# Background

Given the changes in energy consumption that have occurred over the course of the COVID-19 pandemic, the need for alternative approaches to CalTRACK and similar, simple pre-post regression methods for estimating EE impacts is paramount. While adding comparison groups typically improves the accuracy of these energy saving estimates, there are three, main logistical challenges:

* **Privacy of non-participant customer data.** Current California laws and regulation exist to protect the privacy of AMI/smart meter data for individual customers[[1]](#footnote-1). While EE program participants consent to sharing their consumption data for the purposes of evaluation and settlement, no such agreement exists for non-participants. The regulations limit the ability to use non-participant data for settlement of pay-for-performance programs and create tradeoffs between accuracy, transparency, and access to data, particularly for third-party program implementers. Relatedly, disclosure of participant and non-participant consumption data in any public setting is limited to aggregations of customers that do not violate the 15/15 rule. This rule requires that any public disclosure of consumption patterns be aggregated such that at least 15 customers are included in the aggregation and that no one customer makes up more than 15% of the total consumption.
* **Transparency Challenges.** Many evaluation methods that rely on a comparison group require extensive calculation in order to construct the comparison group from a large pool of eligible non-participants. Additionally, they require familiarity with statistical matching methods and access to high-powered servers with statistical applications to produce the matching variables and construct the comparison groups. This complex system, when operated by an experienced statistician, can result in precisely-matched comparison groups. However, it can hinder independent review and/or replication of the findings. Transparency is limited further if reviewers or implementers do not have access to the non-participant data due to privacy concerns.
* **Complexity and frequency.** PG&E and third-party EE program implementers target a wide range of customer segments and geographic areas. Program enrollment occurs throughout the year and different customer sites have unique pre-intervention periods used for baseline periods. Because developing comparison groups typically relies on finding non-participants with similar consumption profiles and demographic indicators during pre-intervention periods, it can require constructing distinct comparison groups for each implementer, participating customer segment, and treatment period; a proposition that adds complexity to existing program administration processes.

To address these concerns, PG&E has implemented profiles of aggregated groups of non-participants, called granular profiles (GPs). These profiles are comprised of 200 or more non-participant premises that have similar characteristics or segmentation. Published on a monthly basis, these profiles are hourly (or daily for gas profiles) aggregations of actual historic consumption for groups of customers. Because these profiles are reliant on actual consumption patterns of non-participants, they can account for exogenous changes in energy use when combined with participant consumption and treatment data.

This document provides the methodology for the construction of the granular profiles for PG&E, with particular focus on how segmentation is developed.

# Granular Profile Construction

The procedure to generate the granular profile relies on identifying eligible premises that are not enrolled in any energy efficiency programs, applying additional eligibility, size, and volatility screens, then aggregating granular consumption data according to pre-specified segmentation. This section describes the procedure for determining customer segmentation, sampling of granular profile participants, and finally constructing the profiles themselves.

PG&E originally produced granular profiles according to segmentation developed for their PY2019 M&V for Residential Pay-for-Performance and Commercial On-Bill Financing Alternative Pathway programs. These segmentation strategies will be replaced by the new residential and commercial granular profile segmentations in the upcoming months. New segmentation was performed in order to better reflect the diversity of residential load shapes and offer a wider variety of non-residential industry segments. The new segmentation also includes enhanced quality control and validation processes. For reference, both segmentation strategies are defined in this and the following section. While each paragraph in the below highlights a dimension of customer segmentation, note that all granular profiles are constructed for all segmentation dimensions combined; for example, a residential granular profile will represent solar customers in climate zone X with summer size bin Y.

## Define the segmentation Strata

### Residential: Motivation

In determining the new segmentation strategies for residential granular profiles, PG&E and DSA defined several high-level principles for selecting customer segments. These principles originated in conversation with Recurve to ensure that all parties were confident that program estimates could be produced accurately.

* **Diversity of load shapes:** Per Recurve, residential load shapes need a variety of load profiles
	+ Construct segmentation on load shape variables
* **Capture solar and non-solar profiles:** Solar customers look significantly different to non-solar customers
	+ Produce separate shapes by solar status
* **Consider base characteristics that influence consumption:** Granular profiles should be separated by climate zone and/or geography to reflect differences in weather and economic conditions
	+ Produce segments by climate zone groups and electric heat status
* **Simple to calculate:** Should not require undue calculations for PG&E
	+ Simple segmentation – load shape variables easy to calculate
* **Data sufficiency:** Profiles should have enough customers so that outliers and account churn can be accommodated
	+ Require sufficient premises in each segment to ensure profile robustness

The segmentation strategy that resulted from these principles relies on grouping residential premises in according to four categories: Climate Zone Groups, Solar Status, Premise Characteristics, and Load Shape Characteristics.

