# OpenEEmeter – Hourly Model

## Lead Contributors

- Armin Aligholian
- Travis Sikes
- Jason Chulock
- Adam Scheer
- Brian Gerke

# 1 Introduction

## 1.1 Motivation

The OpenEEmeter 4.0 hourly model, implemented from the CalTRACK 2.0 methodology, has been used to measure hundreds of demand-side portfolios. Over time, certain limitations have emerged that hinder its effectiveness, particularly for buildings with solar photovoltaic (PV) systems. For instance:

- Solar PV Challenges: The OpenEEmeter 4.0 hourly model, now called the legacy hourly model, does not incorporate solar irradiance data. Due to a lack of solar irradiance data inaccuracies, particularly under varying cloud conditions, for customers with solar PV systems. For example, baseline data dominated by sunny days led to overestimation of solar PV generation during cloudy periods in reporting data.
- Performance and flexibility concerns: For meters without solar PV, the model's performance is sufficient, but the inability to leverage supplemental data limits its adaptability and development of potential improvements.
- Overfit model: The legacy hourly model tends to perform considerably better in the baseline data (training data) than in reporting period data (test data) indicating an overfitting issue.
- Efficiency: Given the improvements that were able to be made to the daily model in OpenEEmeter 4.0, it was suspected that significant efficiency improvements could be made to the hourly model to make it less costly to run.
- Difficult to use: The legacy hourly model was not developed with a cohesive API in mind, resulting in an unintuitive architecture that is difficult for users and developers alike.

Addressing these shortcomings formed the foundation for developing the new hourly model.

## 1.2 New Model Goals

The new model was developed to address the limitations of the legacy system while maintaining or improving its existing strengths. The primary objectives of the new model include:

1. Improved Solar PV Prediction: Better handling of meters with solar PV by integrating solar irradiance data to capture variability caused by solar irradiance and cloud coverage.
2. Performance Retention and Enhancement: Maintain or exceed the legacy model's accuracy for non-solar PV meters.
3. Flexibility: Incorporate supplemental data, such as additional time series or categorical variables, when available, to enhance predictions.
4. Faster Computation: Ensure that the model performs computations more efficiently.
5. Enhanced Usability: Simplify the API for easier and more seamless integration into workflows.

By achieving these goals, the new model is positioned as a robust, adaptable, and efficient replacement for the legacy hourly model.

## 1.3 Brief Model Overview

The new hourly model is a performant, flexible, data-driven framework designed to predict electricity consumption for individual meters. Key features of the model include:

Required Inputs:

1. Temperature time series (primary input variable)
2. Energy usage time series (target variable)

Optional Inputs:

1. Solar irradiance time series (solar PV system input variable)
2. Supplemental data: The model can also utilize supplemental time series or categorical variables that the user expects to be predictive of customer energy consumption.

Framework:

The model is built using the Elastic Net framework, a linear regression model with regularization. This regularization enhances predictive power by performing feature selection to limit unnecessary model complexity while maximizing model accuracy.

Prediction Mechanism:

OpenEEmeter 3.0 gets the input data per hour and predicts the electricity consumption for the same hour. Conversely the new model operates on a 24-hour prediction framework. It ingests 24-hour input data (temperature, optional solar irradiance, and categorical variables) for a day to predict electricity consumption for the same day 24 hours.

This architecture ensures that the new hourly model not only outperforms the legacy model in critical areas but also provides an extensible solution for future energy measurement challenges.

# 2 Data Input Overview

The accuracy and reliability of the new hourly model relies on well-prepared, high-quality input data. The model requires two key datasets: an hourly temperature time series and an hourly electricity consumption time series, the latter serving as the target variable. These inputs must meet specific criteria in terms of granularity, coverage, and quality to ensure robust predictions.

## 2.1 Data Granularity and Coverage

The model operates at an hourly resolution, meaning all input data must be recorded hourly. Data coverage is equally important, with the following criteria:

- The model's framework requires a baseline dataset of up to 365 days, covering at least 90% of each month's data to capture long-term patterns such as seasonal trends.
- Each day can have up to 50% missing data, but no more than 6 missing hours consecutively. For each meter, days with more missing data compared to these thresholds will be flagged and excluded from the model training. The rest of the meter data will be used to fit or predict with the model, provided that it still meets the sufficiency criteria above.
- Gas data must never go negative
- Any zeros in electricity data are viewed as errors and changed to NaN

## 2.2 Data Preparation and Cleaning

To manage these requirements, a custom-built **Hourly data class** was developed to validate and preprocess the input data. Error messages are generated if any failures occur so that minimal effort is required by the user to check their data. This class performs several critical tasks, including:

1. Validation of Inputs: Ensures that both hourly temperature and electricity consumption time series are present and that they meet all sufficiency requirements.
2. Timestamp Continuity: Identifies and fills any gaps in the datetime index to maintain a continuous time series.
3. Handling Missing Data: Uses a correlation-based interpolation method to estimate missing values, provided the missing data falls within acceptable thresholds (e.g., fewer than 6 consecutive hours missing per day). Interpolated values are flagged to ensure users can locate such data points. Additional information can be found below.
4. Solar Irradiance Data Integration: When available, hourly solar irradiance data is incorporated into the feature space as an additional input. This inclusion enhances the model's accuracy for meters with solar PV systems by capturing the variability in solar energy production.

5. Warn of potential data issues such as extreme outliers in the data

## 2.3 Correlation-based interpolation

For each input time series features, i.e. temperature, and the electricity consumption time series, we interpolate all missing values. The methodology is based on autocorrelation of each feature/target. We use autocorrelation over the time series to find the *N* largest peaks (where *N* is heuristically defined) to get the lag/lead index for each feature (i.e., the lookahead/lookback period on which the data is the most self-similar). The lag/lead values at each missing time step for each feature are averaged to replace the missing value. This methodology ensures that we capture the pattern of the time series to replace the missing value. Figure 1 (lower) illustrates the performance of the interpolation method for 10% missing data for a single meter's data. This methodology has been tested up to 50% missing data to ensure that any future data sufficiency changes will not be limited by the interpolation methodology.

