Key Elements of Reliable and Complete Population NMEC Measurements: Disqualification, Assigned Savings, and Interpolation

8/26/2025
Updated 6/17/2026

I. Summary

Since its creation in 2020, the *Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption* (NMEC Rulebook)[1] has provided an essential framework to enable Population NMEC programs. With the NMEC Rulebook guidance in place, Population NMEC has evolved from a niche set of pilot programs to a primary deployment mechanism for flexible and accountable demand side initiatives.[2]

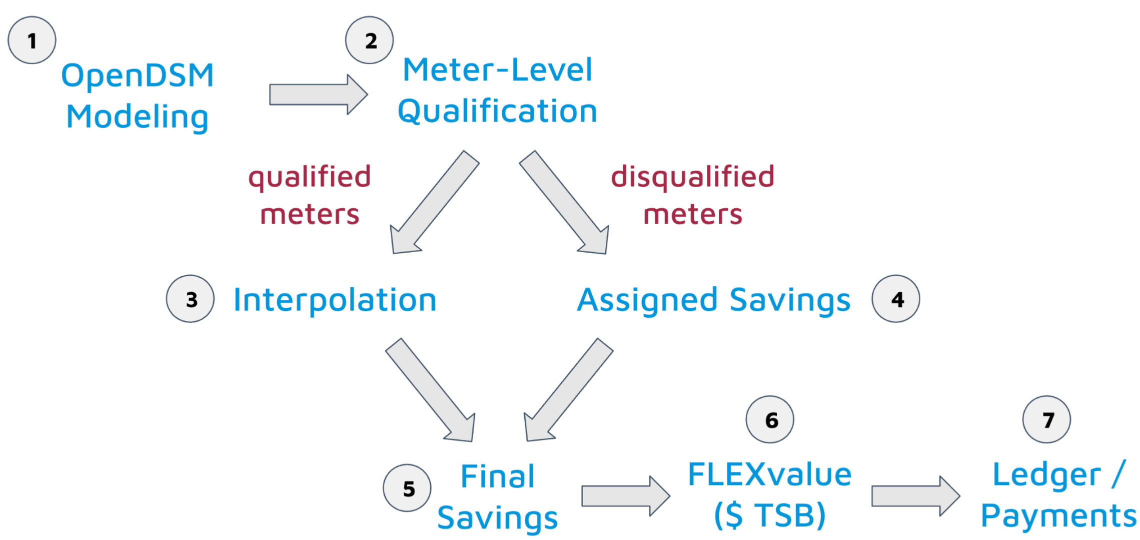
While the NMEC Rulebook establishes many fundamental guardrails, it does not fully contemplate several important issues that arise in most programs. Here we cover three such related issues and how Recurve's products address them:

1. **Disqualification** of meters with counterfactuals that cannot be trusted to serve as a reasonable estimation of consumption in the absence of the program
2. **Assigned Savings** calculated for the meters that are disqualified from a measurement
3. **Interpolation** of absent savings values that arise from missing consumption or weather data

Together, these steps ensure the quality of a population NMEC measurement, while also providing proper credit to all installed projects.

The following schematic outlines the relationship between these topics and the sequence that procedures are applied in Recurve's software.

Recurve Software Computational Pipeline:



[1] *Rulebook for Programs and Projects Based on Normalized Metered Energy Consumption (version 2.0)*, California Public Utilities Commission, 2020.

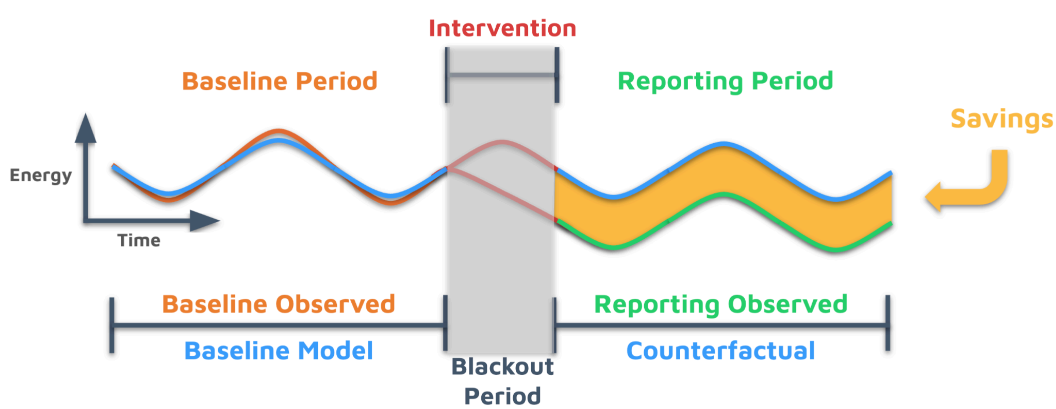
[2] California's Market Access initiative and the measured pathway of the Inflation Reduction Act's Hope for Homes Program are clear examples of increasing reliance on Population NMEC programs as energy resources on par with supply side options.

After data ingestion and formatting, Recurve's computational pipeline begins with the creation of baseline models using the open-source, Linux Foundation Energy OpenDSM package.[3] The modeling outputs of the OpenDSM step are needed to determine which meters are qualified and which should be disqualified from the measurement due to their risk of creating instability or bias in the population level results. This disqualification process is described in Section II. Disqualified meters are then taken into an assigned savings pathway (Section III). Qualified meters enter an interpolation procedure, detailed in Section IV, which fills in the missing savings values that may arise due to absent meter or weather data.

Once the interpolation and assigned savings steps are completed, a full set of time-series energy savings values is assembled for all projects. These data flow into the open source FLEXvalue engine to determine the Total Systems Benefits (TSB). The TSB are the basis for aggregator payments, which is determined in the last step of the process.

II. Disqualification

The schematic below shows the elements of a meter-level OpenDSM load impact calculation.[4] Based on a meter's pre-program consumption data and weather data, a "baseline period" model is created. After the program intervention, that model is projected forward as a counterfactual using the "reporting period" weather. The counterfactual represents the model's prediction of what energy consumption would have been if the program intervention had not taken place. Often, this counterfactual is "corrected" based on an analogous set of calculations applied to a matched comparison group.[5] The difference between the counterfactual and the reporting period observed usage is taken as the energy savings.



For stable and valid population NMEC results, it is critical to ensure that the individual meters subject to this measurement can be properly modeled and the predictions those models generate constitute a reasonable comparison to the observed reporting period consumption.

---

[3] https://lfenergy.org/projects/opendsm/

[4] Meter level results are aggregated to the population level in a Population NMEC measurement

[5] See the open source GRIDmeter methods (*Comparison Groups For the COVID Era and Beyond*) and code (https://github.com/recurve-methods/comparison_groups)

To this end, disqualification criteria should be established that remove meters when the associated counterfactuals should not be trusted as the basis of a savings calculation. Disqualification criteria often include data sufficiency requirements,[6] model fit metrics, outlier thresholds, and screening for "Non-Routine Events," such as the installation of a major new load that is not associated with the program, among others. In Recurve's experience, when applying a complete set of disqualification criteria, attrition of 20% is common in a population NMEC measurement, though the specific number will depend on many factors. With this said, programs that successfully screen for baseline period disqualification triggers in an enrollment eligibility step can retain a much higher fraction of meters in the final measurement.

Below are the disqualification criteria Recurve typically applies along with brief explanations and common default values. These default values can be modified depending on the context and needs of a program.

Long Term Baseline (AKA Energy Efficiency or Project-based)

A. Analysis Period Disqualification Criteria

The analysis period is defined as the start of the baseline period through the end of the reporting period. The following criteria are not tied specifically to the baseline period, the blackout period, or the reporting period.

- **Modified load:** Meters can be disqualified based on the addition or subtraction of a major new load during the analysis period that is not associated with the program. For example, the counterfactual would no longer provide a valid comparison to observed consumption after a customer installs rooftop solar PV during the reporting period. While some forms of modified load can be detected using meter data alone, others benefit from knowledge captured by program implementers or administrators.
- **Dual participation:** Meters should be disqualified if the customer has participated in multiple programs that impact long-term counterfactuals or reporting-period observed consumption during the analysis period.

