Applying Energy Differential Privacy To Enable Measurement of the OhmConnect Virtual Power Plant

A study of Demand Response during the California August 2020 blackouts.

Prepared By: Marc Paré, Mariano Teehan, Stephen Suffian, Joe Glass, Adam Scheer, McGee Young, Matt Golden

Prepared for:

U.S. DEPARTMENT OF ENERGY

RECURVE
SHAPE THE FUTURE OF ENERGY

Table of Contents

Abstract

To scale the cost-effective deployment of demand side energy resources, utilities and other parties must access and utilize population consumption data for the comparative measurement of load impacts. However, insufficient and dated privacy protections for customer data have inhibited the release of data for the development of robust, standardized methods needed to create a common playing field for demand flexibility markets. Without the data resources needed to develop and execute consistent methods, dozens of competing measurement techniques have been adopted, resulting in ambiguity and confusion that ultimately discredits demand side solutions to modern grid and climate challenges.

New tools are available that are capable of providing mathematically rigorous privacy protections while retaining the utility of the underlying data. In this work we use open-source Comparison Group sampling methods along with Energy Differential Privacy tools to assess the load impacts of a recent demand response event that was administered by OhmConnect to address an emergency electricity shortage on California's grid. These open-source resources have been developed in partnership with the National Renewable Energy Laboratory and supported by the Department of Energy.

Within Energy Differential Privacy, the choice of the core privacy parameter, ε , is an ongoing area of research and development. The question is not simply a technical one, but a negotiation between the mathematical, legal, political, and social aspects of data privacy. At its heart, the choice of ε represents the balance struck between data privatization and accurate information. In differential privacy, as protection is achieved via the addition of, noise to underlying data or analysis results. However, noise also introduces error, which can decrease the value of measured resources in the market.

In this work, we erred on the side of caution, picking epsilon values representing a high degree of privacy protection (lower epsilon equates to greater privacy protection), setting the comparison group at ε =0.843. This resulted in an error bound of 2.6% in the load impact measurement. Simulation of privacy threats against energy usage data, and especially derivatives like savings or demand flexibility (load shape impact), suggest that differential privacy guarantees dramatically reduce privacy risk. With continued stakeholder input and consideration of societal value versus customer privacy risk, along with improved privacy engineering, we believe significant room exists to further improve precision while maintaining robust privacy protections.

Research Summary

To effectively deploy behind-the-meter flexibility to meet the grid challenges posed by short-term disaster events, medium-term disruptions such as COVID shutdowns, and long-term structural changes including decarbonization and increased renewable generation, utilities and other parties must be able to access and make use of population energy consumption data.

Historically, however, a lack of smart meters and insufficient privacy protections for customer data have inhibited the development of methods needed to produce confident, predictable, and transparent telemetry into the impacts of these increasingly essential flexibility resources.

To solve this problem, Recurve, supported by DOE and in partnership with the National Renewable Energy Laboratory (NREL), Lawrence Berkeley National Laboratory (LBNL) and MCE, recently released the GRIDmeter. This open-source methodology identifies comparison groups via stratified sampling on key usage parameters to enable a high level of accuracy and confidence in behind-the-meter resources. These data-driven approaches represent a robust solution to endemic measurement problems that have injected uncertainty into markets and inhibited access in both the energy efficiency and demand response markets.

Essential to this innovation is the application of Energy Differential Privacy. This mathematically rigorous framework allows privacy-protected energy consumption data to inform comparative measurement, while ensuring robust privacy protection at an individual customer level. The application of Energy Differential Privacy goes beyond comparison group analyses. Recurve outlined several use cases in "Differential Privacy for Expanding Access to Building Energy Data" that includes customer targeting for enhanced demand-side program impacts, the assessment of COVID impacts to energy consumption, and a host of use cases for which population energy data can provide essential information are all made more secure, and therefore possible, with the incorporation of differential privacy methods.

This report explains the Energy Differential Privacy framework for the protection of individual customer energy consumption data privacy while still deriving critical information. When used for a new class of demand flexibility analytics, the methods described here can enable accurate hourly measurement of event-driven demand response and predictable long-term load shaping through energy efficiency and other mechanisms.

We then summarize how Energy Differential Privacy was applied to real data, quantifying the impact of OhmConnect's virtual power plant of participating residential customers. This VPP was deployed to mitigate last summer's major grid event that resulted in power outages across the State of California.