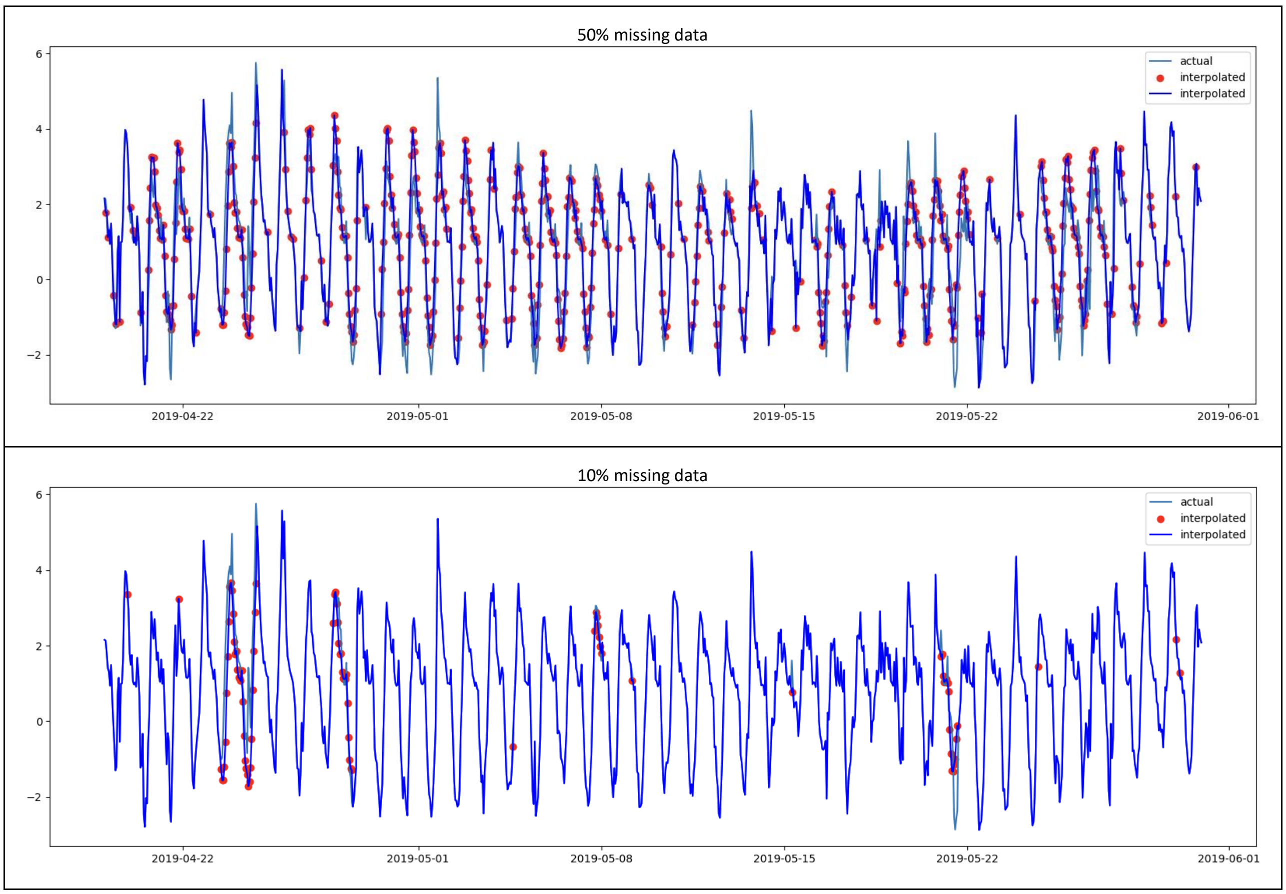


**Figure 1.** Example plots of energy usage data with upper panel) 50% data removed and lower panel) 10% of data removed and used as ground truth for the correlation-based interpolation methodology.

By enforcing these constraints and systematically addressing data quality issues, the model can be sure to have a clean and reliable dataset. This robust preparation process enhances predictive performance and ensures flexibility for incorporating additional data in the future.

# 3 Model Overview

## 3.1 Feature Generation

Feature generation is a crucial aspect of the new hourly model, ensuring that the input data is transformed into meaningful features that capture both linear and non-linear relationships among the input variables and energy consumption. The model incorporates various types of features, including time series, categorical variables, and their combinations to predict hourly electricity consumption with high accuracy.

### 3.1.1 Temperature

Building energy consumption generally has a strong response to temperature which varies considerably in different temperature regimes as the need to heat and/or cool the building changes. Hourly temperature is therefore a required input for each day. To better capture the variation in building energy consumption behavior in different temperature ranges, especially at temperature extremes, temperature values are divided into bins, each of which is assigned an independent slope and intercept in the model. There are two categories of temperature bins in the model:

- Low and High Temperature Bins: Energy usage behavior can change dramatically in these areas. For example, an AC could be turned off, continue operating as normal, struggle to keep its setpoint, or hit maximum capacity which would all change the rate of response to temperature. To account for this, the lowest 4.25% and highest 4.25% of hourly temperature values are modeled with a non-linear term in addition to the linear slope and intercept. For these bins, exponential growth and decay rates are included to capture deviation from linear behavior, where supported by the data.
- Intermediate Temperature Bins: Temperatures between the low and high bins are spaced linearly, using increments of 12°F to create additional bins, each of which is assigned an independent slope and intercept representing the response of energy consumption to temperature in that specific temperature bin.

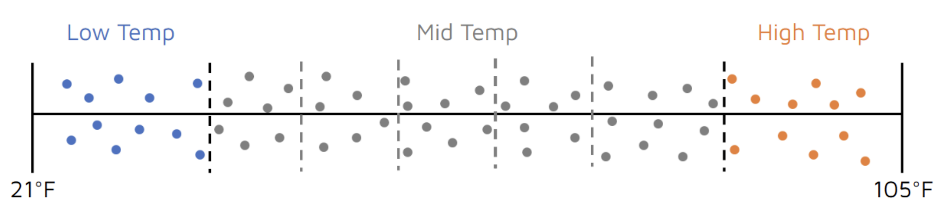


**Figure 2.** Temperature binning schematic showing how intermediate temperatures and extreme temperatures are binned

The figure above shows an example of hourly temperature for a sample meter. All of the temperature bins have a linear impact on the energy consumption modeling (slope and intercept). The way that they are each given a unique intercept is to create features for each bin such that every data point is a True/False flag (read as 0 and 1). When the Elastic Net applies a coefficient, this becomes either 0 or 1*coefficient which is an intercept. Likewise unique slopes are created by multiplying the temperature by the intercept column for each temperature bin. Since this coefficient is multiplied by the actual temperature, it acts as a slope.

### 3.1.2 Non-Linear Temperature Features

The lowest and highest temperature bins have extra components in the features space to account for nonlinearities. To model the nonlinear behavior a function of the form below is used where *A* is set such that the function is 1 at the extremes of the temperature bins, *s* is positive and negative for two different features, *T* is the standardized temperature time series, and *k* is the rate determined heuristically.

$$f(T) = A\left(\exp\left(\frac{s}{k}T\right) - 1\right)$$

An example of the non-linear features is shown below.

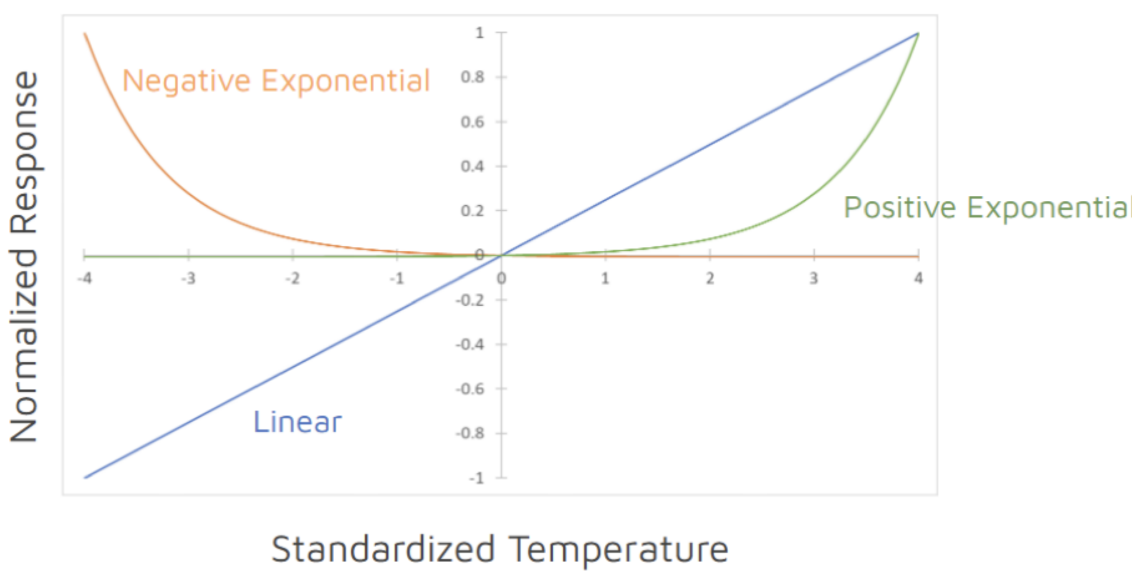


**Figure 3.** Example of response of positive exponential (green curve) and negative exponential (orange) compared to a linear response (blue).