The load shape segmentation is a new addition to the residential granular profiles and addresses concerns that Recurve had regarding how to accurately capture peak and net peak savings. This statistic is calculated by computing the percent of total daily consumption that occurs between 1pm and 4pm on summer weekdays with no heating or cooling load (based on temperature conditions). This allows for more accurate assessments of customer occupancy and produces a diversity of granular profile load shapes. Shown in Figure 1 are example granular profiles produced from this segmentation. Each cluster of five load shapes represents one of four quartiles of load shape segmentation for Non-Solar customers in the Coastal climate zones. The five load shapes in each segment represent the four quartiles of annual consumption plus the electric heat customer group.

Figure 1: Example Load Shape Segmentation Results



### Commercial: Motivation

Commercial granular profiles reflect the diversity of industries in PG&E’s territory. Rather than using load shape segmentation – as was done for residential premises – the non-residential segmentation strategy involves segmenting customers into sizes and industries within a climate zone group. The industry mapping is done on the basis of premise NAICS codes and includes mapping for the groups listed in Table 2.

Table 2: List of Commercial Granular Profile Industry Segments

|  |  |
| --- | --- |
| Agriculture and Pumping | Manufacturing |
| Automotive and Repair | Miscellaneous/Unknown |
| Education | Office |
| Full-Service Restaurants and Bars | Personal Care Services |
| Gas Stations & Convenience Stores | Property Management |
| Government-Institutional | Religious |
| Grocery | Retail |
| Health | Transportation, Communications, and Utilities |
| Limited-Service Restaurants | Warehouse and Wholesale |
| Lodging and Entertainment | Construction |

Granular profile premises are also segmented in to climate zone groups and customer size, as determined by the premise’s rate code.

### Climate Zone Groups

To improve reporting across a wider variety of weather conditions, both commercial and residential granular profiles will now be produced for each of four climate zone groups:

1. Coastal (Zones 1, 3, and 5)
2. Inland (Zones 2 and 4)
3. North Central Valley (Zones 11 and 12)
4. South Central Valley (Zone 13)

This segmentation improves on the 2 climate zone groups available for non-residential customers in the prior version of the granular profiles, while also accommodating an additional dimension of segmentation for residential customers (that is, adding load shape segmentation).

### Residential: Solar Status

Residential granular profiles are constructed separately for customers with installed solar PV and for those without. The information about whether a premise has solar installed comes from PG&E’s interconnection database, which documents the interconnection date and installed capacity for each premise.

### Residential: Customer Size and Electric Heat

Residential granular profile candidates are segmented in to groups based on electric heat status, as defined by the electric end-use rate code within PG&E’s customer database. For sites without electric heat, they are further segmented based on size bins within each climate zone group. The size bins are constructed on the basis of relevant premise characteristics: installed solar capacity for solar customers, annual kWh for non-solar customers, and annual therms for gas customers. A summary is shown below.

Table 3: New Residential Premise Size Segmentation

|  |  |  |  |
| --- | --- | --- | --- |
| Bin  | Electric - Solar | Electric - Non-Solar | Gas |
| Electric Heat | All Electric Heat | All Electric Heat | N/A |
| Bin 1 | 0-25th Pctile PV kW Capacity | 0-25th Pctile Annual kWh | 0-25th Pctile Annual Therms |
| Bin 2 | 26-50th Pctile PV kW Capacity | 26-50th Pctile Annual kWh | 26-50th Pctile Annual Therms |
| Bin 3 | 51-75th Pctile PV kW Capacity | 51-75th Pctile Annual kWh | 51-75th Pctile Annual Therms |
| Bin 4 | 76-100th Pctile PV kW Capacity | 76-100th Pctile Annual kWh | 76-100th Pctile Annual Therms |

As was the case with the old segmentation, the cut-points for each size bin are saved so that any new customers will be added to the same size bin, instead of recalculating quartiles of usage from a different sample.