Differential Privacy Use-Case

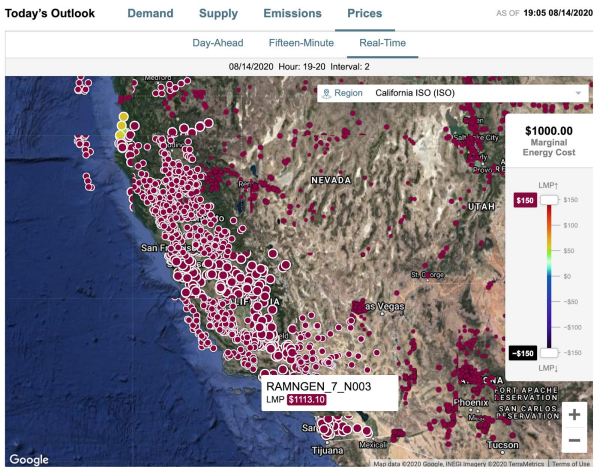
This work focuses on a real-world example of how differential privacy is unlocking the data necessary to facilitate accurate and reliable methods for measuring the impact of demand response during a period of extreme grid stress in California.

Recurve is fortunate to have the partnership of OhmConnect, one of the most advanced aggregators of peak load reduction for residential customers, to provide an extremely relevant test case for the application of Energy Differential Privacy, and MCE, California's first community choice aggregator.

Measuring OhmConnect's VPP

OhmConnect is a Residential Demand Response Virtual Power Plant developer based in California with over 150,000 customers nationwide. OhmConnect pays people to save energy in response to grid events and is in turn paid by utilities and CAISO for demand response and resource adequacy. During periods of peak usage, OhmConnect engages users through a combination of gamification, economic rewards, and direct control of grid-edge devices to reduce demand during peak events.

On August 14th, 2020, high temperatures pushed the California grid into a significant event that resulted in statewide blackouts. In an emergency bid to enhance supply, the California Independent System Operator's (CAISO) real-time marginal pricing skyrocketed to over $1000/MWh (compared to a typical price of around $35/MWh). However, day-ahead prices remained below the price cap of $1000/MWh. Behind-the-meter demand response resources were also dispatched to help reduce demand and stabilize the grid.



However, OhmConnect and other Demand Response providers could not deliver their full potential due to the combination of a lack of available data and consistent methods for accurate measurement, and complex market rules, arbitrary market caps, and heavily discounted value for delivered load reductions. The day ahead LMPs did not reach the price cap during the rolling blackouts, and as a result, many Third Party Demand Response Providers did not dispatch their resources. Despite these barriers, OhmConnect did dispatch when called upon, taking on hundreds of thousands of dollars of liability that the company has been unable to recover.

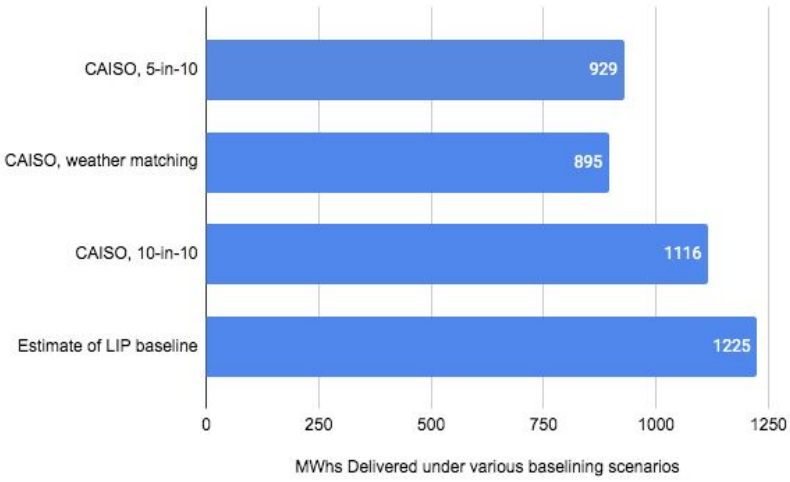
It is clear that under such conditions as encountered on August 14, any amount of energy saved or shifted off-peak is valuable. However, gauging the impacts of a demand-side intervention requires establishing a robust counterfactual that estimates customer energy consumption in the program's absence.

Current tactics for quantifying the impacts of behind the meter resources such as those used to assess Demand Response and Resource Adequacy rely on methods designed before the ability to access and process smart meter data became widely available. In addition to the technical challenges of applying outdated methods to the granular data sets now available through California's investment in smart meters, the lack of an adequate privacy framework has restricted access to the very data sets needed to identify appropriate comparison groups and thus enable accurate measurement.

Without population datasets to conduct timely measurement, regulators, utilities, and implementers have had to rely on outdated methods to determine where, when, and how much a given program or intervention has affected energy consumption.