B. Baseline Period Disqualification Criteria

The following criteria are applied to the baseline period (the 365 day period before the program intervention):

[6] Often, the predominant reason meters are disqualified is insufficient baseline period data. The OpenEEmeter (precursor to OpenDSM) working group deliberated the topic of data sufficiency and arrived at qualification criteria. When using the OpenEEmeter methods, the threshold for missing data is 10% of a year for the 2.0 daily model. In other words, 90% of days in a 365-day baseline period must contain non-zero values. For monthly or billing data, this often effectively translates to a maximum of one month or billing cycle with missing data allowed. For the CalTRACK 2.0 hourly model, the working group established that 90% of usage data must be present in each month of the year to form a valid model. These existing data sufficiency requirements were adopted for the new OpenDSM hourly model..

- **Baseline data sufficiency:** Meters must have enough baseline period data to build reliable models. Baseline data sufficiency criteria are often distinct for monthly, daily, and hourly interval data. The following defaults, based on the OpenDSM sufficiency criteria, apply to observed meter data, temperature data, and, in the case of the solar hourly model, solar irradiance data, as well as to the presence of all types of data jointly:
  - *Monthly*: Meters must have 11 months with all expected data
  - *Daily*: Meters must have at least 90% of days (> 327 days) with all expected data
  - *Hourly*: Meters must have at least 90% of days with at least 90% of expected data, and 90% of hours within each calendar month must have all expected data.
- **Baseline model fit:** Meters must have consumption patterns that allow for the generation of a model with a reasonable fit. OpenDSM computes a model's Coefficient of Variation of Root Mean Squared Error (CVRMSE), its Percentile Normalized Root Mean Squared Error (PNRMSE), defined as the RMSE divided by the interquartile range of a meter's interval consumption, among other metrics, and disqualifies meters exceeding certain thresholds in these values[7]. Recurve's default threshold values match the OpenDSM criteria.
  - *Hourly*: OpenDSM hourly model CVRMSE between 0 and 1.4 OR PNRMSE < 2.2.
  - *Daily/billing*: OpenDSM daily/billing model CVRMSE between 0 and 1.0 OR PNRMSE < 1.6
  - Legacy models prior to OpenDSM 1.0 disqualify meters with a *daily* model CVRMSE > 1.0 for all time granularities.
- **Comparison group selection:** If comparison group correction is being used, meters must have adequate comparison pool matches available with sufficient data and model fit to allow construction of a comparison group.
- **Modeling breaking points**: Certain computational issues can arise that result in broken models. For example, a meter may have a pattern of missing data that prevents a model from being formulated.

C. Reporting Period Disqualification Criteria

The following criteria are applied to the reporting period (the 365 day period after the blackout end date or beyond):

- **Reporting year 1 data sufficiency:** These are analogous requirements to those established for the baseline period. However, this criterion must be assessed on an ongoing basis as the first reporting year progresses.

[7] See the OpenDSM documentation for more information: https://opendsm.energy/stable/documentation/

- **Savings outliers:** Very high or very low savings relative to the counterfactual often indicates that a building's consumption has changed in ways that are unassociated with a program. Default: Reporting year 1 savings (savings/counterfactual) should be within the bounds of -0.5 to 0.5.[8] Outlier thresholds can and should be revisited based on statistical analysis of the program being measured or similar past programs.
    - In some cases, after review, it may be determined that a savings outlier is actually showing reasonable or consistent savings in respect to the intervention. In this case, the meter may remain qualified.
- **Consumption outlier:** In a mature program it is risky if any single meter accounts for a very high fraction of the total consumption of all customers in the program. When portfolios reach 100 projects it is recommended to disqualify any meter that accounts for more than 10% of total portfolio counterfactual, though this setting can depend on the context of a specific program.

Event-based (AKA Demand Response)

Event-based measurement uses slightly different terminology for the periods involved in constructing a measurement, reflecting conceptual differences in how the measurement is constructed. In place of the "baseline period" above, we refer to the "training period", and instead of the "reporting period" we refer to the "measurement period".

A. Training Period Disqualification Criteria

The following criteria are applied to the model training period :

- **Training data sufficiency:** The standard training period to measure an event is 45 days prior to the week in which the event is called. The following criteria apply to the data in this period.
    - At least 35 days within the 45-day training period must contain data for 90% or more of expected readings for each of meter, temperature, and solar data
    - Each hour of the week must have meter data from at least three different days
    - The meter data must have sufficient variability: at least 10% of the data points must have unique values
    - If a behind-the-meter solar or battery installation date falls during the training period for a given meter, all data prior to this date will be treated as missing for the purposes of data sufficiency and model training
- **Model fit sufficiency:** Meters must have consumption patterns that allow for the generation of a model with a reasonable fit. Recurve's platform computes the CVRMSE and PNRMSE of each model and applies the following model fit sufficiency criteria, which are rooted in the OpenDSM hourly model criteria:
    - CVRMSE between 0 and 1.4 OR
    - PNRMSE less than 2.2

[8] For electrification programs, a building's gas usage may be removed all together and electric consumption will increase substantially. Therefore, corresponding outlier ranges of fractional savings for gas should be closer to -0.5 to 1 and -1 to 0.5 for electric.

- **Comparison group selection:** Each event participant must have a sufficiently large set of non-participant comparison-group meters that meet all of the training period requirements and the measurement period data sufficiency requirements. Participant meters for which a suitable comparison group cannot be constructed are disqualified.

B. Measurement Period Disqualification Criteria

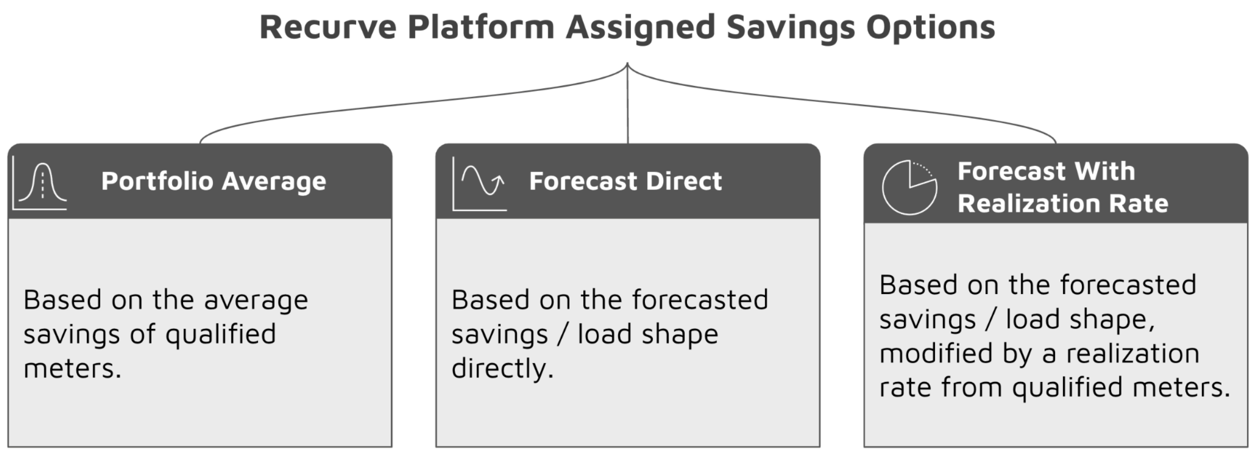
- **Measurement Period Data Sufficiency:** The measurement period consists of the 24 hours ending at midnight on the day the event was called. Within this period, the following data sufficiency requirements must be met.
    - All event intervals must have valid and complete meter, temperature, and solar data
    - Non-event intervals must contain data for 90% or more of expected readings for meter, temperature, and solar data, as well as the joint dataset including all three sources
    - If a behind-the-meter solar or battery installation date falls during the measurement period for a given meter, all data subsequent to this date will be treated as missing for the purposes of data sufficiency and measurement.
- **Dual Participation:** Meters that are dispatched in more than one event on the same calendar day (e.g., due to dual program enrollment) will be disqualified from measurement, and the reason will be documented.
- **Impacts Outlier:** Meters with load impacts that are very large outliers (fractional load impacts more than 3x the interquartile range above the 75th or below the 25th percentile of the dispatched population) are disqualified.

III. Assigned Savings

While it is important to disqualify meters from a savings calculation that could compromise the accuracy or trustworthiness of a population NMEC measurement, it is equally important that all completed projects are given due credit. Therefore, methods are needed to assign savings for disqualified meters. This section details three strategies Recurve has available for the assignment of savings to disqualified projects.