### Residential: Customer Load Shapes (Electric Only)

A new feature for residential electric granular profile segmentation is customer load shapes. This dimension was added to improve matching across a range of residential occupancy patterns. The steps to construct the load shape variables are:

1. Find days with no heating or cooling load in the last year. These days have an average temperature between 59F and 61F and a daily maximum temperature of less than 77F across a simple average of coastal and inland weather stations.
2. For each premise on each of these days, compute the percentage of the daily total kWh consumed between 1pm and 4pm (HE14-HE16).
3. Average the percentage consumed across the days identified in Step 1 for each premise.
4. Within each climate zone group, solar status, and size bin, construct four quartiles representing equal bins of premises ranked based on the average percent of consumption occurring in the midday hours.

This allows for more accurate assessment of customer occupancy and produces a diversity of granular profile load shapes. As with the customer size bins, the cut-points of the initial segment development are saved to be reapplied in the future.

### Commercial: Industry and Size

The segmentation for commercial customers is similar to that which was done originally: the combination of industry and customer size, which is determined by the customer’s rate class. However, additional industries have been added and profiles are constructed for each customer size, provided that there are enough customers in the segment, size group, and climate zone group to support it. In cases where there are not enough customers to have separate small and medium/large size bins for a given industry and climate zone group, the groups are combined across customer sizes as shown in green in Table 4.

Table 4: Commercial Customer Counts and Combined Industry/Size Segmentation



Rates are defined as they were in the old segmentation. Customers in A1, A6, AG1, AG2, AG4 and AGA rates are classified as small while customers on A10, A15, B19, E19, B20, E20 AGB, AGC, and AG5 rates are classified as medium/large. Customers on any other set of rates (transportation, streetlights, electric vehicles, standby or special rates) are excluded.

## Sample GP participants for Each Segment

### Valid Granular Profile Participants

The initial production of granular profiles relies on identifying valid premises that can be sampled in to the profiles. Additionally, sites should be screened every month to ensure they still meet the criteria defined in Table 5.

Table 5: Granular Profile Candidate Exclusions

|  |  |  |
| --- | --- | --- |
|  | Electric | Gas |
| Residential | * Full panel of data for the prior year
* No change in solar status (no addition of onsite solar, no adding of incremental capacity)
* No EV rates
* No other DERs (e.g. batteries)
* No EE participation in the last 12 months
 | * Full panel of data for the prior year
* No EE participation in the last 12 months
 |
| Non-Residential | * Full panel of data for the prior year
* No DERs (solar, battery, or other) onsite
* No EV rates
* No EE participation in the last 12 months
 | * Full panel of data for the prior year
* No EE participation in the last 12 months
 |

These exclusions were applied both to the old and new customer segmentation strategies. Additional construction and validation were done to produce the new granular profiles, which are described below.

### Picking Initial Sample

To pick the initial sample of commercial granular profile customers, DSA used the full commercial population – removing any invalid premises as per Table 5 – to assess segment coverage and to perform the sampling for granular profiles. 300 unique premises per granular profile were picked. For residential, PG&E provided a stratified random sample of residential sites meeting the exclusion criteria above. From that sample – stratified by climate zone group, solar status, and electric heat status – 400 premises were selected.

### Removing Large/Influential Sites

Any commercial premise with greater than 5% of the total consumption in that segment was removed. For residential, any site with annual consumption greater than the 99th percentile or less than the 1st percentile of the group was excluded.

### Selecting Primary Granular Profile Participants

Once outliers were removed, 200 premises were randomly sampled within each segment to produce the primary premises for the granular profile. If there were not 200 premises in the segment, all customers were selected. This only occurs for a small subset of commercial industries.

### Picking Alternate GP Participants

A feature of the new GP production process is pre-selecting sites that will be used to replace any premises that violate the eligibility criteria at any future point. The alternates are selected from the initial 300 or 400 premises and have the same size restrictions applied. The alternates are selected for each GP participant to closely match the existing premises on the basis of annual kWh. That is, for each primary premise, a list of five alternate premises are developed that are as close as possible in size to the primary customer. An alternate premise can act as an alternate for multiple primary premises. More detail about the substitution procedure for primary GP participants can be found in Section 3.