The California Load Impact Protocols illustrate this problem. With 149 pages of flowcharts and narratives describing general methods for a range of use cases to measure demand response resources, the Protocols give multiple answers to the same question from different agencies and evaluators, even when looking at the same event.

This chart, produced by OhmConnect, demonstrates the variance of the load impact measurements currently used for different California applications.



Depending on which calculation method is used, very different values result – each of which is correct within the context of its implementation code and hundreds of embedded engineering choices that are not specified in advance. In fact, due to a lack of precise specifications for any of these methods, such as multiple applications for "same-day adjustments," different engineers can commonly get different answers, even using the same beginning scenarios.

Lack of a consistent measurement is a significant impediment to companies throughout this space. In an article in Microgrid Knowledge covering their recently announced Alphabet (Google) funded $100M virtual power plant, OhmConnect CEO Cisco DeVries commented that,

> "Depending on which agency we're dealing with, there are five different ways that they count how our energy reduction is measured. And they change those all the time...In order to create a long-term investment in flexible load, it's important to know how to count the megawatts. This affects what the consumers are paid as well as the types of devices that are controlled."

All of these approaches are limited by a lack of data and rather simplistic and easily biased methods. For example, the 10-in-10 calculations measure the DR baseline by taking an average of the customer's last ten days of usage at the DR event time. This approach fails to account for the fact that the event day is usually much hotter than other days by its very nature. Lacking access to population data for a comparison group, a "same day adjustment" is made to scale the event to that day's energy use. There are multiple methods to calculate this adjustment, each producing different results with a range of known biases. In addition, the adjustment is often capped, though the caps can vary based on baselining methodologies (80-120% or 71-140%).

The problem is compounded by the fact that a customer's baseline is affected explicitly by other DR events. This means that responding to an event on one day can reduce the value of future event periods and create a dynamic where long-term and consistent energy efficiency or seasonal load shaping to fight the duck curve undermines the value of event-based DR.

Convergence Data Analytics' 2018 Load Impact Evaluation for OhmConnect's DR Resource report summed up the need for a comparison group and more advanced methods:

Conclusions and Recommendations

Based on our evaluation of the 2018 dispatch of OhmConnect's behavioral DR resource, we provide the following recommendations:

Recommendation	Description
Investigate more robust baselining and comparison group methodologies for estimating event impacts	OhmConnect's DR events have short durations, so comparison day baselines with same day adjustments (SDAs) can be used for estimating impacts for the handful of hours around event windows. Joining a long list of evaluators, we have documented that raw 10 in 10 baselines would not have been serviceable without SDAs. But more importantly, baselines can't establish that the event savings were caused by the event as convincingly as randomized or even synthetic controls. For a pioneering behavioral program to convince its skeptics that its results are legitimate, adopting the most rigorous methods would go a long way.

These issues are not specific to OhmConnect, affecting market access for a broad set of VPP providers. Leap provides a platform that includes everything from residential smart thermostats to commercial EV charging stations, and pumps at municipal water treatment facilities.

According to Jason Michaels, Chief Commercial Officer at Leap, "Almost by definition, current baseline methods understate performance on the days when the grid has the greatest need for demand response, resulting in reduced incentive to support the grid in future events. More accurate methods for measurement and verification will help companies like Leap bring more flexible demand from local distributed energy resources to help balance the grid."

Recurve FLEXmeter Settlement Quality Measurement

In this work, Recurve combines the OpenEEmeter and GRIDmeter methods and code with the Energy Differential Privacy methods that have been developed with the partnership of DOE, NREL, and MCE. When applied consistently, this approach can overcome current measurement barriers and improve confidence in the delivered DR resource. In this way, DR resources can be correctly valued and fully deployed to stabilize and decarbonize the grid. To demonstrate this approach, Recurve carried out a load impact analysis of a fraction of OhmConnect's participants during the hours of the August 14th demand response event.

Recurve's measurement methods involve a two-step calculation that is described in greater detail below. In short, a meter-level hourly baseline is created for every treated and comparison pool customer using the hourly CalTRACK methods and the The CalTRACK calculations yield hourly counterfactuals for each meter based on past consumption.

Next, the GRIDmeter methods and code draw a comparison group sample from the comparison pool. At the heart of these methods, multidimensional stratified sampling on key usage characteristics is utilized to produce a comparison group that accurately reproduces the load shapes across the entire distribution of treatment customers.

Using the CalTRACK model outputs, the difference between the observed hourly consumption and the counterfactual is measured for both treatment and comparison group customers. Taking the difference (known as the "difference of differences") between these two sets of measurements yields the comparison-group adjusted hourly load impacts.

