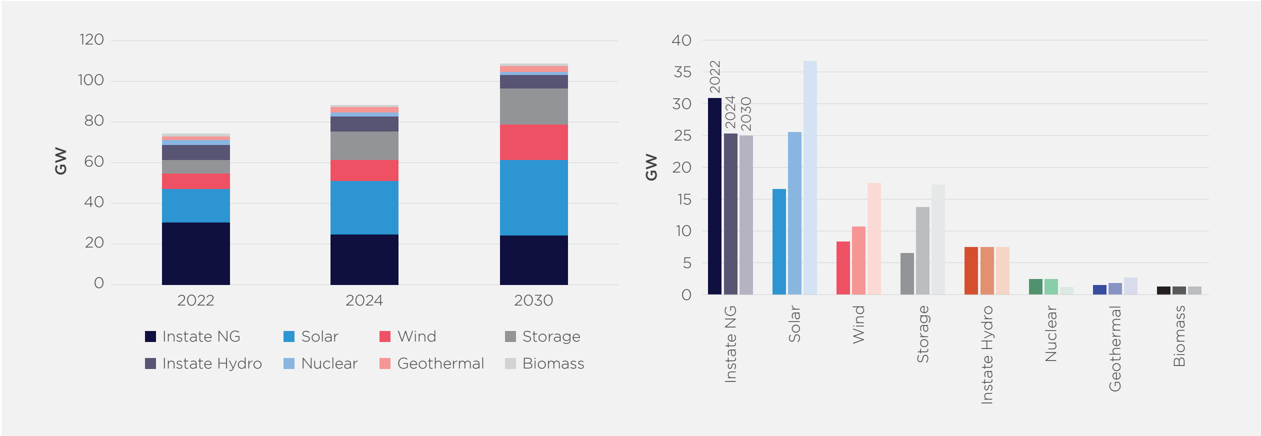


Figure 2: CAISO Total Installed Capacity Based on CPUC's 2022 IRP 25MMT Scenario

Adjusted Net Load Profiles

To illustrate the impact on system operations from the proposed system changes discussed in the previous section, the author tracks changes to CAISO's Adjusted Net Load profile, which includes additional factors beyond wind and solar. In this study, Adjusted Net Load (ANL) is defined as:

$$ANL_i = Load_i - wind_i - solar_i - Hydro_i - non\_dispatchable\_thermal_i.$$

For a given ANL profile, PLEXOS must utilize a combination of NG resources, storage, imports/ exports, and curtailments to arrive at a balanced portfolio (i.e., there is no unserved or dumped energy in any hour).[2] To aid the reader in tracking the evolving changes in CAISO's ANL profile, Figure 3 provides a direct comparison of historical data from 2022 with forecasted data for 2024 and 2030.

In comparing these ANL profiles, one can observe how the projected increase in wind and solar is expected to significantly reduce the ANL throughout the year. These changes reduce the role of natural gas-fired generation and imports from neighboring balancing authorities (BAs) in setting prices.[3] Despite the addition of over 10 GW of new battery capacity starting in 2024, storage is still projected to be outpaced by solar at a rate of 2:1, as displayed in Figure 2. Consequently, for any excess power still available after fully satisfying the storage fleet's demands for charging energy, CAISO will need to rely on either exports or curtailments to maintain stable grid operations. Should neighboring BAs be unable to absorb this excess power due to their own increase in new solar projects, CAISO will have to implement additional curtailments, absent more storage capacity coming online.

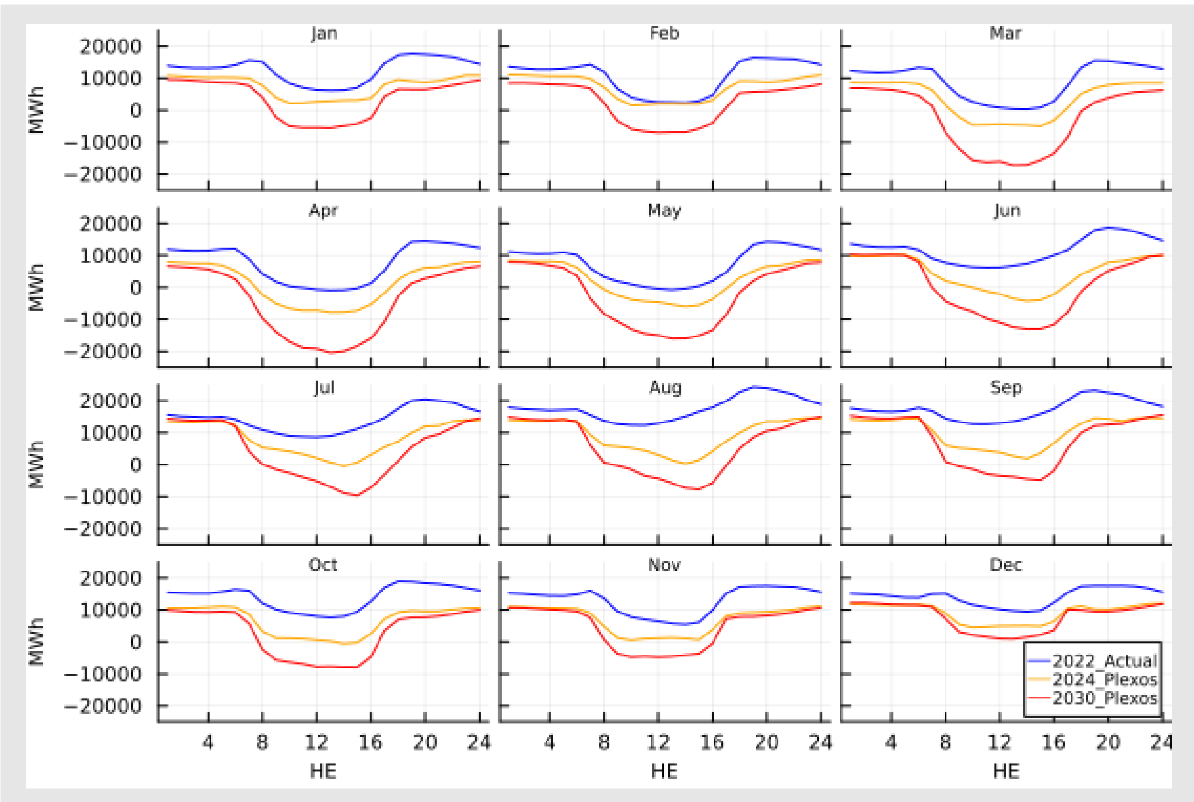


Figure 3: 2022, 2024, and 2030 Hourly Adjusted Net Load (MW) Profiles

[2] Wind and solar reflect the gross output of the solar and wind facilities pre-curtailments. Forecasted curtailments increase significantly in the model as one goes further out in time, going from 1,569 GWh in 2024 to 28,472 GWh in 2030.
[3] This statement primarily concerns the economic aspects of energy pricing under standard grid operations and does not apply to scenarios when the grid is challenged with high peak demands due to severe temperatures.