In Figure 3, a linear response (blue) is shown vs standardized temperature for comparison. The two additional non-linear features are shown (orange and green). In the extreme temperature bins, all of these components are used to estimate the energy usage with respect to temperature. To do this each of components has an additional coefficient multiplied to them and a net intercept. This means that the scale and sign affect the response of energy usage and all three of these responses provide a very flexible set of features to capture linear and non-linear behavior if the data supports it.

### 3.1.3 Temporal Clusters

The energy consumption of an individual building depends on its occupancy and utilization, which tend to vary on a regular schedule. For instance, offices may be occupied during the week and unoccupied during weekends, while residences may have an opposite pattern. The legacy hourly model made a rough estimate of occupancy as a pre-fitting step and includes time week as a variable to address this; however, this approach tends to be coarser than is desirable to capture the nuances of building utilization throughout the year. Another approach involves allowing different models on weekdays versus weekends, or by season, as in the OpenEEmeter 4.0 daily model. However, some buildings (e.g., theaters and restaurants) may have usage variation that does not follow a strict weekday/weekend pattern. The detailed information available in hourly data allows us to account for this variation in a more flexible way by identifying *clusters* of similar daily load shape patterns throughout the year. In this way, if a particular building has distinct operational patterns on, e.g., Mondays and Tuesdays in winter compared to the rest of the year, the model can capture these differences.

In order to do that we use day of the week and month of the week patterns as the initial guess and industry standard for electricity usage pattern behavior segmentation. Then for each of the 84 combinations of these patterns, (Monday, January), (Tuesday, January), …, (Sunday, December), we take the median usage of each segment. We call these daily median patterns as the representative of each segment. Below is the clustering methodology:

Clustering steps and algorithm:

- A discrete wavelet transform with 5 levels using a Haar mother wavelet is used to transform the time series data out of the temporal domain and into a series of wavelet coefficients.
- The coefficients are then input into a principal component analysis (PCA) to reduce the dimension of the wavelet coefficient to capture the most important parameter coefficients. PCA coefficients are included up to 72.5% minimum variance ratio explained.
- Next, a bisecting k-means clustering algorithm is used to cluster the PCA coefficients for 2 to 24 clusters.
- The optimal number of clusters is selected using the Variance Ratio Criterion.

Results show that the majority of meters cluster into 2–3 groups, significantly reducing complexity. Examples below show the effectiveness of the clustering algorithm. For instance, in Figure 4, the initial guess of 84 distinct categories is shrunk down to two electricity usage patterns. A more complicated clustering result is shown in Figure 5, illustrates 3 clusters exhibiting a unique combination of seasonality, afternoon peak, and solar generation magnitude behavior found specifically for this meter.

The temporal clusters are turned into features using methodology similar to temperature binning. Unique intercepts and slopes for each temporal cluster are the coefficients that are multiplied by the binary indicators of cluster membership and interaction between the binary cluster membership and temperature, respectively.

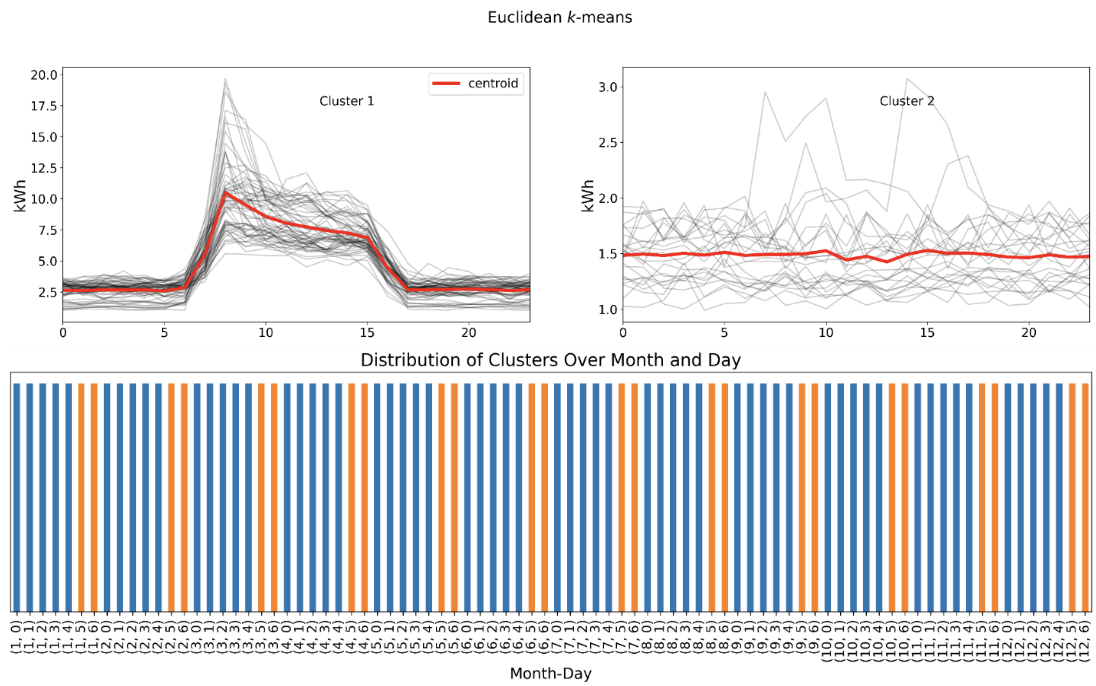


**Figure 4.** An example of temporal clustering where cluster 1 (blue) represents the weekdays and cluster 2 (orange) the weekend.