The three strategies are as follows:

1. **Portfolio Average:** Assign project savings based on the portfolio average
2. **Forecast Direct:** Assign project savings based on the project's forecast
3. **Forecast With Realization Rate**: Assign project savings based on the project's forecast multiplied by a portfolio-level realization rate determined from the qualified projects

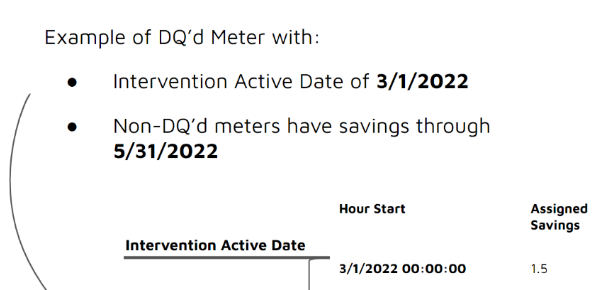


With any method, each meter is assigned to a single sub-program portfolio (often referred to as a canonical portfolio). This canonical portfolio can be used to further break down assigned savings calculations.

In the Portfolio Average method, the savings averages can be calculated at a full program level, or at an aggregator level. In addition, each meter is assigned to a single canonical portfolio. Canonical portfolio assignment is often performed with categorical variables such as program year, measure packages, or income status. This canonical portfolio can be used to further break down assigned savings calculations. The realization rates determined in the Forecast With Realization Rate method can also be broken down in these ways. This optionality ensures that one aggregator's results need not influence the outcomes for another aggregator and allows for distinctions between different pathways established within a program.

Each of the three strategies is most appropriate for particular sets of circumstances.

1. The Portfolio Average approach is best for mature portfolios with a largely homogenous participant base. For example, a residential portfolio with 1,000 projects would be expected to yield calculation of stable average savings results. However, this is *not* the best option when the program serves a wide variety of customers such as a non-residential portfolio with customers that span a wide range of total consumption. Assigning savings to a large office building with 100 MWh annual consumption based on the average from a portfolio that mostly serves small businesses would not be advisable.
2. The Forecast Direct strategy is best utilized for nascent portfolios that lack the volume of past projects needed to calculate reliable savings averages or realization rates.
3. The Forecast With Realization Rate option is best suited for portfolios with enough completed projects to calculate reliable realization rates, but with a heterogeneous customer base. Note, this method is NOT advisable for portfolios where a given fuel type could have negative savings, such as electrification, because



realization rates with negative forecasted savings are nonsensical.

In each of these methods, savings are assigned on a sliding window from the intervention active date through the last known date of savings from qualified meters. As a result, savings will be assigned continuously over time, rather than assigning a full year of savings at once. This enables continuous reporting of assigned savings, and opens the door for ongoing payments from assigned savings.

The first and third options utilize the meter-based performance to the extent possible, which is often an objective of population NMEC programs as doing so preserves the accountability of the measured results. Collectively, the three approaches cover the large majority of expected cases.

IV. Interpolation

Even among meters that pass all disqualification criteria, consumption or temperature data points can be missing, resulting in empty savings values. Recurve has studied several datasets and we observe that among meters with sufficient data, raw meter data and/or weather data are missing 0.5 - 3% of values, resulting in the corresponding savings values being eliminated.

While this may appear to be a small problem, for competitive programs that operate on the margins of cost effectiveness and aggregator profitability, the failure to capture all benefits can represent a significant impediment. Additionally, gas and electric data often have different degrees of completeness. Therefore, for electrification or fuel switching programs in particular, these imbalances can skew savings, avoided cost, and greenhouse gas impacts.

Recurve has designed, tested, optimized, and implemented a highly accurate interpolation algorithm to restore missing savings data points from these issues. This document gives two anonymous examples, describes the interpolation approach, and gives the results of testing.

A. Types of Missing Data

Missing savings values can arise for at least three distinct reasons:

1. Missing treatment meter observed data in the reporting period
2. Missing temperature data, which prevents computation of a counterfactual
3. Missing comparison group correction, which can arise from the elimination of outlier data points or missing comparison group meter observed data in the reporting period

B. Examples

Test Case 1:

The table below gives a summary of the savings data points from a recent dataset that Recurve analyzed.

Fuel/Interval	Savings Data Points	Count	% Total
Electricity/Hourly	Non-Null	262,042	97.89%
	Null Treatment Observed	4,510	1.68%
	Null Temperature	1,087	0.41%
	Null Corrected Counterfactual	38	0.01%
	Null Total	5,635	2.11%

In this case, more than 2% of all savings values were null due primarily to missing values in the reporting period consumption data.

Test Case 2:

The table below gives another breakdown of null savings values from a dataset Recurve recently analyzed. In this case, 0.47% of electric and 0.43% of gas savings values were null.

Fuel/Interval	Savings Data Points	Count	% Total
Electricity/Hourly	Non-Null	1,215,509	99.53%
	Null Treatment Observed	2,835	0.23%
	Null Temperature	2,772	0.23%
	Null Corrected Counterfactual	148	0.01%
	Null Total	5,755	0.47%
Gas/Daily	Non-Null	81,475	99.57%
	Null Treatment Observed	334	0.41%
	Null Temperature	5	0.01%
	Null Corrected Counterfactual	10	0.01%
	Null Total	349	0.43%

In this case, the null electric values were split fairly evenly between missing reporting period meter values and missing temperature values. Because the gas data were of a monthly interval, the missing temperature values were largely mitigated by taking average daily values per the OpenDSM method specifications.

C. Technical Challenge and Solution

Simple forms of interpolation are not appropriate for energy and savings data. For instance, filling in missing data with average annual values ignores seasonal patterns. Assuming a linear relationship between missing data points would fail to capture a customer's predictable monthly or hourly consumption patterns. Further, because missing savings values can arise due to missing temperature or missing observed data, it is insufficient to only interpolate temperature in a pre-modeling step.

An ideal solution would be simple, inexpensive, and equally applicable across all cases, from hourly electric to monthly gas and from a single missing data point to potentially a block of many missing points.

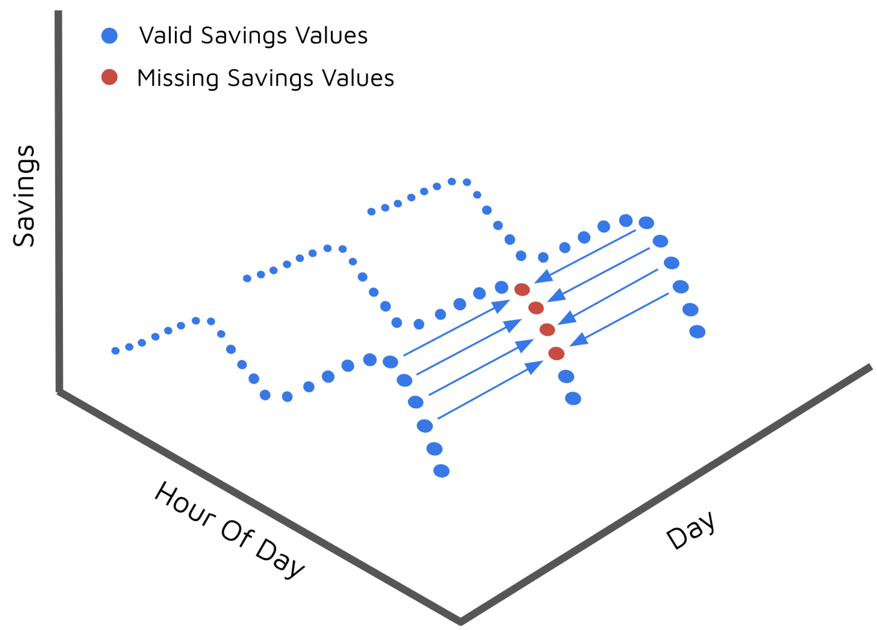
i. Core Concept: Partitioned Moving Average

The central element of the solution Recurve has developed is a partitioned moving average. Taking an hourly electric measurement as an example, a moving average is partitioned along the following axes:

- Meter
- Weekday vs Weekend
- If Hourly Savings: Hour of Day

Partitioning into these categories ensures that if a particular meter is missing a 6 pm weekday datapoint then only non-null 6 pm weekday measurements from the same meter will be used to create the interpolated value.

A simplified schematic of this interpolation concept is shown below.