Types of Model Uncertainty

It is important to acknowledge that no model - no matter how sophisticated - can serve as a 'crystal ball' because they are all subject to uncertainty. Uncertainty arises from multiple sources and each type imparts its own unique signature on the model with varying impacts to model fidelity. Uncertainty can be categorized as either structural or parametric.[4] Parametric uncertainty is unavoidable because it arises whenever input parameters, such as natural gas prices or hourly demand profiles, can't be populated with perfect information about future conditions. Conversely, structural uncertainty arises from the model's imperfect representation of the real-world system it aims to emulate. Parametric uncertainty is relatively straightforward to understand because it arises whenever actual input values deviate from what is assumed in the model. Structural uncertainty, however, is more intricate, often resulting from oversimplified assumptions or partially complete model formulations.

Conducting a backcast study can significantly enhance one's understanding of the impact of both types of uncertainty on the model. In modeling the system with historical actual values defined for key input parameters, modelers can gain insights in two ways: 1) they can estimate what the value of perfect information is, and 2) they can isolate the impact of structural uncertainty by removing the effects of parametric uncertainty. Thus, in a well-designed backcast study, any residual discrepancy between modeled and actual prices can be attributed to structural uncertainty.

[4] Xiufeng Yue et al. A Review of Approaches to Uncertainty Assessment in Energy System Optimization Models. Energy Strategy Reviews. August 2018.

Isolating the Impacts of Structural Uncertainty

While fundamental models offer users numerous advantages, they are susceptible to various forms of structural uncertainty. The primary sources of structural uncertainty in FPA's PLEXOS model are detailed in Table 1. Each item listed is categorized as either a 'Simplifying Assumption' or 'Incomplete Information.' For this exercise, Simplifying Assumption designates when the model incorporates a reduced formulation of the real-world object it is simulating, whereas items marked as Incomplete Information flag instances when the author lacked access to the information required for a proper model configuration. A brief description of each source along with an explanation of its contribution to the model's uncertainty is also included in the table.

Model Item	Type	Description	Significance
Zonal geographic configuration	Simplifying Assumption	The model uses a zonal configuration of WECC, aggregating individual busbars into geographic regions	In a zonal model, all intra-zonal congestion is ignored and all generators in the location see the same market price
Transport model for transmission	Simplifying Assumption	The model uses a transport model for transmission operations. A simple line loss model is assumed, and path flows are not dependent on other line activity	In a DC OPF, resistance and reactance are defined for each line and path flows are dictated by engineering principles
Linear unit commitment	Simplifying Assumption	The model uses a linear commitment logic, allowing for the commitment of partial units during the optimization.	To reduce run times, the model does not enforce integer commitment, keeping the problem formulation as an LP rather than a MILP
Aggregate units	Simplifying Assumption	The model aggregates various resources by geographic region, including solar, wind, geothermal, biomass, small hydro, and large hydro.	This aggregation, using a single generation profile for regional resources of the same type, overlooks any intra-region variability
	Simplifying Assumption	The model aggregates various resources by geographic region, including solar, wind, geothermal, biomass, small hydro, and large hydro.	This aggregation, using a single generation profile for regional resources of the same type, overlooks any intra-region variability
Single start times (no hot, warm, cold)	Incomplete Information	The model lacks information on startup transition times so is unable to differentiate between hot, warm, or cold starts.	Hot starts require less fuel and time to come online and become dispatchable in contrast to cold starts, which require more fuel and time.

Model Item	Type	Description	Significance
Startup / shutdown profiles	Incomplete Information	The model does not have any constraints defined to limit operations during startup or shutdown	During these periods, ramp rates are usually constrained, making the unit non-dispatchable and ineligible to set prices
Cycling constraints	Incomplete Information	The model does not have active constraints defined to limit cycling activity at generators and storage facilities	The model is free to cycle units as long as min up/down constraints are honored. It's not uncommon, however, for facilities to be subject to active cycling constraints that differ from what's assumed in the model
Hedged NG prices vs spot prices	Incomplete Information	During the backcast run, the model sets the fuel price for all NG units equal to CAISO's market power mitigation reference prices, as posted on their OASIS site.	Actual fuel prices a generator incurs vary significantly because of multiple factors, including whether a hedge agreement is in place or not
Actual bidding behavior of market participants	Incomplete Information	The model assumes all generators are bid into the market at marginal cost	Market participants commonly bid submit generator offer curves at prices other than their marginal price

Table 1: Sources of Structural Uncertainty in First Principles Advisory's WECC Regional PLEXOS Model

As highlighted in Table 1, multiple sources contribute to structural uncertainty in the PLEXOS model, complicating the already challenging task of forecasting energy prices. This is especially true under strained supply-demand conditions because PCMs struggle to accurately reflect scarcity pricing – when elevated price premiums arise due to a reduced availability of excess supply reserves amid strong demand. These market conditions can be triggered by numerous factors, such as high demand levels, simultaneous large outages at generation or transmission facilities, and limited fuel availability. Although production cost models are based on fundamental engineering principles, they typically do not include market psychology factors in their formulations. And as experienced industry professionals are already aware: market pricing doesn't always align with market fundamentals, especially during these high-risk periods.

Post-Processing Production Cost Model Results

Acknowledging inherent limitations in PCMs, some organizations already apply post-processing techniques to their model results to reflect future pricing conditions that are more likely to materialize. For example, as part of the CPUC's Avoided Cost Calculator (ACC) proceeding, the consulting firm E3 implements post-processing techniques to adjust energy prices from the CPUC staff's PCM runs, which are done using Astrape's SERVM model. As described by E3, SERVM is a "production simulation model representing a theorized and optimized view of the day-ahead energy market."[5] To compensate for the model's tendency to produce overly optimized results, E3 incorporates a 'scarcity scaling function,' which is a statistical technique that adjusts implied market heat rates and energy prices to better reflect actual prices when the grid operates close to its maximum limits.

Although it's likely other organizations utilize similar post-processing techniques, the specifics of these methods often remain confidential due to the commercially sensitive nature of the information. Consequently, the author has not come across any other existing studies that publish their post-processing procedures. At the same time, the recent advancements in artificial intelligence have led to a growing adoption of machine-learning tools in multiple industries, including the electric power sector. This inspired the author to conduct a study evaluating the efficacy of applying machine learning as a post-processing tool to refine price forecasts generated by a production cost model such as PLEXOS.