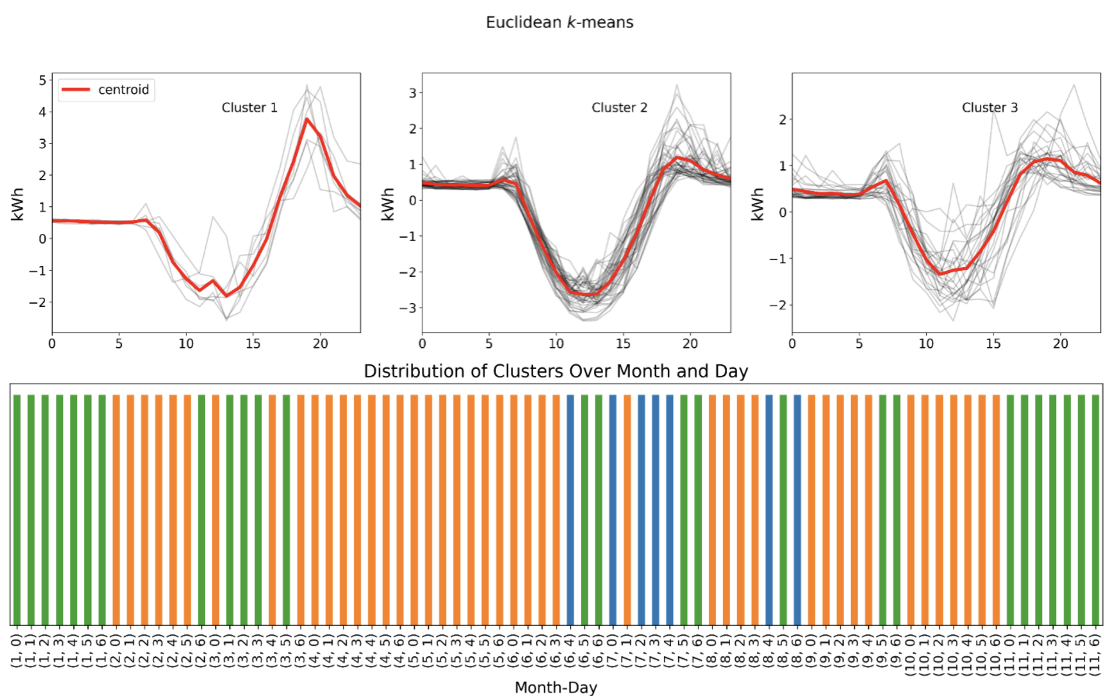


**Figure 5.** An example of temporal clustering where cluster 1 (blue) shows unique summer usage, cluster 2 (orange) shows mostly moderate seasonal usage, and cluster 3 (green) shows primarily winter usage with interspersed days throughout the rest of the year.

### 3.1.4 Solar Irradiance

Global Horizontal Irradiance (GHI) is used as a time series input when available to inform the new model of solar irradiance. GHI provides a direct proxy for solar generation, which is approximately linearly proportional to PV production. It is currently up to the user to procure their own GHI data. This can be done with free, but delayed, sources such as NREL or paid through a service such as Solcast and should be location-specific either by weather station or directly over the meter location. The model allows a linear response to GHI that is the same across all hours of the year.

### 3.1.5 Supplemental Data

Supplemental data is a research and development input which is supported but should not be used for OpenEEmeter compliant measurements. This is the case because it's impossible to know what a potential user might input as a supplemental feature. If certain features are negotiated and agreed upon by all parties then it is reasonable to be included in measurements, but should not be referred to as OpenEEmeter compliant.

Nevertheless, optional supplemental datasets can be provided by the user, such as pump schedules or EV charging schedules. These features are treated as linear inputs when complete and available. These datasets help refine predictions for specific use cases by incorporating operational patterns into the feature set. It should be mentioned that supplemental data can be either time series, which should be available for all of the timesteps without missing values, or it can be used as categorical variables which need to be available for all of the days in the dataset.

As an example, the hourly electricity consumption of a meter during the baseline period is shown in Figure 6. A change point in electricity consumption can be observed around November 11, 2019, attributed to the installation of solar PV.

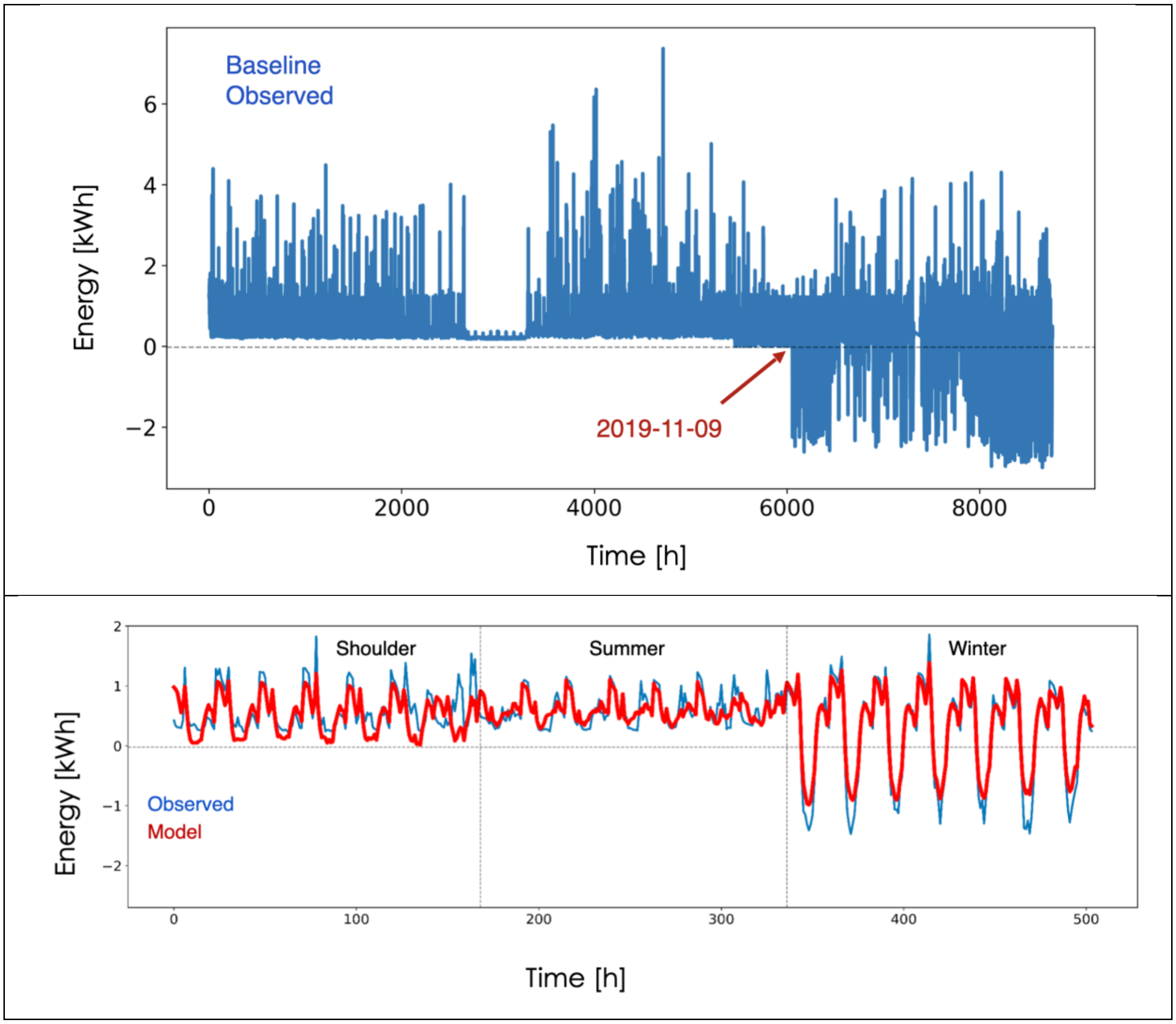


**Figure 6.** An energy consumption plot of a building likely undergoing a solar PV installation in the baseline period. The upper panel) depicts the observed time series and lower panel) shows the corresponding seasonal hour-of-week loadshapes for the model fit and the observed data.

In this example, the solar PV installation date was used as a categorical variable to train the model. The categorical value is assigned as 0 for days before the installation date and 1 for days after it. This simple yet important information is shown to significantly improve the model's performance, as demonstrated in **Figure 7**.