Details of how to formulate the moving average (window size etc.) are the subject of empirical testing and are described below.

ii. Counterfactual, Observed, and Savings Interpolation

When both fields are present, savings are calculated as the difference between a meter's counterfactual value and its observed reading. Therefore, when one of these fields is missing, the resulting savings are also missing. Both observed and counterfactual values are expected to be temperature-dependent. Because the difference between observed and counterfactual ultimately matters, when one of them is missing it is advisable to interpolate both of them.

Interpolating only one would be prone to error on account of mismatching temperature dependencies among other unknown factors. Interpolating both values also ensures that no matter the root cause of a missing savings value, its repopulation is consistent and retains the relationship between counterfactual, observed, and savings.

D. Testing and Analysis

In order to test the accuracy and stability of the interpolation strategy, the following approach was formulated:

1. A seeded sample of N random points that had non-null savings values is taken from qualified meters within the test dataset.
2. For these N points, the *corrected counterfactual*, *observed*, and *corrected savings*[9] values were assigned as null.
3. For these N points, the original *corrected counterfactual*, *observed*, and *corrected savings* values were retained for residual analysis.
4. The interpolation algorithm (detailed below) is applied to generate interpolated *corrected counterfactual*, *observed*, and *corrected savings* values.
5. The interpolated values are compared to the original values in a statistical analysis.

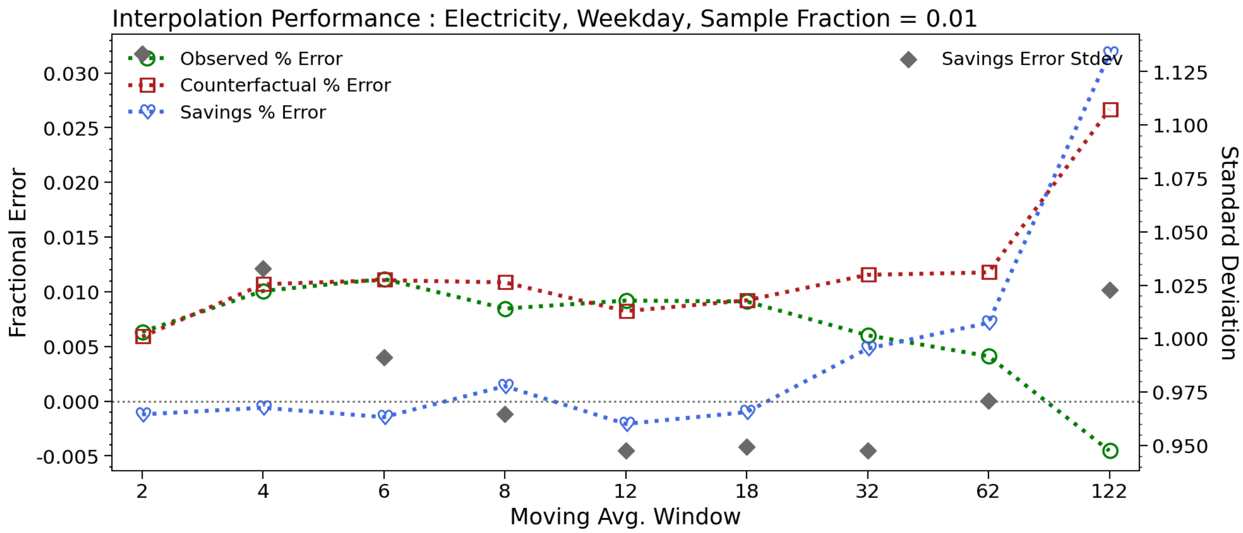
This analysis was conducted independently for electric meters and gas meters. Electric samples were isolated for weekends and weekdays. For both electric and gas, samples were chosen corresponding to scenarios of 1%, 10%, and 30% missing data. For each sample, the interpolation algorithm was run with moving average windows of 2, 4, 6, 8, 12, 18, 32, 62, and 122 points, with an equal number of points preceding the missing value and following the missing value.

For example, with a moving average window of 8 points, a missing electric value that occurs on a weekday at 3 pm will use the 4 closest non-null preceding weekday 3 pm points and the 4 closest non-null following weekday 3 pm points to construct the interpolated value.

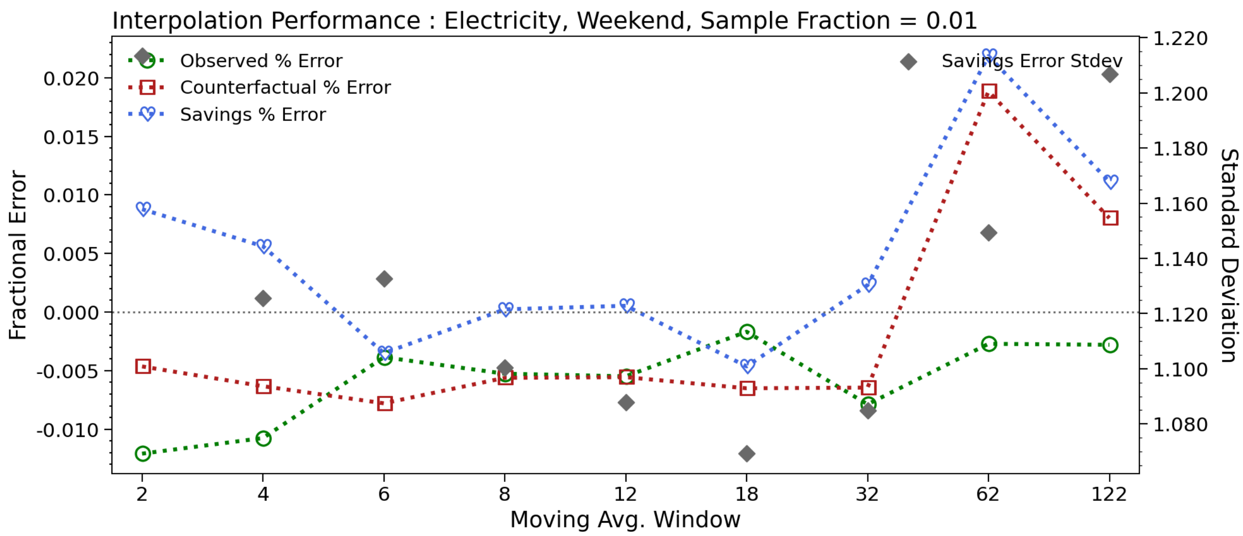
The figure below shows fractional error for interpolated counterfactual (red squares), observed (green circles), and savings (blue hearts) as a function of moving average window (x-axis) for the electric weekday sample. The gray diamonds show the standard deviation (kWh) of the savings error. At 8,723 points this sample represents 1% of the available dataset.

From moving average windows sizes of 2 - 18, the counterfactual and observed fractional error hover between 0.005 and 0.011, with the fractional savings error very close to 0 (-0.003 - 0.002). Fractional savings error is calculated as *sum*(*savings residual)/sum(original corrected counterfactual)*. With larger windows (32 - 122), a diversion is observed between counterfactual and observed error, leading to higher absolute error of the savings values.

[9] The term "corrected" indicates values have been calculated using a comparison group.