[5] E3. 2022 Distributed Energy Resources Avoided Cost Calculator Documentation: For the California Public Utilities Commission. June 2022, which can be accessed here: https://www.cpuc.ca.gov/-/media/cpuc-website/divisions/energy-division/documents/demand-side-management/acc-models-latest-version/2022-acc-documentation-v1a.pdf

Methodology

The primary objective of this study is to integrate PLEXOS with a machine-learning model to enhance the accuracy of the model's price forecasts by yielding outputs that are more accurate across a diverse range of operating conditions. Figure 4 presents a high-level overview of the study methodology.

Methodological Framework for the PLEXOS-Machine Learning Hybrid Forecast Platform

Step 01

Data Collection and Preparation

- Populate the PLEXOS XML database for both the backcast and forward model runs.
- Utilize data from multiple sources, including California agencies (CPUC, CEC, CAISO) and the EIA.
- Gather historical actuals for the backcast model from CAISO's OASIS site via GridStatus.io.

Step 02

Model Execution

- Conduct backcast model runs in PLEXOS to generate energy price forecasts for CAISO system operations for calendar years 2021 and 2022.
- Conduct forward model runs in PLEXOS to generate energy price forecasts for CAISO system operations for calendar years 2024 and 2030.

Step 03

Preparation of Training Dataset for the Machine Learning Model:

- Calculate the load-weighted CAISO DLAP (Day-Ahead Locational Marginal Price) from the outputs of the backcast and forward model runs.
- Prepare the training datasets for the machine-learning model.

Step 04

Machine Learning Calibration

- Perform in-sample testing using 2022 data to configure and calibrate the ML model.
- Perform out-of-sample testing using 2021 data to evaluate the model's performance.

Step 05

Post-Processing

- Apply post-processing adjustments to the original PLEXOS price forecasts based on the systemic relationships discovered in Step 4.

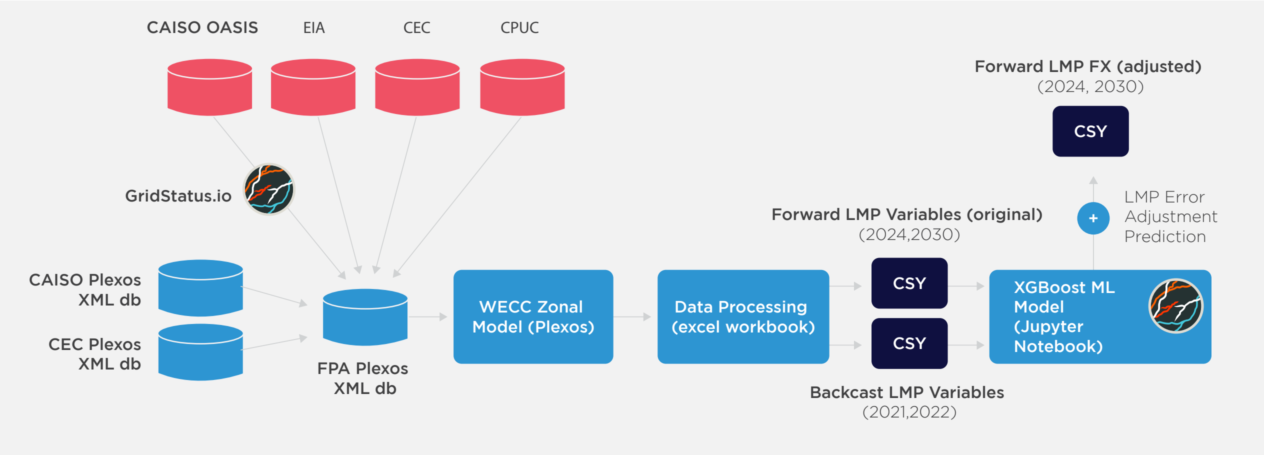


Figure 4: Graphical Outline of First Principles Advisory's Post-Processing Methodology

Results

Historical Actuals and Initial Plexos Output

Figure 5 displays the CAISO historical load-weighted Default Load Aggregation Point (DLAP) prices alongside the initial results from the PLEXOS backcast run for 2021 and 2022. Comparing actual prices with modeled results highlights the PCM's difficulty in mirroring system performance during operationally challenging days. Given the numerous sources of structural uncertainty listed Table 1, these results are not unexpected. Moreover, these challenges are not unique to PLEXOS, as all PCM models are also affected by these factors.

CAISO experienced multiple periods of significant market stress in both 2021 and 2022, as demonstrated by an examination of the actual hourly prices throughout the year. In 2021, February's Winter Storm Uri resulted in extreme price spikes, and the summer months underwent recurring bouts of constrained supply. Similarly, in September 2022, a late-season heatwave, reminiscent of the August 2020 blackouts, stretched CAISO's grid to its limits. As new all-time peak record loads were being set with demand exceeding 52,000 MW, grid operators issued a Level 3 EEA warning on September 6th, and day-ahead prices soared to nearly $1,400/MWh. In December, a surge in natural gas prices across CAISO and the broader WECC region closed out the year with elevated electricity prices.

Figure 6 reorganizes the same data included in Figure 5 to present it as a heatmap with a modified legend scale to zero in on more typical prices. While the PLEXOS model tends to slightly overestimate midday prices and underestimate afternoon-evening prices, it generally performs well in capturing market behavior on most representative days throughout the year. These results are indicative of a reasonably configured backcast model and suggest that the effects of structural uncertainty are greater during challenging operational periods and less so when sufficient balancing reserves are available in the model.

For long-term planning and procurement, assessing the value of a PCM model such as PLEXOS should be based on its ability to simulate unit commitment and economic dispatch under normal system conditions. Nonetheless, the ability to accurately forecast prices during market scarcity and oversupply is still valuable, especially if the frequency and magnitude of these types of conditions increase over time. While price forecast accuracy is important for asset valuation studies of all technology types, for resources such as battery storage and natural gas combustion turbines it is critical. This is because these resources are expected to be the marginal, price-setting unit for the majority of their operating hours. Thus, the author turned to machine learning to explore the additional value that might be derived from pairing PLEXOS with an ML model.