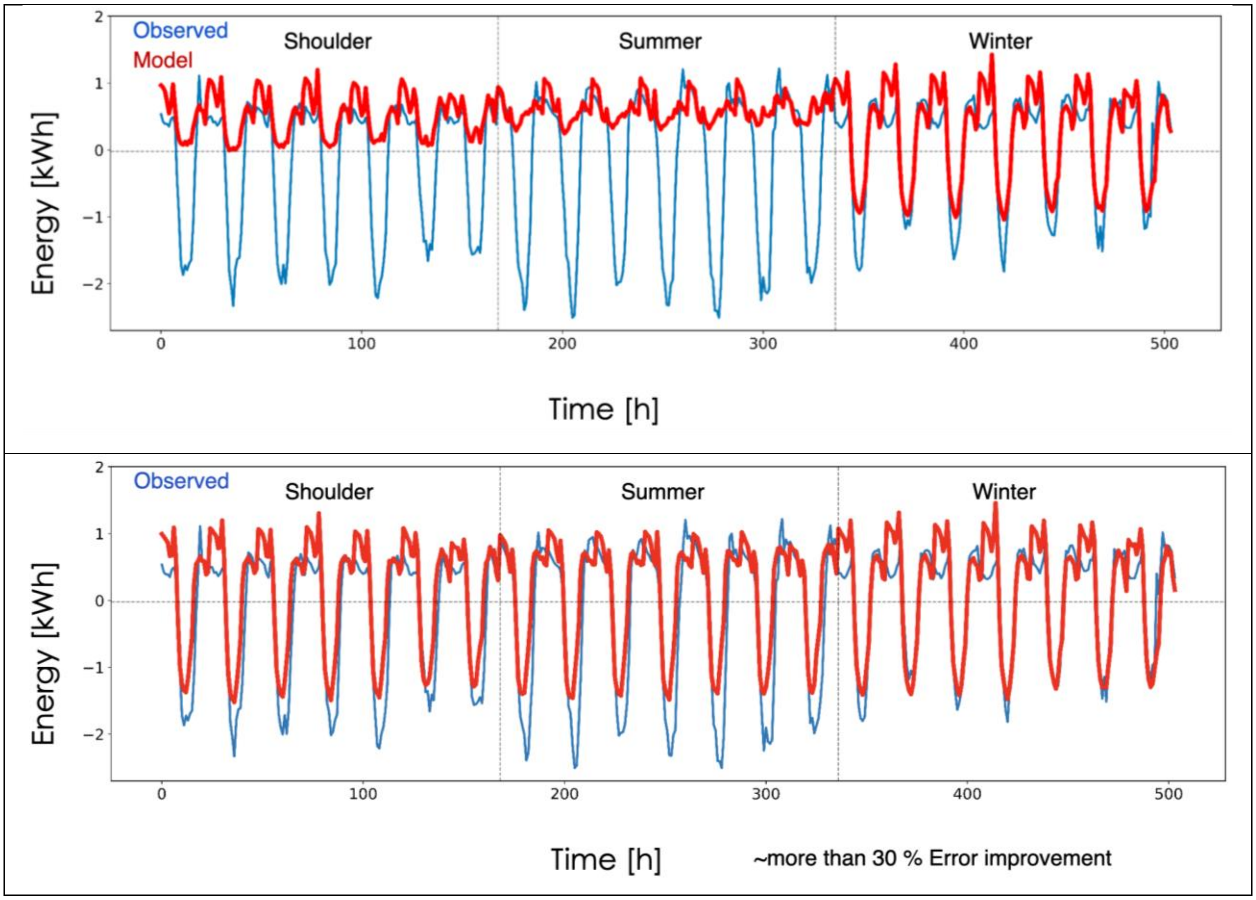


**Figure 7.** An example of how model performance can be significantly enhanced from the standard model upper panel) without utilizing the solar installation date compared to the lower panel) with solar installation date.

### 3.1.6 Total Number of Features

To help aid in understanding how the model ingests data the following equation has been given to calculate the number of features for a given fit:

$$2\,[T_{bins} + T_{nonlinear} + t_{clusters}] + 24\,GHI + N_{supplemental}$$

where $T_{bins}$ is the number of temperature bins, $T_{nonlinear}$ is the number of nonlinear temperature features for the low and high temperature bins, $t_{clusters}$ is the number of temporal clusters, $GHI$ is either 0 or 1 if GHI is included or not, and $N_{supplemental}$ is the number of supplemental columns provided. The number of features in the model varies depending on the specific characteristics of each meter. Since temperature binning and clustering are optimized individually for each location, the number of temperature bins and clusters are unique to the meter being analyzed. Furthermore, if solar irradiance (GHI) or other supplemental data is available, the feature count adjusts accordingly. This adaptive approach ensures that the model remains tailored to the specific data and characteristics of each meter, further enhancing its predictive performance.

## 3.2 Model Framework

### 3.2.1 Building the 24-Hour Input-Output Framework

To enhance the model's ability to capture intricate lead-lag relationships between temperature, solar data, and electricity consumption, we implemented a 24-hour input-output segmentation framework. Historically, M&V models have predicted electricity consumption per hour without considering correlations among different hours. To improve on this, the new approach allows the model to process a full day's data as input and output, enabling it to automatically recognize and model correlations across different time lags for each meter. The elastic net model gets a feature vector ($1 \times N$) as an input and uses a matrix ($N \times 24$), rather than a vector, to map the input feature to 24 hours output (electricity usage prediction). This matrix of coefficients allows the model to learn the lead/lag correlation between the input data and the output data. For instance, electricity consumption for hour 11 can be impacted by the temperature of all of the hours in the day based on the behavior of the meter in the baseline data. This framework improves prediction accuracy and computational efficiency by leveraging matrix multiplication within Elastic Net, streamlining the fitting process.

### 3.2.2 Elastic Net

In developing the new hourly model, we chose Elastic Net regression as the core modeling framework after an extensive literature review of energy consumption M&V techniques. During the review, we evaluated a range of models, from simple linear regression to decision tree-based models all the way to advanced approaches like neural networks and transformers. Simple models, such as linear regression, were deemed inadequate due to their inability for feature selection and tendency to overfit on the baseline data when all desirable features are included. While decision tree-based models could better handle nonlinearities, they often struggled with overfitting in large datasets and performed poorly when needing to extrapolate. Advanced models like neural networks and transformers were highly capable but came with a significant computational burden, making them impractical for use across millions of meters.

Elastic Net offered the perfect balance between simplicity, interpretability, and computational efficiency. Its ability to be coerced to handle both linear and nonlinear relationships in data made it a highly effective choice for modeling energy consumption. Furthermore, its computational efficiency ensured that it could scale to large datasets without becoming prohibitively resource-intensive.

Elastic Net is a regularized regression method that combines the strengths of L1 (Lasso) and L2 (Ridge) regularization. This dual regularization approach offers several advantages:

1. Feature Selection and Sparsity: The L1 penalty encourages sparsity in the model by shrinking some coefficients to zero, effectively performing feature selection. This is especially useful when dealing with a large number of features, as it helps focus on the most relevant predictors.
2. Handling Multicollinearity: The L2 penalty helps stabilize coefficient estimates in the presence of multicollinearity, where predictors are highly correlated. By blending L1 and L2, Elastic Net balances sparsity with robustness.

3. Flexibility: Elastic Net allows for a mix of linear and nonlinear modeling by adjusting its regularization parameters, making it adaptable to various types of data.
4. Computational Efficiency: Elastic Net is computationally efficient, particularly when applied to large datasets. Its matrix-based optimization methods are well-suited for high-dimensional data, making it scalable for applications involving millions of meters.
5. Interpretability: Despite its ability to model complex relationships, Elastic Net remains interpretable, a key requirement for ensuring trust in energy consumption models.