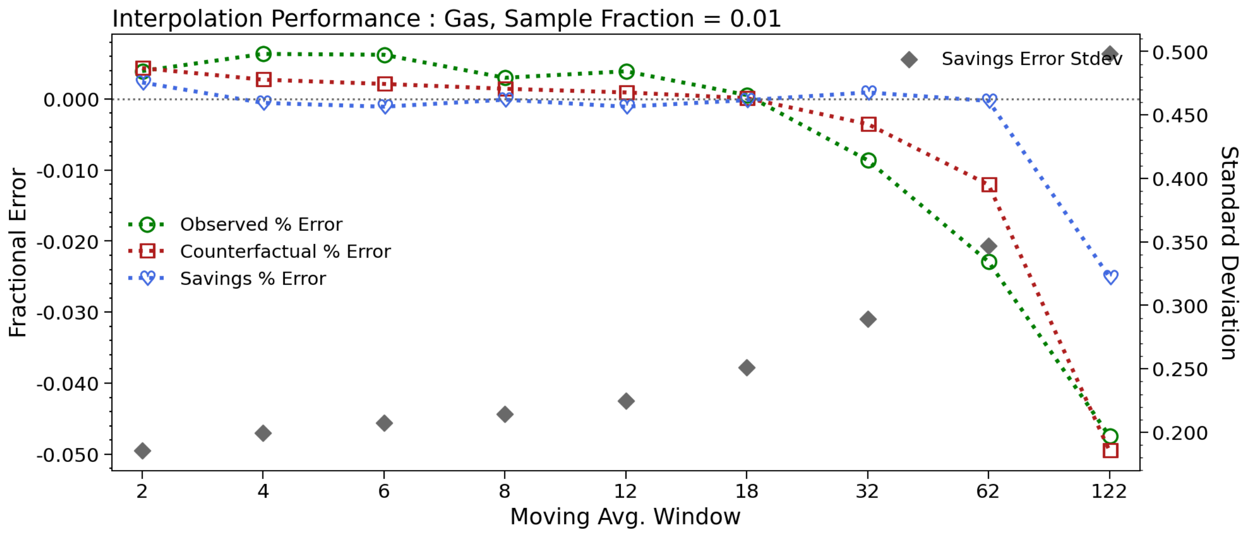


The next figure shows results for the electric weekend sample. This sample also represents 1% (3,489 points) of the available dataset. The lowest residual of the savings values is observed with moving average window sizes of 8 and 12 points. Again, with window sizes of 18 and under, fractional errors are all 0.012 or less.



The next figure gives results for the gas sample. This sample also represents 1% (818 points) of the available dataset.

Extremely low fractional error in the savings values is observed for all moving average window sizes except for the largest window of 122 points.



Analogous figures are given in the appendix for sample sizes that correspond to 10% and 30% of missing values. In these cases generally good performance is also observed among the smaller window sizes.

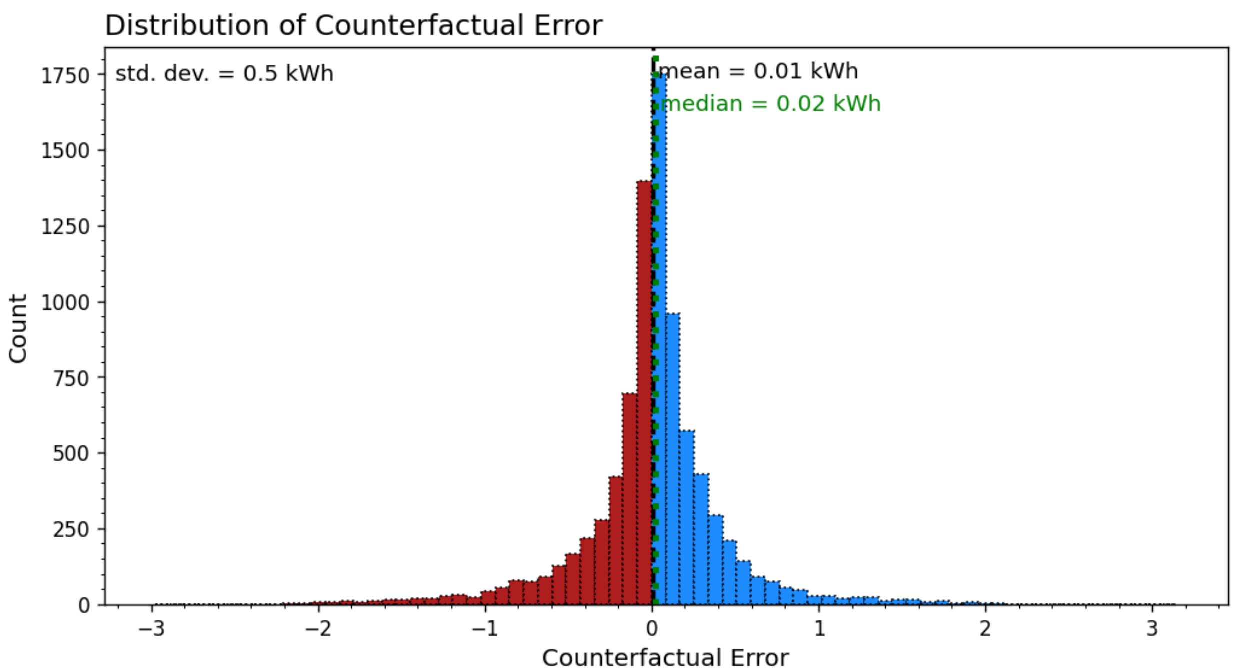
The table below summarizes results by moving average window size. Values in this table are computed from averages of the electric and gas 1% and 10% samples. (The 30% sample is extreme and was analyzed mainly to understand the limits of the algorithm in rare cases where a great deal of data is missing but a meter is still taken as qualified.)

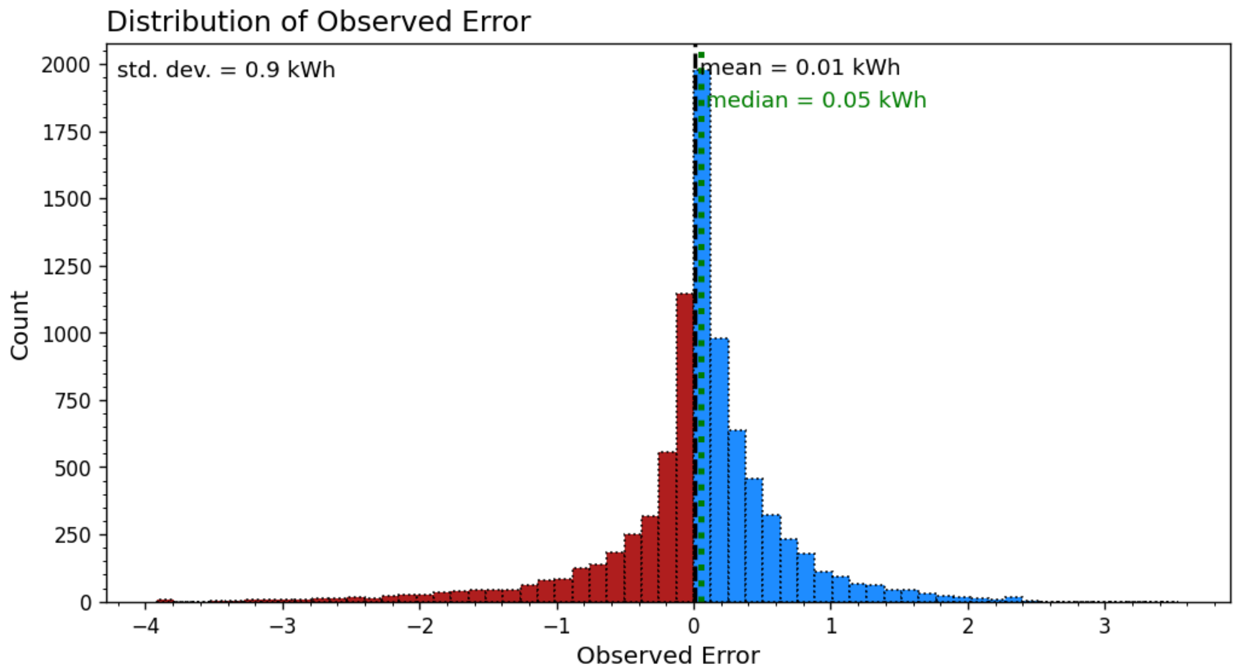
Fuel	Moving Avg Window	Avg. Abs Savings Error*	Avg. Median Savings Error (kWh/Hour or Therms/Day)
electricity	2	0.32%	0.01107
electricity	4	0.18%	0.02113
electricity	6	0.16%	0.02955
electricity	8	0.11%	0.03075
electricity	12	0.19%	0.03692
electricity	18	0.26%	0.04155
electricity	32	0.49%	0.04510
electricity	62	1.37%	0.04712
electricity	122	1.66%	0.05381
gas	2	0.31%	0.00000
gas	4	0.12%	0.00029
gas	6	0.16%	0.00053
gas	8	0.10%	0.00027
gas	12	0.17%	0.00024
gas	18	0.15%	0.00088
gas	32	0.14%	0.00117
gas	62	0.24%	0.00239
gas	122	3.00%	0.00610