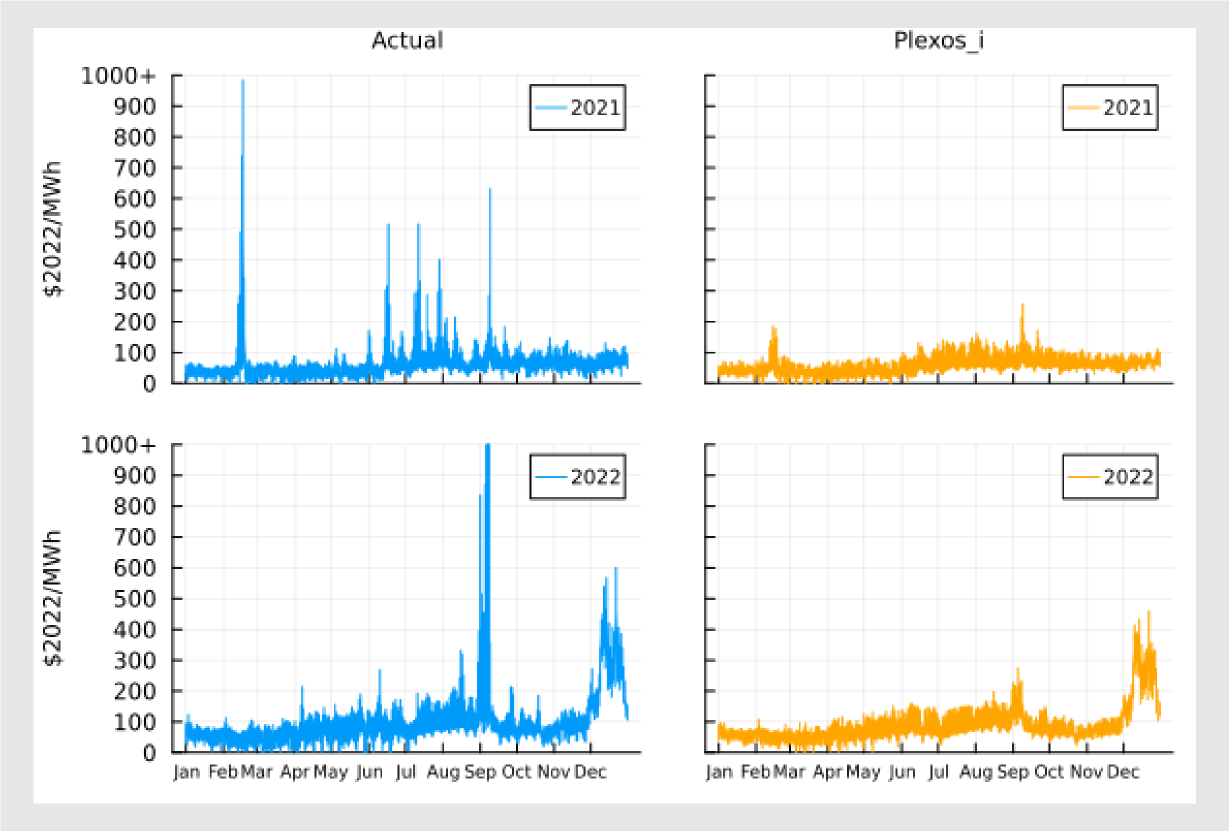


Figure 5: Actual Historical and PLEXOS Load-Weighted CAISO DLAP LMPs ($/MWh)

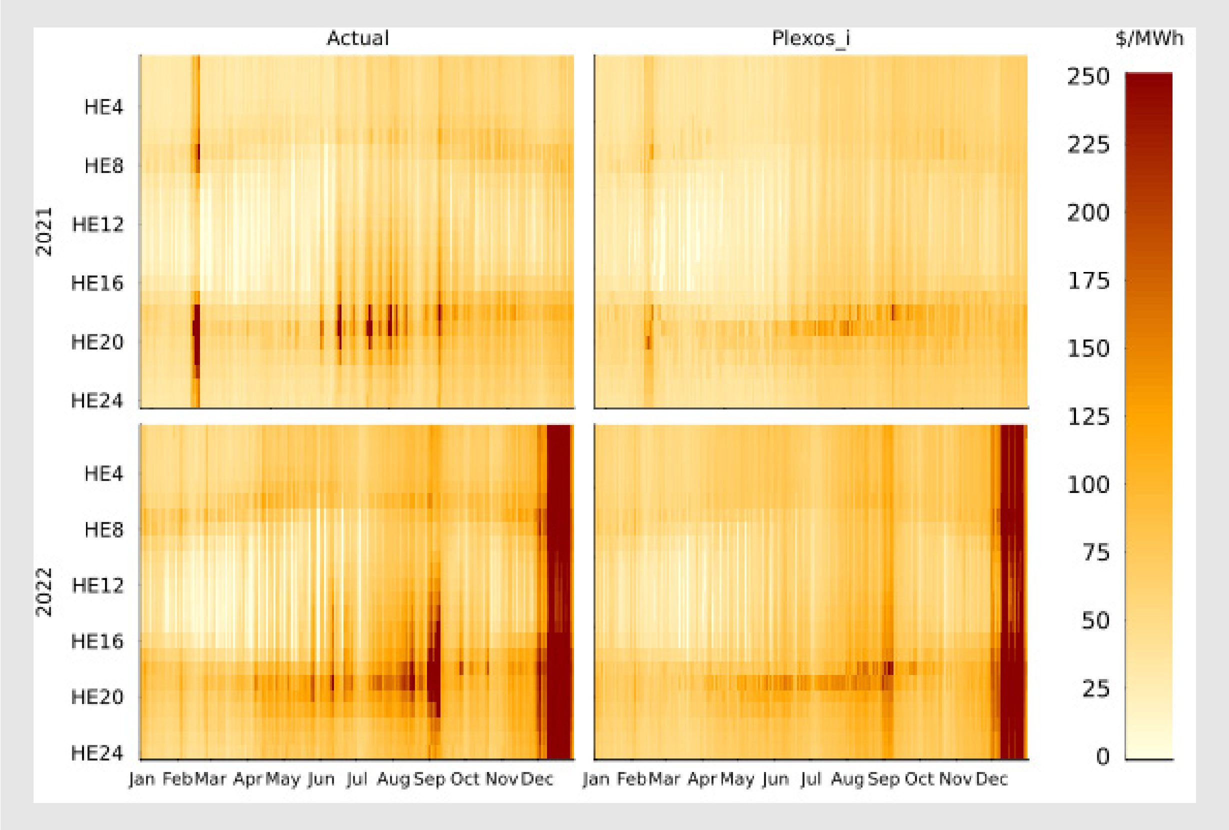


Figure 6: PLEXOS Modeled Load-Weighted CAISO DLAP LMPs ($/MWh)

Machine Learning Model

For the machine learning segment of this study, the author collaborated with the data service provider GridStatus.io.[6] In addition to supporting the retrieval of historical data from CAISO's OASIS website to populate the backcast model, GridStatus.io also built a machine learning model that utilized the 'XGBoost' algorithm. XGBoost, which stands for eXtreme Gradient Boosting, is recognized by NVIDIA as "the leading machine learning library for regression, classification, and ranking problems.'"[7] This versatile, open-source algorithm is designed to handle complex, non-linear data interactions and is equipped with multiple features, such as overfitting risk mitigation and model parameter tuning assistance.