## Get AMI data for the extract of sampled sites & Screen outLiers

The initial sample of granular profile candidates were screened to have a full panel of data during the prior year. Large sites were removed from each granular profile segment, however individual reads for primary participants will also be removed. These reads, which may contain spurious meter reads are identified based on the 3rd and 97th percentile of that premise’s full year of data. Reads that were removed were those with greater than 3 times the 97th percentile or:

* Less than 0 net kWh for non-solar customers, or
* Less than 3 times the 3rd percentile for solar customers. As solar customers will have negative net kWh reads, this screen removes large negative readings.

## Aggregate to a single time series per segment

For the valid primary premises and valid net meter reads (that is, delivered minus export channels) were averaged for each segment, date, and hour. The final dataset contains each granular profile, a date and hour indicator, and both an average customer consumption read and count of valid premises that reading contains. For gas profiles, the profiles are constructed on a daily basis rather than hourly.

# Granular Profile Maintenance

This section documents the updates to the granular profile maintenance process that will be put in place for the new profiles.

### Monthly Maintenance

Each month, granular profile primary premises and all alternates should be assessed to determine if they are still eligible to be part of the profile. This involves determining if any site now no longer meets the criteria outlined in Table 5. If any primary premise no longer meets these criteria, an alternate will be selected from the pre-defined list. The specific steps for this replacement procedure is as follows:

1. Start with the file containing current primaries and all assigned alternates
2. Identify any primaries and all alternates that are now ineligible
3. For those that are ineligible, assign their most similar eligible alternate to be the new primary
4. Go through the list of new primaries and ensure there are no duplicates
5. This can happen because a secondary can be assigned to multiple primaries and could then be substituted for both if they both become ineligible
6. If any duplicates occurred drop the primary for all after the first occurrence of the customer and pull in the next alternates
7. Repeat the process until all primaries are unique
8. Update the list of alternates to remove the premises that have been promoted to primaries.

Once the list of primaries has been updated, the new month of AMI data can be produced. When new premises are added to the primary list for a given segment, AMI data for the old primary should be truncated at the end of the prior month. That is, if during the September 2022 update to the granular profiles, it was found that a primary customer installed solar at their site as of August 15th, their AMI data should run through July 31st, and their selected alternate’s data should be substituted starting August 1st. The application of outlier screenings to the new AMI data should still apply as all primaries and alternates are required to have a full year of data from which to do segmentation and assess customer volatility.

### Monthly Validation

As part of the updated granular profile production process, additional validation of each profile is done on a monthly basis. The checks are detailed in code, but contain:

1. Visual inspection of profile loads for any trends, outliers, etc.
2. Flagging any observation that has less than 90% of the target number of participants (e.g. fewer than 180 premises for commercial profiles which should contain 200 premises each)
3. Flagging any observation that violates a temperature-load trend (using local regressions to highlight major deviations in average daily consumption)
4. Flagging any observations that violate short-term trends in daily consumption (using local regressions to highlight major deviations over time)
5. Assessing any profiles or dates with multiple flags (typically more than 5% of days having one or more flag)

In the case of concerns with a given profile, the underlying profile data should be investigated to see if there is a specific primary premise creating the issue. In that case, that premise may be substituted with a valid alternate.

### Annual Maintenance

On an annual basis, the list of alternates should be replenished to ensure there are sufficient premises available to backfill ineligible primaries over time. While this process should happen at least annually, it can be done more frequently if all alternates have been depleted. .

1. Start with most recent list of all eligible premises that are not already in a GP.
2. Compute each premises granular profile segmentation using the prior year of their data.
3. Sample 300 (or as many are available) potential alternates for each segment
4. Filter out large customers based on customer annual usage
	* 1. If a customer makes up more than 5% of the candidate GP group’s total annual usage then toss them out and check again.
5. Assign alternates to the primaries from those that remain in the candidate pool
	* 1. Assign alternates to primaries based on which are closest to each primary in size (annual usage)
		2. Each Primary gets 5 alternates and alternates can be assigned to multiple primaries at a time.
1. Including CPUC Decision D.97-10-031 (1997) Direct Access Proceeding (15/15 rule), Senate Bill No. 1476 (2010) Chapter 5 Privacy Protections for Energy Consumption Data, and CPUC Decision D.14-05-016 (2014) Decision Adopting Rules to Provide Access to Energy Usage and Usage-related Data While Protecting Privacy of Personal Data (2014) [↑](#footnote-ref-1)