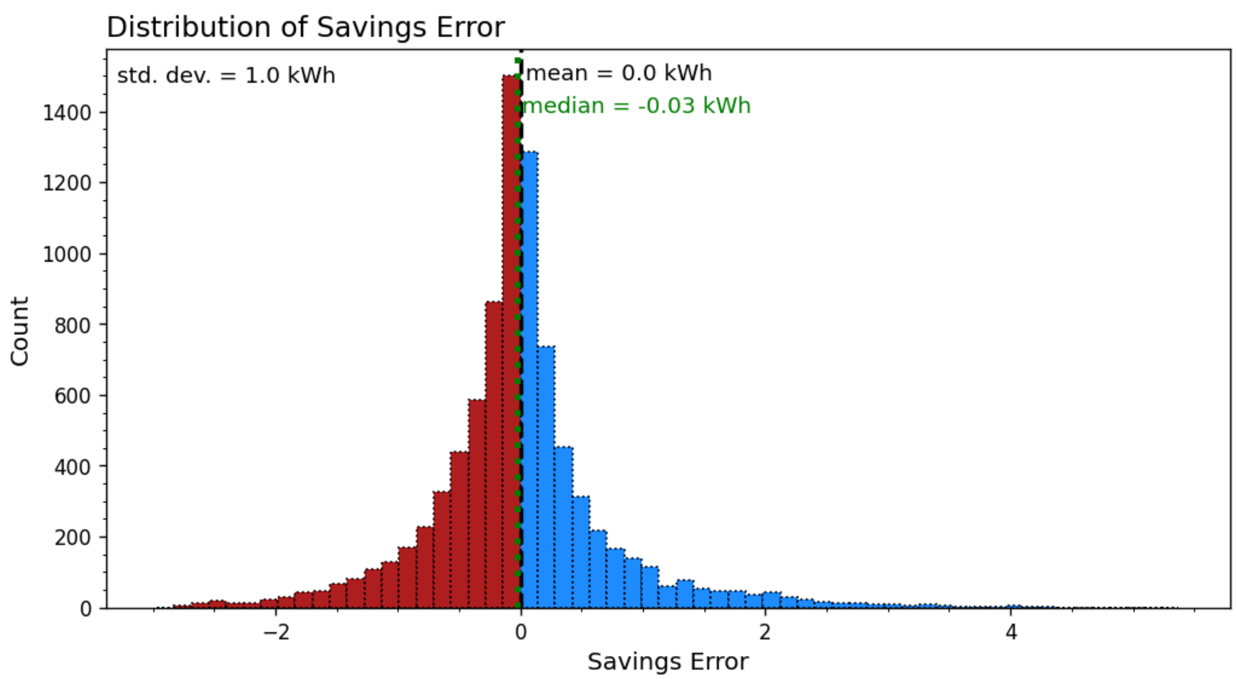
*Savings Error / Actual Corrected Counterfactual

In both the electric and gas samples the mean of the absolute savings error is lowest at a moving average window size of 8 points. The table also shows that absolute values of median savings error tend to be lower for smaller window sizes. This behavior may reflect the influence of data points far from the mean, which are more likely to influence the moving average as the window size grows. This is consistent with the observation that as window size grows, a higher degree of skew is observed in the distributions of observed residuals (see figures below). As long as mean error values are stable and near zero, the algorithm can be considered successful.

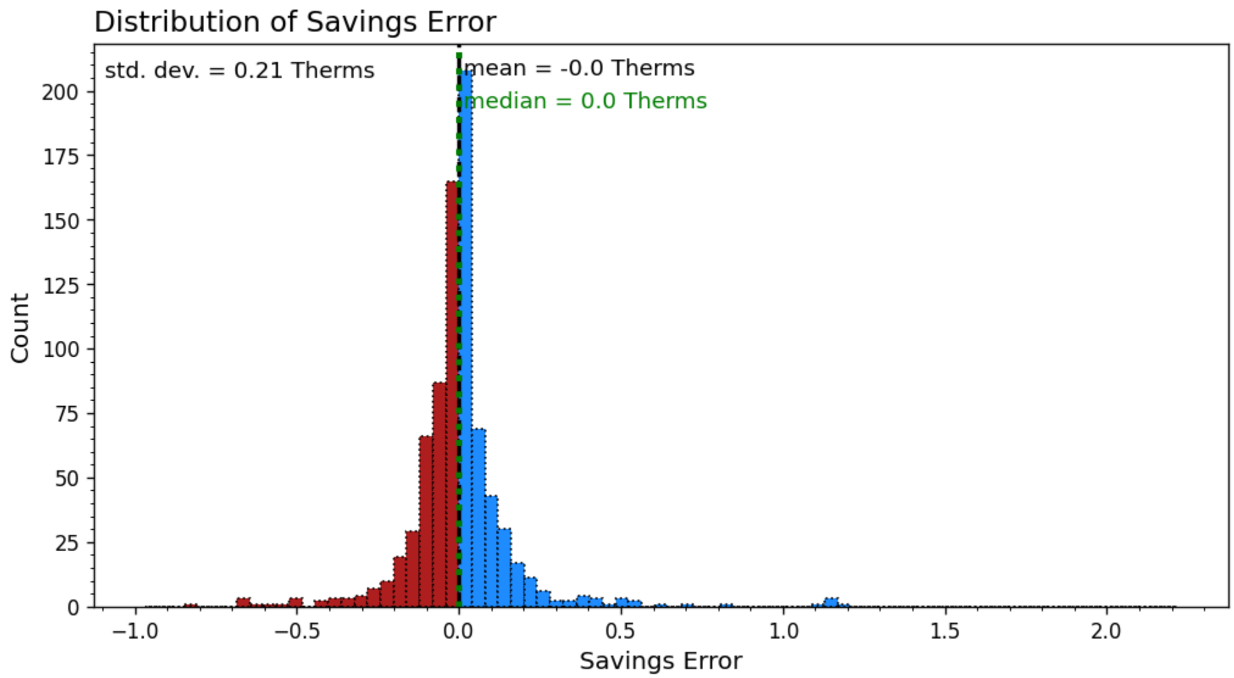
The following three figures show the distribution of counterfactual, observed, and savings residuals that result from applying the interpolation method in the 1% electricity weekday sample. These distributions are highly symmetric and centered around zero.