Machine Learning Calibration

The ML model built by GridStatus.io offers multiple user-configurable settings to optimize its performance. The model can be trained in one of two ways:

Option 1: Predict forecast error (i.e., $LMP_{actual} - LMP_{initial\ fx}$) based on the set of explanatory variables provided, which then gets added to the initial PLEXOS forecast to arrive at the final value (i.e., $LMP_{final\ fx}$);

Option 2: Directly predict the final forecast i.e., $LMP_{final\ fx}$) by including the initial PLEXOS forecast (i.e., $LMP_{initial\ fx}$) as one of the explanatory variables.

Moreover, users can specify the in-sample dataset (training year) and the out-of-sample dataset (evaluation year) for model evaluation. To help manage the impact large statistical outliers have on the model, a limit can be set to cap the maximum hourly price correction applied to the original PLEXOS forecast.

Through a mix of expert judgement and iterative testing, the author determined the optimal settings that minimized fitting error. Option 1 was selected as the preferred training method. Furthermore, the model was trained using 2022 data and evaluated against 2021 data. During the training exercise, the author observed the model displayed heightened sensitivity to extreme outliers in the training data, which manifested downstream as overfitting problems and adversely affected model performance. To address this, historical prices were capped at $500/MWh to strike a balance between minimizing overfitting while still capturing an appropriate amount of market scarcity conditions. Additionally, the maximum correction parameter was set at $250/MWh to ensure the ML-adjusted forecast remained within a reasonable range of the original PLEXOS forecast. Further investigation is needed to optimize the trade-off between capturing extreme market risk premiums and minimizing overfitting risk.

Specifics regarding the selection of the training variables are undisclosed due to confidentiality. However, it's worth mentioning that only variables fundamentally tied to the underlying changes in the system - such as adjusted net load, system curtailments, and effective market heat rates – were selected. Because the grid is expected to undergo a significant transformation over the next decade, the model deliberately avoids using any parameters not directly tied to a fundamental system operating characteristic. For example, calendar-based parameters such as month, day, or hour were not eligible as training variables. The strategy used to select the variables that ultimately trained the ML model was intentionally designed to accommodate the significant changes in CAISO's projected adjusted net load profile, as previously discussed.

[6] https://www.gridstatus.io/
[7] https://www.nvidia.com/en-us/glossary/xgboost/

Machine Learning Testing Results

This section discusses the results from the out-of-sample testing of the ML model. The forecast from the original PLEXOS backcast study is denoted as PLEXOS_i, and the forecast modified by the ML model is denoted as Plexos_ML.

Figure 7 illustrates the 2021 month-hour average profiles for actuals, PLEXOS_i, and Plexos_ML.

Key observations include:

- The original PLEXOS forecast generally aligns with the actuals, mirroring the overall trend for most hours across each month. This indicates the backcast model provides a reasonably accurate depiction of the CAISO system when key input parameters are accurately defined.
- Both PLEXOS_i and Plexos_ML struggled to capture the severe market stress experienced in February 2021 due to Winter Storm Uri.
- The ML model's performance during peak afternoon and early evening hours in summer months was inconsistent in that it reduced systemic under-forecasting biases in some months (e.g., July) but over-corrected in others (e.g., August).
- The ML model effectively accounted for PLEXOS' tendency to under forecast system curtailments of excess solar energy and adjusted mid-day prices downward in non-summer months in response.

Applying the ML revisions to the original 2021 PLEXOS forecast achieved a 13% reduction in the root mean squared error (RMSE). At first glance, this improvement might appear modest, but it's important to note that the backcast study already captured a significant level of market representativeness, as illustrated in the figure. A feature importance analysis identified adjusted net load and gas prices in CAISO and the desert southwest as the three primary variables with the greatest impact on the ML model's performance.

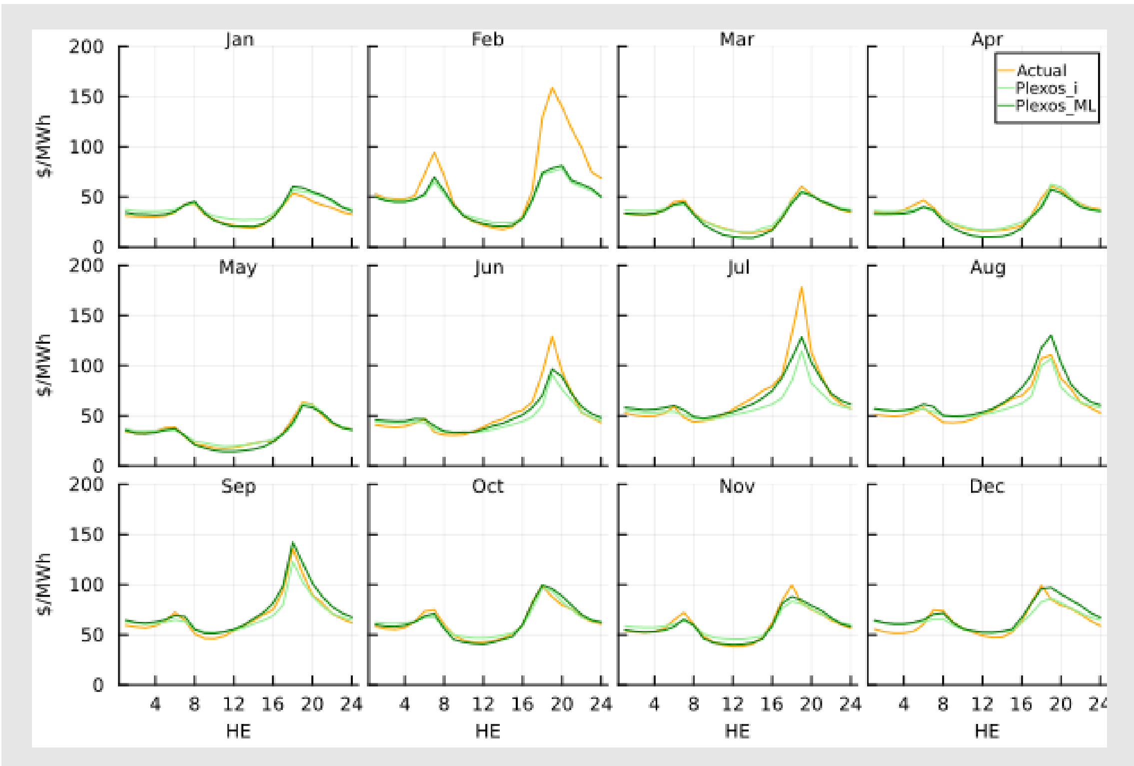


Figure 7: 2021 Month-Hour Average Profiles of Actual, PLEXOS_i, and Plexos_ML Forecasts

[7] Based on CAISO Production and Curtailment data, which can be accessed here: https://www.caiso.com/informed/Pages/ManagingOversupply.aspx.