These advantages, combined with its versatility and scalability, make Elastic Net an ideal framework for the new hourly model.

## 3.3 Hyper-Parameter Optimization

To fine-tune the new hourly model, we conducted extensive hyperparameter optimization using data from over 35,000 residential and commercial meters. This dataset included both non-solar and solar meters, ensuring the optimization was robust and representative of diverse real-world scenarios. The key hyperparameters optimized during this process included the Elastic Net parameters, **alpha** and the **L1 ratio**, **temperature bin width, percent of temperature data to be placed in the extreme temperature bins**.

Elastic Net Hyperparameters:

1. Alpha: This parameter controls the overall strength of the regularization applied in the model. Higher values of alpha increase the regularization effect, shrinking coefficients more aggressively, which can help prevent overfitting but may reduce the model's ability to capture complex patterns.
2. L1 Ratio: This parameter defines the balance between L1 regularization (Lasso) and L2 regularization (Ridge).

The optimization of these hyperparameters was conducted at a population level, ensuring that the values selected were optimal at a population level across both residential and commercial sectors. These hyperparameters serve as the foundational tuning parameters for the model and remain constant across all meters. While the hyperparameters are consistent across the population, the primary model coefficients and parameters, such as the weights assigned to individual features, are unique to each meter. These coefficients are determined based on the baseline data and input features for each meter, allowing the model to adapt to specific usage patterns and environmental factors.

# 4 Disqualification Criteria

## 4.1 Normalized Error Metrics

Historically CVRMSE and NMBE have been the normalized error metrics by which a meter's model is judged. For reference these equations are presented below as Eqns. 1 and 2.

$$CVRMSE = \frac{\sqrt{\overline{\epsilon^2}}}{\overline{obs}} \tag{1}$$

$$NMBE = \frac{\overline{\epsilon}}{\overline{obs}} \tag{2}$$

where $\epsilon$ is the error between the observed and model values, $obs$ is the observed values, and a bar above a variable represents the mean of that variable. These normalized metrics have been adequate until meters begin to have mean observed values which are negative or close to zero. In the event that the observed values are negative, one could modify the denominator of both Eqns. 1 and 2 to have an absolute value. However, if the mean observed is negative, then the meter must have some kind of behind-the-meter generation such as solar PV generation. At least at a residential level, solar PV systems are commonly designed such that the house is approximately net zero annually. This is where we run into the biggest problem with CVRMSE and NMBE; an observed mean which is close to zero will cause the error metric to blow up. At this point the error metric is no longer useful and meters will be disqualified even if they are predictive with low error.

To address both issues of a negative and a near zero denominator, an alternative has been developed such that it will be applicable for meters with behind-the-meter generation as well as those without. Several choices were considered such as only averaging positive signals, averaging nighttime signals, or using peak load; however, it was decided to instead utilize the difference between percentiles. Percentiles are useful because they will always be positive because the larger percentile will always be larger than the smaller percentile. It is possible for the percentiles to be the same, but in such an event the meter itself would have very little change in its energy consumption.

A range of percentiles were considered but fortunately normalizing the RMSE by the 25% and 75% quantiles were found to correlate best with CVRMSE. These percentile normalized metrics are referred to as PNRMSE and PNMBE and are defined in Eqns. 3 and 4.

$$PNRMSE = \frac{\sqrt{\overline{\epsilon^2}}}{obs_{Q_3} - obs_{Q_1}} \tag{3}$$

$$PNMBE = \frac{\overline{\epsilon}}{obs_{Q_3} - obs_{Q_1}} \tag{4}$$

where $obs_{Q_3}$ is the third quartile and $obs_{Q_1}$ is the first quartile. To see how well PNRMSE correlates with CVRMSE, 1000 non-solar residential customers' error values are compared in Figure 8. It is common to think about CVRMSE as a measure of error relative to how much energy a meter uses. PNRMSE can be thought of in a similar manner except that it is normalized by the variation in a meter's usage.

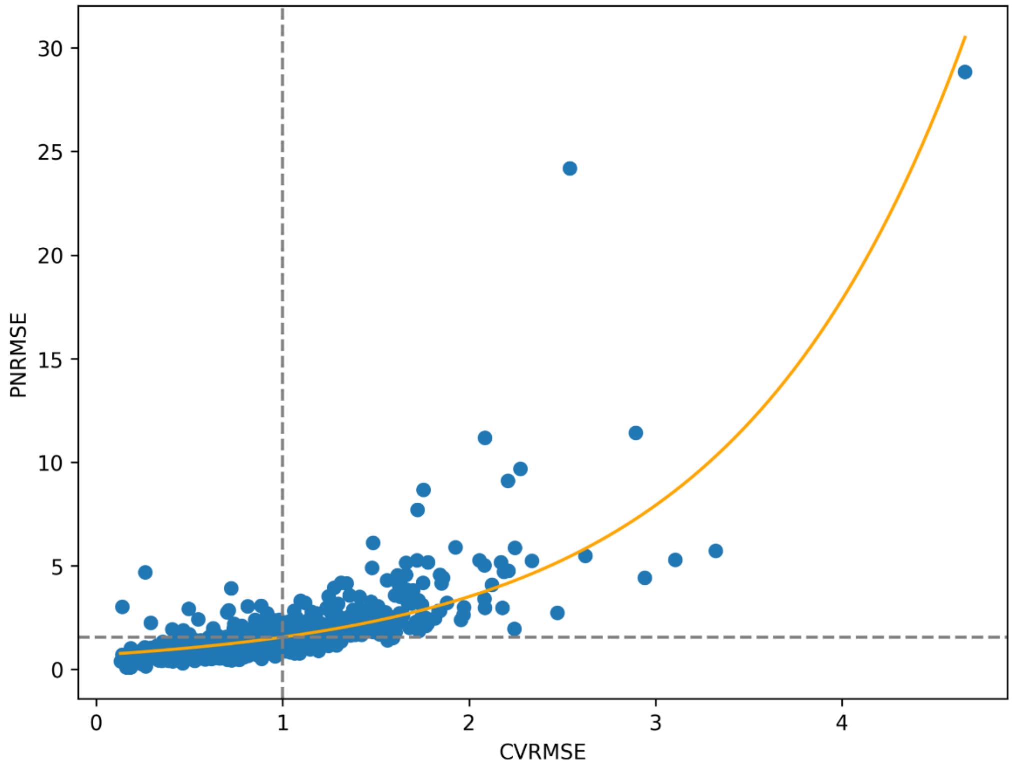


**Figure 8.** Correlation between PNRMSE and CVRMSE for 1000 non-solar meters. Each data point is represented in blue and a best fit in orange. Dashed lines depict the intersection of CVRMSE at 1 with PNRMSE at 1.6 utilizing the fit curve.

Since PNRMSE and CVRMSE are fairly well correlated, it might be reasonable to consider replacing CVRMSE with PNRMSE; however, we are not proposing such a drastic change. At the same time, a CVRMSE threshold of 1.0 is far too conservative for hourly data which shows significantly more variation than daily data. We therefore propose that the CVRMSE threshold for hourly models be 1.4 and PNRMSE be a corresponding 2.2. This value of 1.4 was determined by looking at 1000's of meters which were deemed acceptable by the daily model and looking at their hourly CVRMSE.