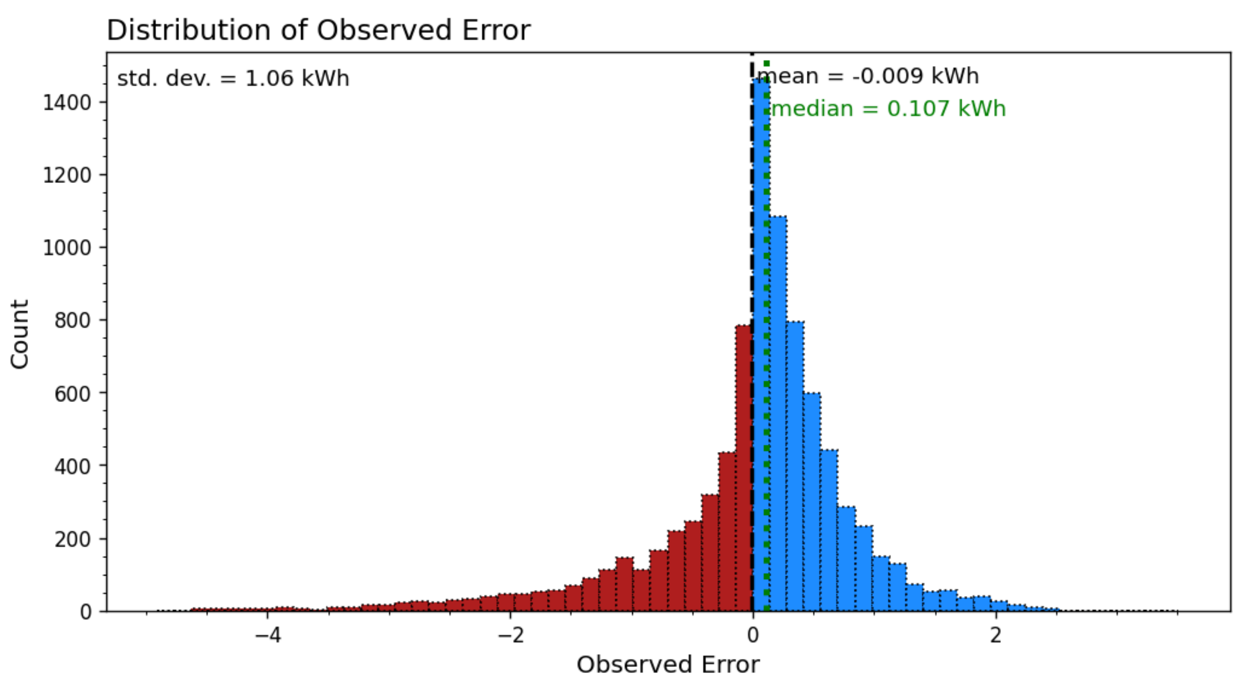




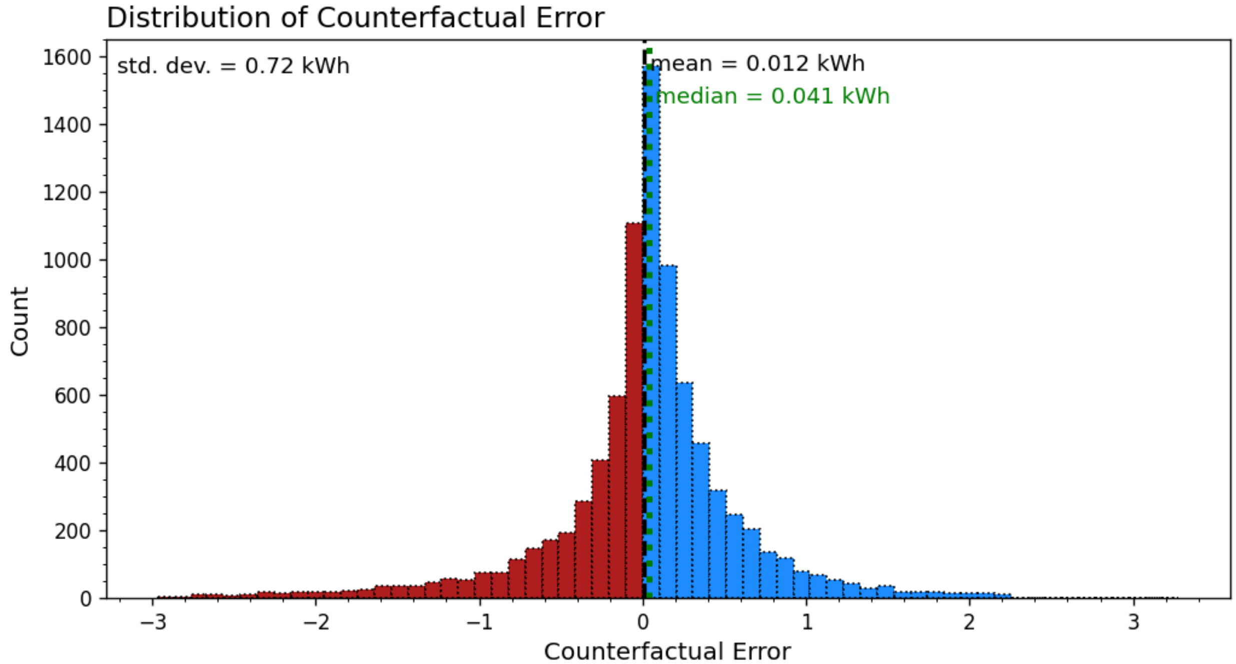
The figure below gives the distribution of gas savings residuals for the analogous gas sample.



When the window size is increased, a higher degree of skew is observed in the negative tail of the observed residuals. However, the mean remains near zero as the median value creeps into positive territory. An example of this is provided in the next two figures (observed and counterfactual error distributions) for the 1% electric weekday sample with a moving average window of 62.

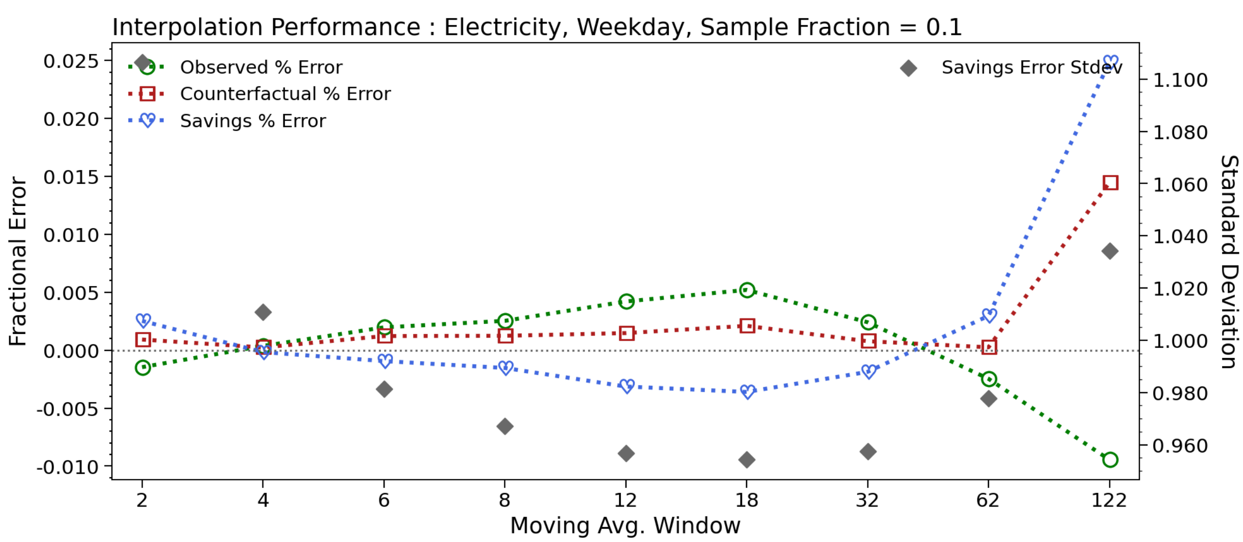


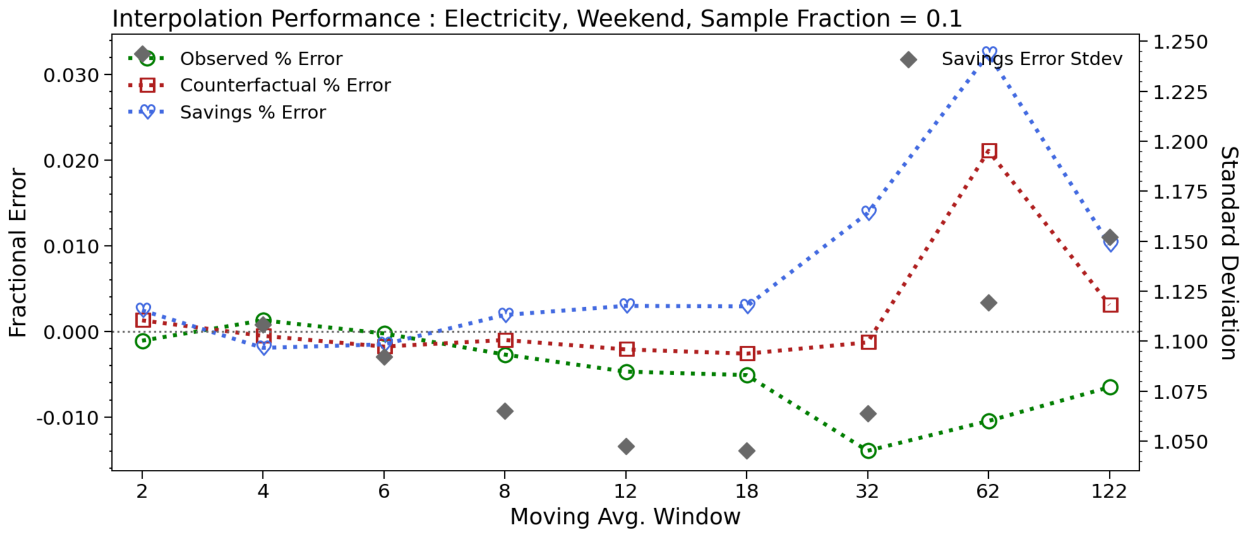
Because the counterfactual results from a model (and therefore will not be prone to high-consumption outliers), the distribution remains quite symmetric even at this large moving average window size.