Forecast Adjustments for Future Periods

Once calibrated, the ML model processed the 2024 and 2030 price forecasts from PLEXOS by correcting for the same systemic biases it identified in the PCM model during its training. Because the ML model was trained only on key fundamental parameters such as adjusted net load, effective market heat rates, and system curtailments, its adjustments can remain relevant despite the extensive expansion of renewable and storage resources that is anticipated in the CPUC's 2022 IRP. Figure 8 and Figure 9 display the 2024 and 2030 month-hour average profiles for the original and revised price forecasts, respectively. In both years, there is a general tendency for the ML model to reduce prices in the middle of the day and increase them in the late afternoon and early evening periods. The magnitude of these adjustments varies by month. Notably, however, the 2030 forecast reveals stronger upward price adjustments during nighttime hours in multiple months. The author attributes this to changes in natural gas prices and increased curtailments of out-of-state wind due to limited transfer capacity on the existing transmission system.

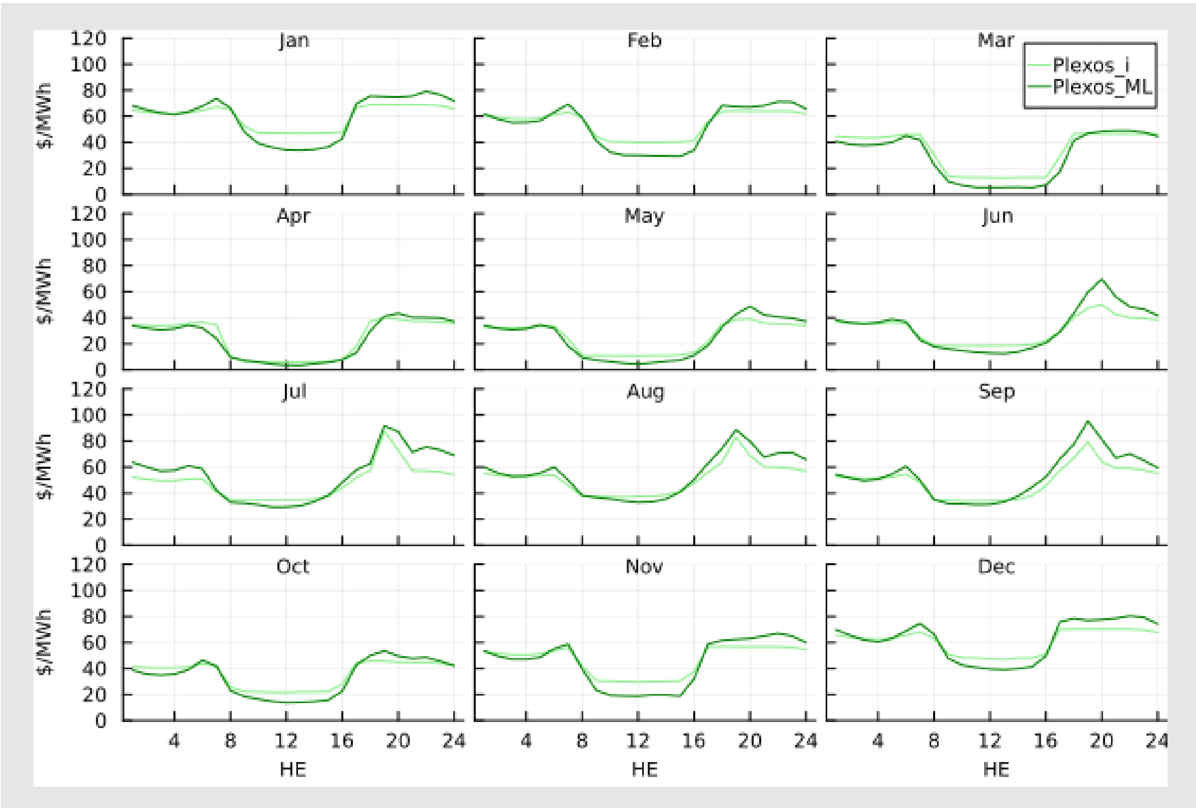


Figure 8: 2024 Month-Hour Average Profiles for initial PLEXOS Forecast (PLEXOS_i) and ML-Adjusted Forecast (Plexos_ML)