There are cases where it might be reasonable to judge a model by CVRMSE or by PNRMSE, but we would prefer to have universal model sufficiency criteria because it is not always clear that generation is occurring. To this end, we propose that a meter's model should be considered sufficient if it would meet either sufficiency criteria. In other words, a meter's model is sufficient for measurement purposes if $CVRMSE < 1.4$ *or* $PNRMSE < 2.2$.

# 5 Population Results

## 5.1 In-Sample Results (mostly)

These results show data from the test data as described in Section 3.3 plus several thousand out-of-sample meters from the same dataset. All of these meters are non-participant meters hence the expectation is that the overall change from baseline to reporting year will be approximately zero. In the interest of a fair comparison, only reporting year results will be shown. This means that overfitting will not affect these results as it is known that the legacy hourly model is more overfit than the new model.

To ensure that the new hourly model is as accurate as the OpenEEmeter 4.0 hourly model comparisons are shown in **Figure 9** and **Figure 10** as histograms of percent improvement over OpenEEmeter 4.0.

**Figure 9** shows a CVRMSE improvement distribution for both the solar and non-solar models of approximately 2% and a slightly positive but negligible CVMBE improvement. **Figure 10** shows a negative CVRMSE improvement of -0.5% if looking at the median or -2% if using the mean. The change in CVMBE is also negligible. The residential customers are roughly normally distributed whereas the commercial customers have a long tail in the negative improvement. The residential customers distribution shows a peak of ~2%. The commercial customers have a mode around zero, but the long tail and skew does mean that the new model is performing slightly worse overall. Putting all of this information together leads us to the conclusion that the new hourly model is performing about as well as OpenEEmeter 4.0's hourly model with some minor variation depending upon residential and commercial customers. This indicates that we have achieved our goal of ensuring that the new model does perform as well on non-solar customers as the prior model.

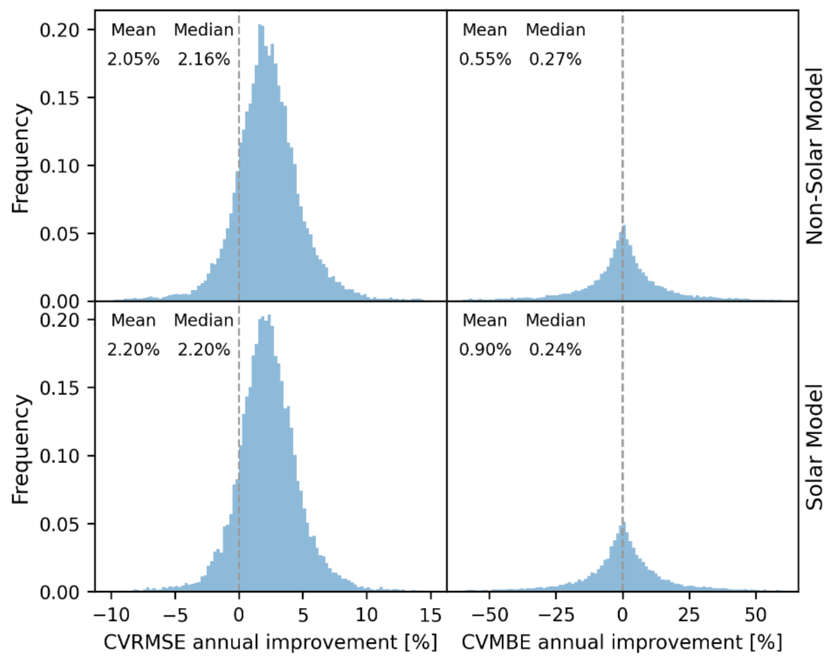


**Figure 9.** Histogram of the new model's improvement over the OpenEEmeter 4.0 hourly model on non-solar residential meters for both CVRMSE and CVMBE (left and right panels) as well as comparisons for both the non-solar model and the solar model (upper and lower panels). A dashed vertical line represents matching the OpenEEmeter 4.0 hourly model.

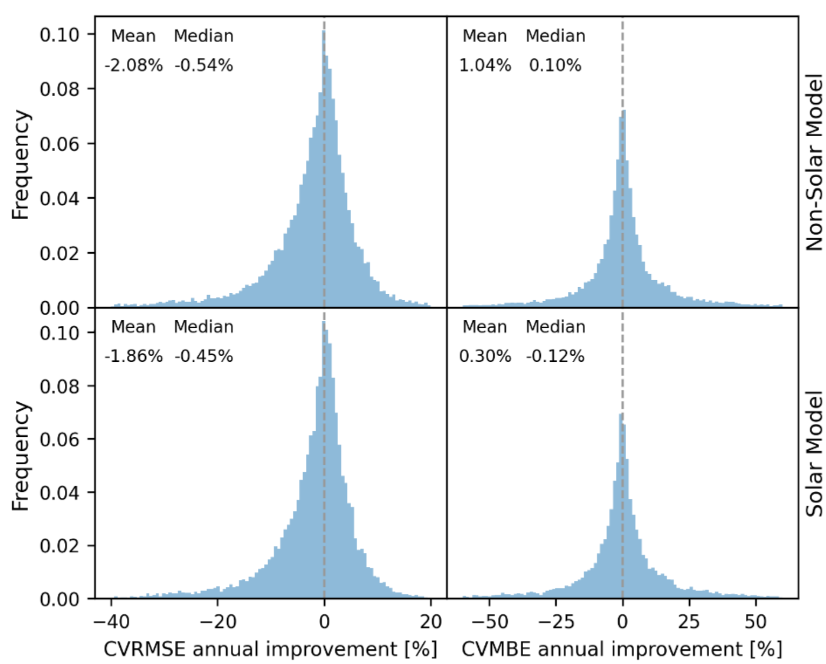


**Figure 10.** Histogram of the new model's improvement over the OpenEEmeter 4.0 hourly model on non-solar commercial meters for both CVRMSE and CVMBE (left and right panels) as well as comparisons for both the non-solar model and the solar model (upper and lower panels). A dashed vertical line represents matching the OpenEEmeter 4.0 hourly model.

The next question is to ask if the goal of improving predictions on meters with solar PV production has been accomplished. To address this **Figure 11** and **Figure 12** show comparisons between the new solar model and the prior model. The bins in both figures have equal sample sizes which leads to more data points to the left side of the plots with increasingly separated data points towards the right because there are fewer cloudy days than sunny days. In the upper panels showing PNRMSE, the ideal result would be for the new model to be lower than the prior model. In the lower panels showing PNMBE, the better model will be closer to the dashed line at 0.