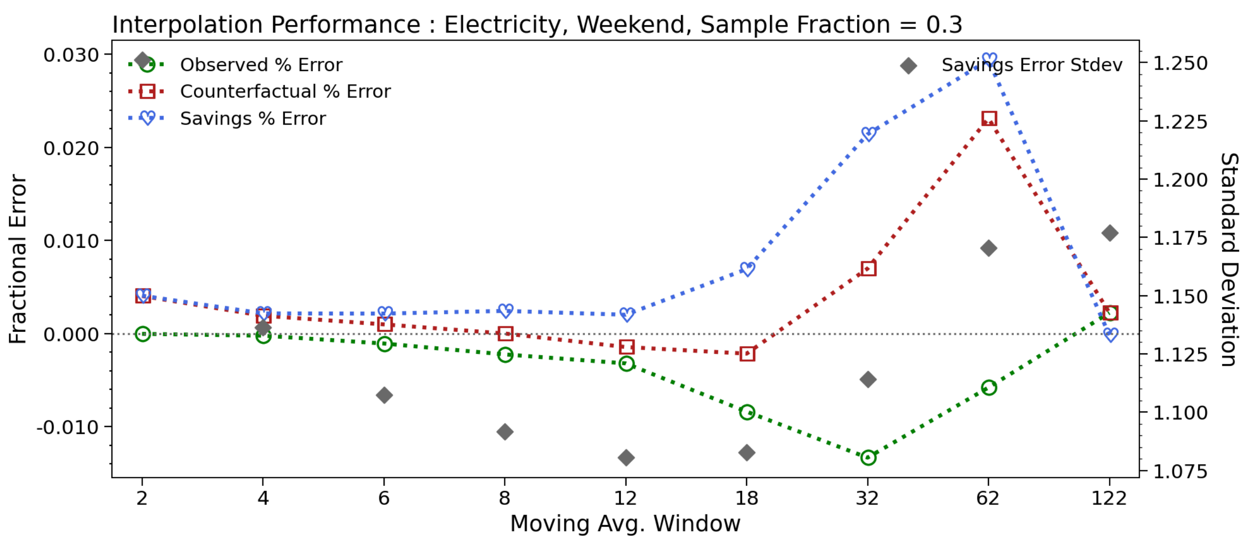
In summary, a window size of 8 performs well for both electric and gas and is recommended for adoption in the interpolation algorithm. This value should be revisited, however, if monthly data ceases to be extrapolated across the days of the month.

E. Additional Figures

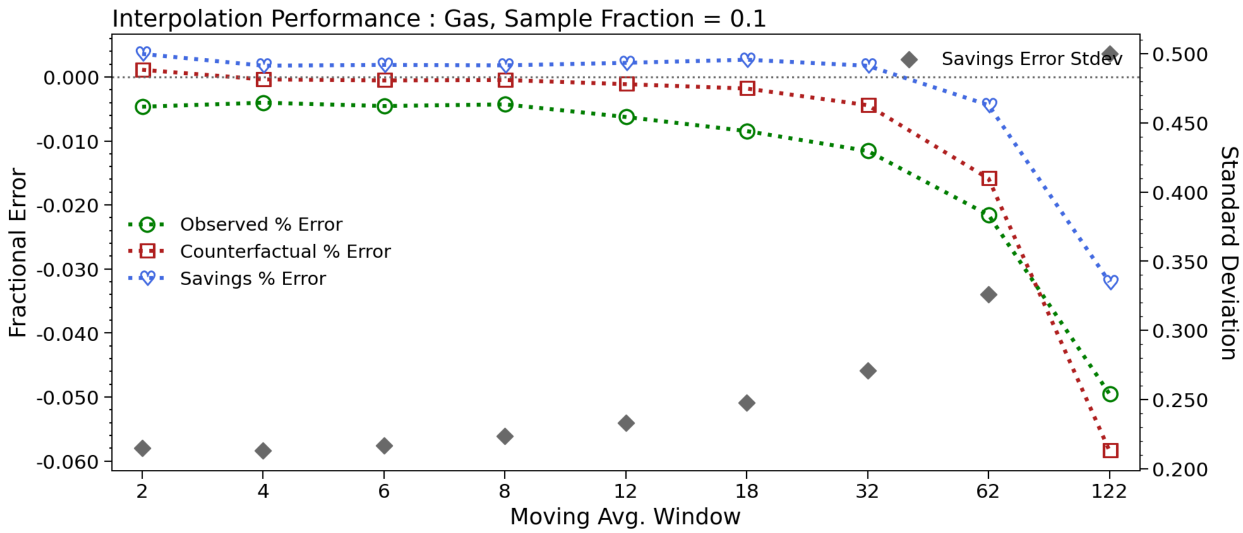




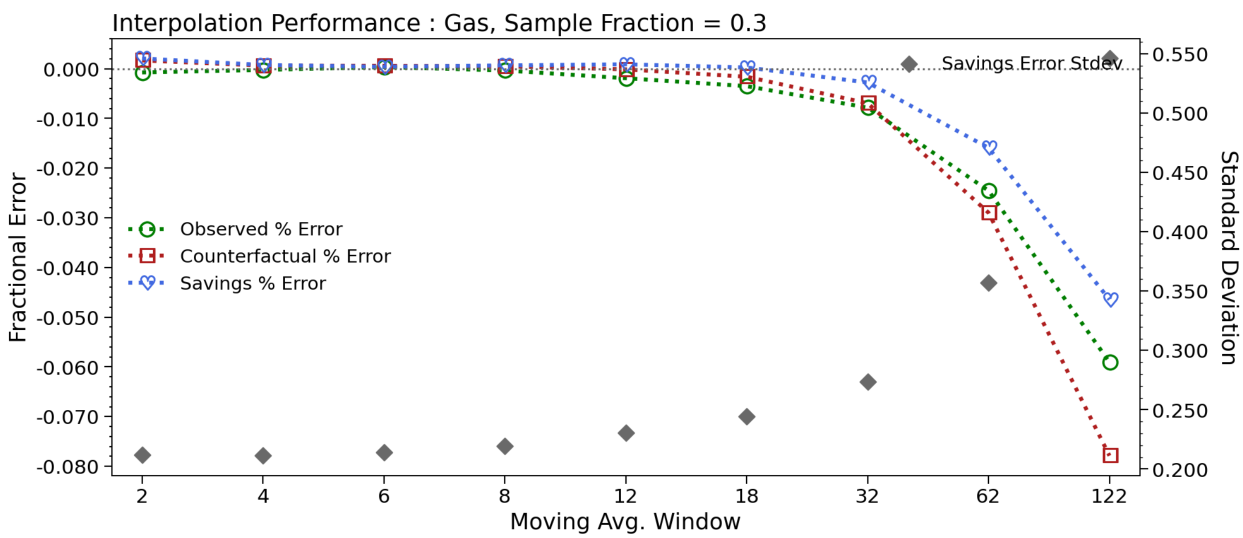
Interpolation Performance : Electricity, Weekend, Sample Fraction = 0.1
Observed % Error
Counterfactual % Error
Savings % Error
Savings Error Stdev
Fractional Error
Standard Deviation
Moving Avg. Window
0.030
0.020
0.010
0.000
-0.010
1.250
1.225
1.200
1.175
1.150
1.125
1.100
1.075
1.050
2
4
6
8
12
18
32
62
122



Interpolation Performance : Electricity, Weekend, Sample Fraction = 0.3
Observed % Error
Counterfactual % Error
Savings % Error
Savings Error Stdev
Fractional Error
Standard Deviation
Moving Avg. Window
0.030
0.020
0.010
0.000
-0.010
1.250
1.225
1.200
1.175
1.150
1.125
1.100
1.075
2
4
6
8
12
18
32
62
122



Interpolation Performance : Gas, Sample Fraction = 0.1
Fractional Error
Standard Deviation
Moving Avg. Window
Observed % Error
Counterfactual % Error
Savings % Error
Savings Error Stdev
0.000
-0.010
-0.020
-0.030
-0.040
-0.050
-0.060
0.500
0.450
0.400
0.350
0.300
0.250
0.200
2
4
6
8
12
18
32
62
122



Interpolation Performance : Gas, Sample Fraction = 0.3
Fractional Error
Standard Deviation
Moving Avg. Window
Observed % Error
Counterfactual % Error
Savings % Error
Savings Error Stdev
0.000
-0.010
-0.020
-0.030
-0.040
-0.050
-0.060
-0.070
-0.080
0.550
0.500
0.450
0.400
0.350
0.300
0.250
0.200
2
4
6
8
12
18
32
62
122