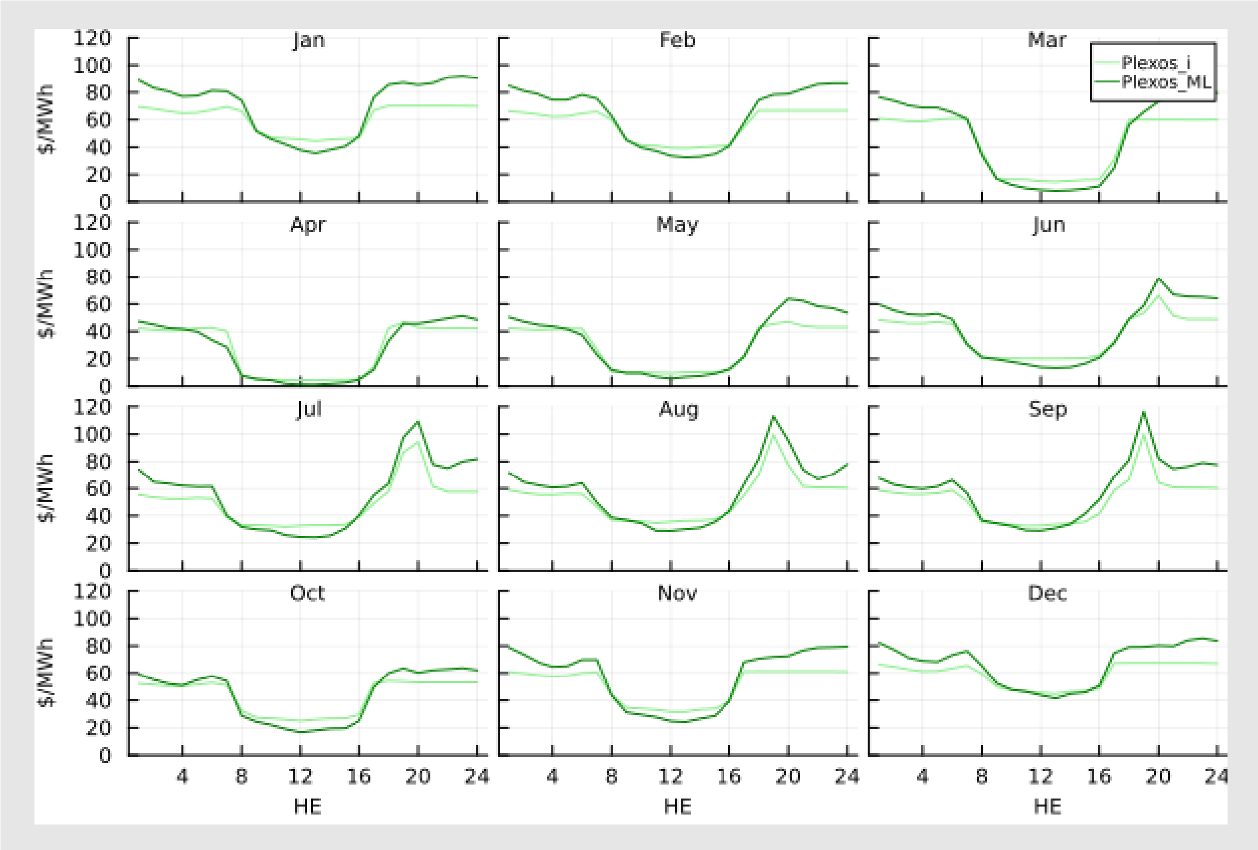


*Figure 9: 2*030 Month-Hour Average Profiles for initial PLEXOS Forecast (PLEXOS_i) and ML-Adjusted Forecast (Plexos_ML)

Given the CPUC's IRP buildout calls for over 10 GW of added storage by 2030, the model gains significant operational flexibility when optimizing its decisions. Unfortunately, this increase in storage also carries the risk of results that are over-optimized because of the PCM's simplified problem formulation, which omits key factors like load and VER forecast errors, voltage and frequency regulation, and market bidding and scheduling activities. With ~14 GW of storage assumed to be online by 2024, batteries assume a greater role in setting prices, and CAISO begins to experience bi-modal pricing patterns in most months outside of the summer peak season. Notwithstanding that price forecasts are dependent on multiple variables that include both changes to the load forecast in addition to an evolving supply stack, PLEXOS exhibits a proclivity to generate relatively stable and subdued prices over the planning horizon. But as demonstrated in Figures 8 and 9, the adjustments applied by the ML model can help correct for this overfitting, thereby allowing for long-term price forecasts that align more closely with more realistic operational conditions.

Figure 10 presents an x-y scatter plot for 2024 and 2030, with PLEXOS_i on the x-axis and Plexos_ML on the y-axis. The plot shows that the ML model generally adjusts prices downward when the PLEXOS forecast is below ~$25/MWh and adjusts them upwards when pricing is above ~$50/MWh. Most upward adjustments are confined to $25/MWh or less. But as the PLEXOS model forecasts higher prices, the magnitude of the ML model's upward pricing adjustments also increases. The wider distribution of upward pricing adjustments compared to the downward revisions highlights the complex nature of price formation processes as system supply conditions begin to tighten. It's important to note that in both calendar years PLEXOS never forecast negative day-ahead prices. This is reflective of the modeling assumption that all solar and wind units are bid into the market at $0/MWh. However, in the event a market premium for renewable energy credits (RECs) is defined in PLEXOS, negative prices will manifest in the model to reflect the opportunity costs associated with these environmental attributes.[8]

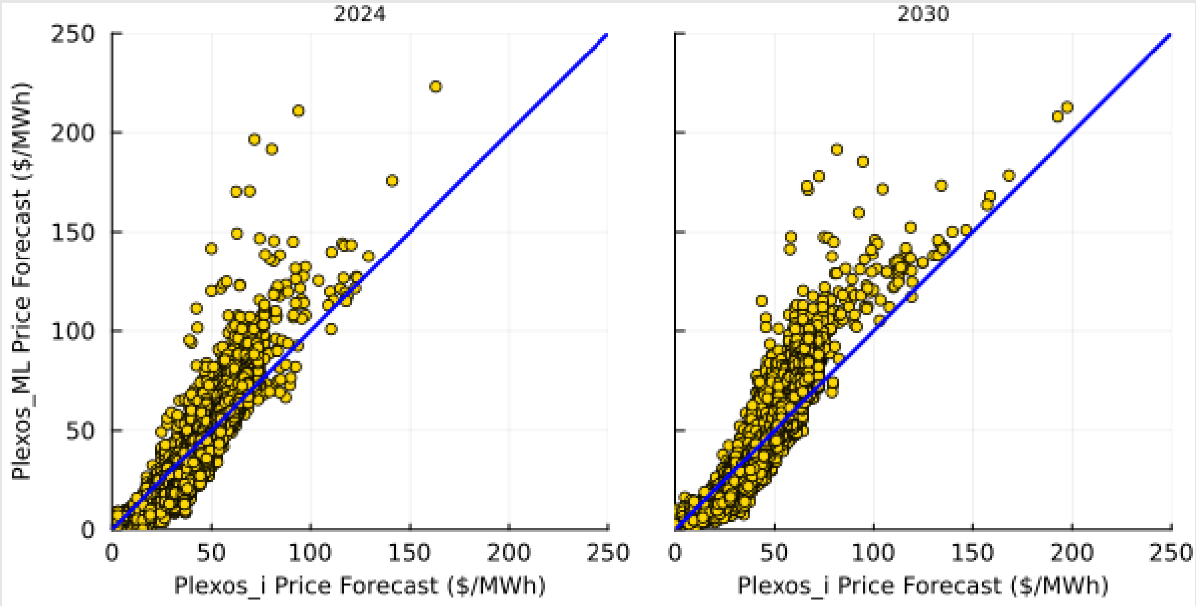


Figure 10: X-Y Scatter Plot of 2024 and 2030 Forecasts

[8] Under this assumption, if RECs are valued at $15/MWh and PLEXOS is actively curtailing a solar or wind facility, the price for that hour will be -$15/MWh.