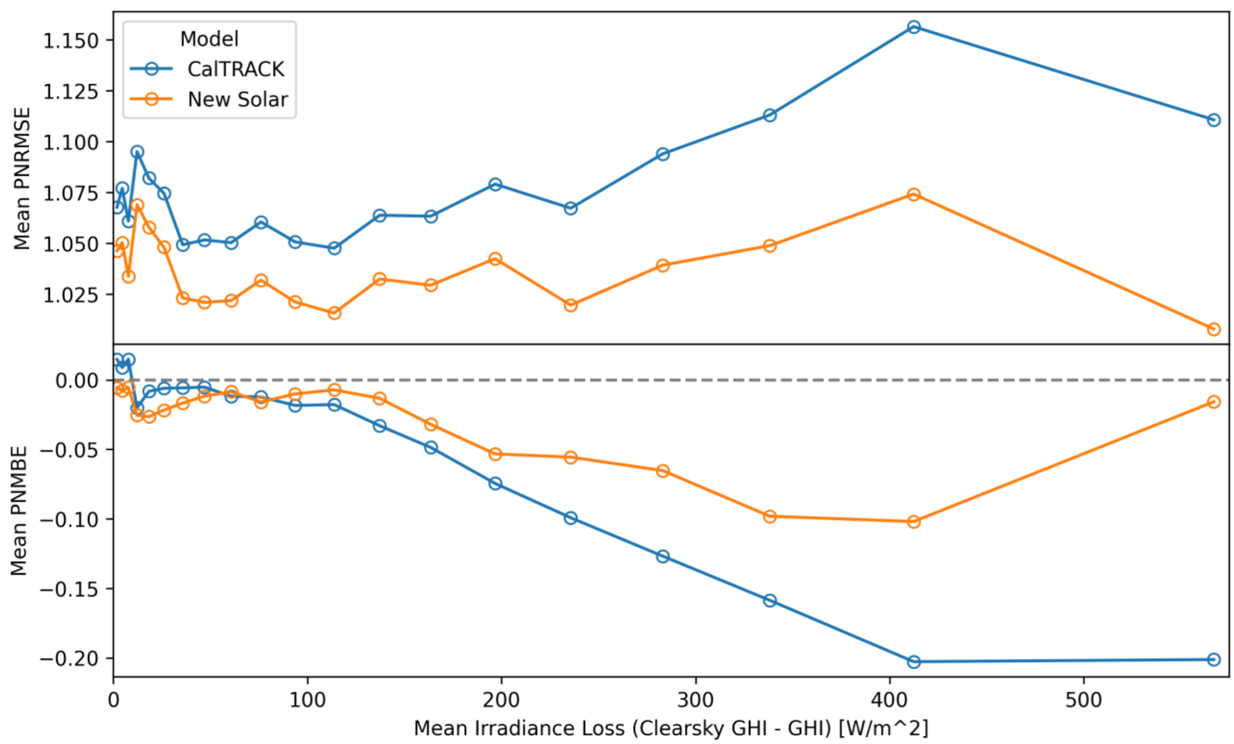


**Figure 11.** A comparison of solar residential meters mean upper panel) PNRMSE and lower panel) PNMBE vs the average irradiance loss for bins with equal sample sizes. The x-axis can be viewed as how cloudy it is with 0 being a sunny day with no clouds and 500 being a very overcast day.

In the upper panel of **Figure 11** showing residential meters' PNRMSE, the new solar model is performing better than the prior model in every bin. Comparing the two models' PNMBE in the lower panel gives a slightly more muddled conclusion. Here the new solar model is performing better in very sunny days, but there is a small segment of 5 points in which the prior model is performing better; however, overall, the new model is performing significantly better. We see similar results for commercial meters.

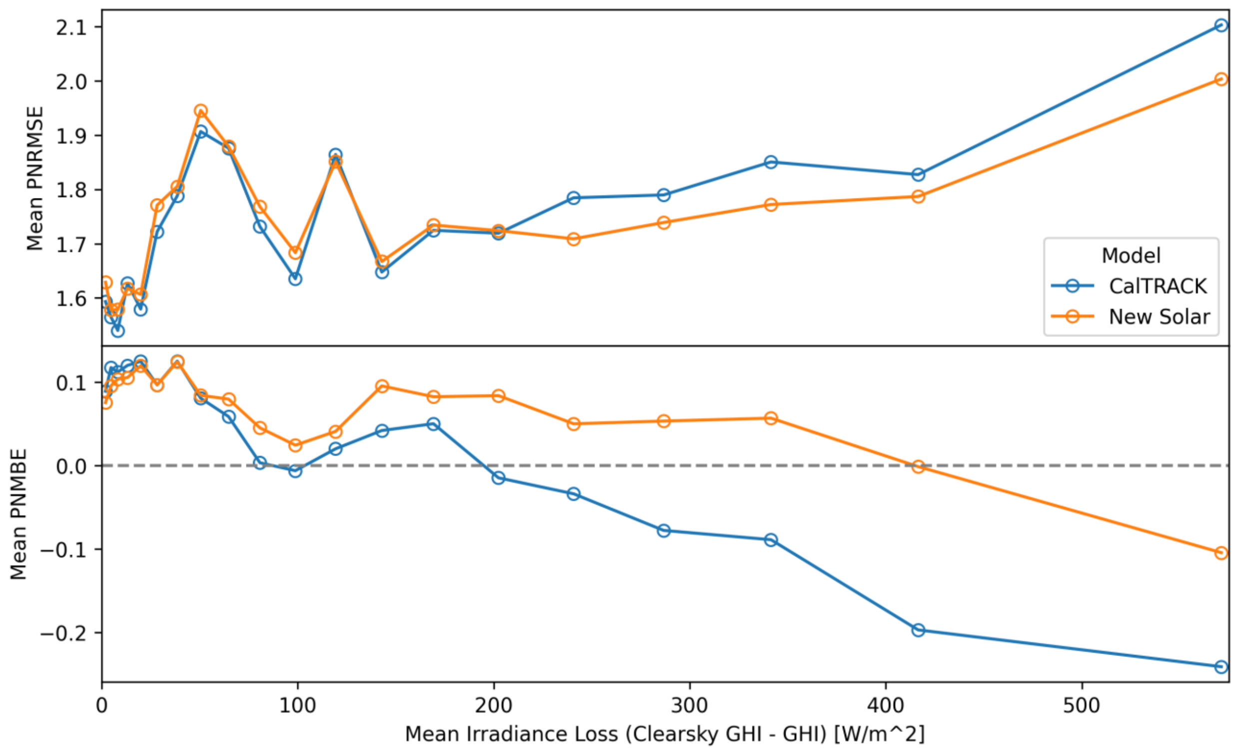


**Figure 12.** A comparison of solar commercial meters mean upper panel) PNRMSE and lower panel) PNMBE vs the average irradiance loss for bins with equal sample sizes. The x-axis can be viewed as how cloudy it is with 0 being a sunny day with no clouds and 500 being a very overcast day.

Commercial meters' results show improvement, but less significant than residential meters. In the upper panel of **Figure 12**, both the new solar model and the prior model largely overlap when it is sunny, but the new model does differentiate itself when it becomes cloudier. A similar result to the residential meters is seen in the lower panel depicting PNMBE, where the new solar model is better on very sunny days, and then the prior model is better in the mid region before becoming worse on cloudy days. The conclusion that we reach with all of this is that the new model is performing better overall, but it's not a complete improvement. There are clearly some tradeoffs being made, but it is a net improvement.

# 6 Conclusion

Based on the extensive analysis conducted at the population level, the new hourly model has demonstrated improvements in both solar and non-solar meter predictions. For solar meters, the model showed a ~3% improvement in accuracy, with the major gains attributed to the inclusion of solar irradiance data (GHI). This enhancement is particularly noticeable on cloudy days (~12% improvement in accuracy), where the GHI data helped the model more accurately account for reduced solar generation. Interestingly, even for non-solar meters, the inclusion of GHI data resulted in an accuracy improvement of around 1.5%. These findings indicate that the new hourly model can be effectively used across various meter types, residential, commercial, solar, and non-solar, provided GHI data is available, making it a versatile solution for different scenarios.

In addition to its predictive accuracy, the new hourly model is three **times** faster than legacy models, offering a significant computational speed advantage. The model's flexibility is another standout feature, as it easily accommodates the addition of new time series or categorical supplemental data, allowing for continuous improvements and customization as new data becomes available.