Impacts to Valuation for Solar and Storage Assets

To evaluate the financial impact of the revised price forecasts, the author conducted an asset valuation study on multiple configurations of storage and solar resources, assuming both a stand-alone and paired configuration. Four distinct projects were analyzed: 1) a 50 MW, 4-hour stand-alone storage facility; 2) a 50 MW, 8-hour stand-alone storage; 3) a 75 MW stand-alone solar facility; and 4) a hybrid project combining a 75 MW solar plant with a 50 MW, 4-hour battery. Figure 11 illustrates the differences in annual net market revenue when switching from the original PLEXOS forecast (PLEXOS_i) to the machine-learning adjusted forecast (Plexos_ML).[9]

As illustrated in the figure, the ML model's adjustments significantly impact each project's financial performance differently. With annual net market revenues more than doubling in both 2024 and 2030, stand-alone storage – both 4-hour and 8-hour duration projects - notably benefit. Thanks to the lower midday charging costs that are then coupled with higher discharge prices later in the day, annual net revenues increase by 125-138%. In contrast, the ML model's daytime price reductions result in a 17% decrease in energy market revenue for stand-alone solar in 2024, which then narrows to 11% decrease in 2030. The solar-storage hybrid project initially experiences a modest revenue gain of 5% in 2024 but later sees a 16% increase in revenue by 2030. Future studies will explore the impact on other resource types like wind and geothermal.

Figure 11: Economic Impact of Forecast Adjustments Applied by Machine Learning Model on Candidate Projects

[8] In this study, annual net market revenue is defined to include all costs associated with charging the storage device but excludes any revenue associated with capacity payments or the provision of ancillary services.

Limitations

While this initial study is informative in revealing the benefits of integrating PLEXOS with machine learning-based techniques for price forecasting, it also highlights several limitations. Price forecasting is inherently complex, and all forecasting methods – both fundamental and machine-learning - have their limitations. As demonstrated in this study, fundamental models such as PLEXOS can be challenged in dealing with structural uncertainty, increasing the risk of the model generating outputs that aren't reflective of actual day-ahead scheduling operations. Conversely, with machine learning, there's a risk that the model may anchor around false or tenuous signatures in the historical dataset that are less relevant in future market conditions, resulting in the assignment of an inaccurate bias to the original price forecast. The difficulty of detecting and correcting for this error is compounded by the opaque nature of these machine learning models, which complicates the understanding of the specific correlations and patterns they employ.

Given the symbiotic nature of the relationship between the PCM and ML models, maximum benefit is achieved only when both models are performing well. A poor configuration in one model can severely limit the performance of the other. Limitations in both the PLEXOS and ML model have been identified, and the following areas warrant additional investigation:

01 Enhance the PLEXOS backcast model by incorporating a richer historical dataset that includes calendar year 2023. Furthermore, additional detailed information on transmission derates, generator availability, and public bidding and scheduling data can be defined in the model to further reduce the impacts of parametric uncertainty. As the accuracy of the backcast study improves, the machine learning model will train on a higher quality dataset.

02 Expand the machine learning analysis by exploring additional explanatory variables or assessing alternative algorithms. Should specific algorithms exhibit superior performance under certain conditions, an ensemble approach (i.e., employing a combination of algorithms) could be implemented to maximize the available benefits.

03 Improve the treatment of curtailments in PLEXOS by adjusting how the model accounts for any opportunity costs associated with RPS REC credits. In addition, update the model's transmission topology to account for new interregional transmission lines that are expected to come online in the next few years and will import out-of-state power into California (e.g., TransWest Express and SunZia Transmission Project).

Next Steps & Additional Applications

Although notable limitations have been identified in this initial proof-of-concept, the findings are promising and lay the groundwork for future enhancements to this hybrid forecasting platform. Moreover, by conducting asset valuation studies on a broader range of project types - including wind, geothermal, and various storage technologies beyond 4-hr and 8-hr Li-Ion batteries – the platform's utility can be greatly expanded to a broader pool of decision makers.

Exploring the integration of forecasts from a fundamental model with those from a machine learning model presents a promising area for further research. This approach could mitigate the inherent limitations of each forecasting method by leveraging their respective strengths. For example, statistical models typically excel in short-term forecasting, while fundamental models are more accurate over the long term. In combining the two forecasting methods, an organization can establish a more comprehensive approach that spans across multiple time horizons. Additionally, machine learning could aid in calculating basis premiums between regional trading hubs and individual nodes, facilitating the transformation of zonal PCM forecasts into nodal forecasts. This could considerably reduce computational demands and costs, although further investigation is required to validate these hypotheses

Conclusion

In this white paper, the author evaluated the potential to enhance long-term day-ahead price forecasts in CAISO by coupling PLEXOS with a machine learning model built off an industry standard algorithm. After using PLEXOS to conduct a detailed backcast study of CAISO operations for 2021 and 2022, the author partnered with GridStatus.io to assess the ML model's ability to correct for a PCM's tendency to over-optimize system operations, particularly during market scarcity and oversupply conditions. After training the ML model on the results from the 2021 backcast study, the author was able to reduce the RMSE of the original 2022 PLEXOS results by 13%.

After calibration, the ML model adjusted the PLEXOS outputs from 2024 and 2030 to assess the impacts on pricing after accounting for the new resources that are expected in the CPUC's 2022 IRP. This includes over 20 GW of new solar, nearly 11 GW of incremental storage, and almost 10 GW of additional wind. The primary adjustments consisted of lowering midday prices and raising late afternoon and early evening prices. Prices in 2030 also saw additional upward revisions from the ML model during the late evening and early morning hours in multiple months. Changes in commodity fuel prices along with additional curtailments of out-of-state wind generation due to congestion are believed to be the primary drivers behind the revisions.

The asset valuation exercise demonstrated that the ML price adjustments have a substantial impact on the financial performance of the projects with effects varying based on the technology type and project configuration. Stand-alone storage facilities experienced a net revenue increase of 125-138%, while stand-alone solar saw reductions of 11-14%. A solar-storage hybrid project experienced moderate revenue gains of 5% in 2024 that increased to 16% in 2030.

While promising in addressing the inherent complexities and uncertainties of price forecasting, this study acknowledges the need for improvements in both the fundamental and machine learning models to fully realize their synergistic potential. Key areas for further investigation include enriching the PLEXOS backcast model with more recent and detailed data, exploring diverse machine learning algorithms, and refining model performance on curtailments and transmission updates. Potential future enhancements to broaden the scope of this hybrid forecasting platform include going beyond just long-term forecasting applications to include short- and mid-term horizons as well. However, further exploration is required to fully validate this expanded functionality.

This study illustrates the capacity of machine learning to augment PLEXOS' day-ahead price forecasts, showcasing the combined strength of these technologies in addressing traditional model limitations and advancing energy price forecasting. As the sector moves towards a more sustainable future, such innovations are crucial for adeptly managing the evolving dynamics of the energy landscape.

Acknowledgements

The author would like to acknowledge Keith Parks for his editorial review and the helpful comments and suggestions he provided.