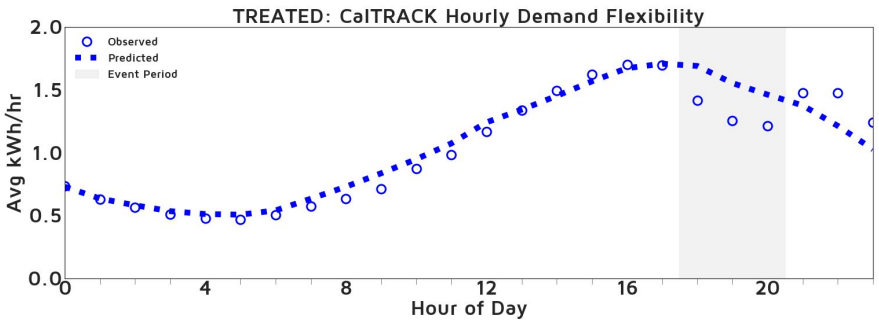
This calculation can be made transparent and auditable by all parties and is significantly more accurate and reliable than any traditional existing demand response methods.

Results from OhmConnect's VPP Dispatch During the August 14th Event

Recurve applied the methods described above to a statistically significant random sample of OhmConnect customers in the MCE territory. Given the emergency situation on the California grid, most of the customers were called upon to reduce load during the full 3-hour window of the event. Often OhmConnect will ask customers to reduce load for shorter windows or will stagger the event hours across different populations.

Revenue-Grade Demand Flexibility

The figure below shows observed consumption (open circles) and CalTRACK hourly counterfactual (dashed curve) results for the treatment group of OhmConnect customers on the date of the demand response event (Aug. 14, 2020).



TREATMENT GROUP:

-17.5%

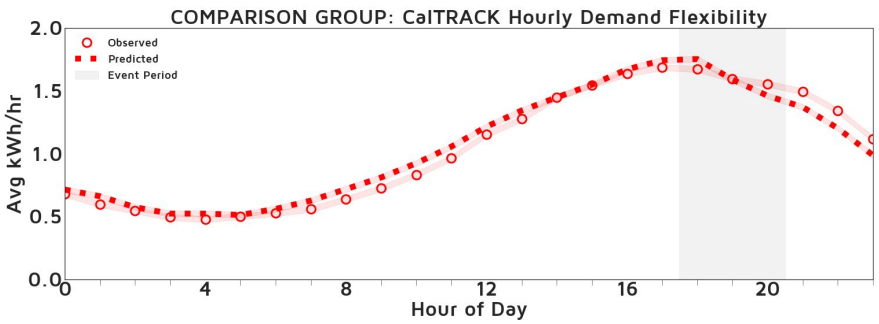
Event Demand Reduction

CALTRACK

LFENERGY OPENEEMETER

The next figure shows analogous results for the comparison group selected by the open-source GRIDmeter. This difference between the curves in the below figure represents the exogenous change in the comparable population. For example, perhaps these event hours were unusually hot (likely), or customers were hearing on the radio about a flex alert and taking action even without OhmConnect. This comparison group enables us to net those effects out.

Note the fuzzy line around the comparison group in the figure below. That "fuzz" is the noise being introduced via Energy Differential Privacy and is precisely what protects individual customer records thus enabling the use of population data.



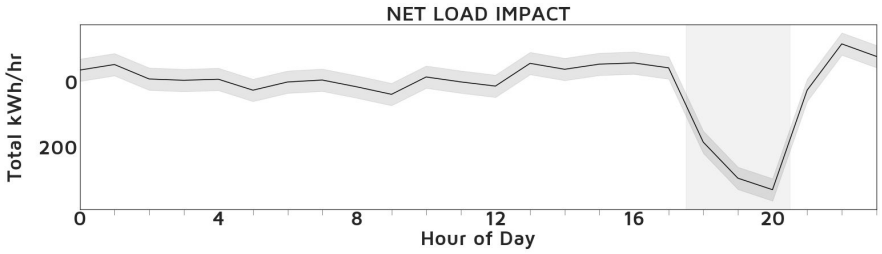
COMPARISON GROUP:

-1.8% ±0.5%
Event Demand Reduction

OPEN GRIDMETER
ENERGY DIFFERENTIAL PRIVACY

Settlement-Quality Net Impact to Loadshape

Finally, the next plot shows the final load impact measurement resulting by combining the measurements of the first two figures. for treated customers.



NET IMPACT:

-19.3% ±0.5%
Event Demand Reduction

≡1.26 kW
Event Demand Reduction

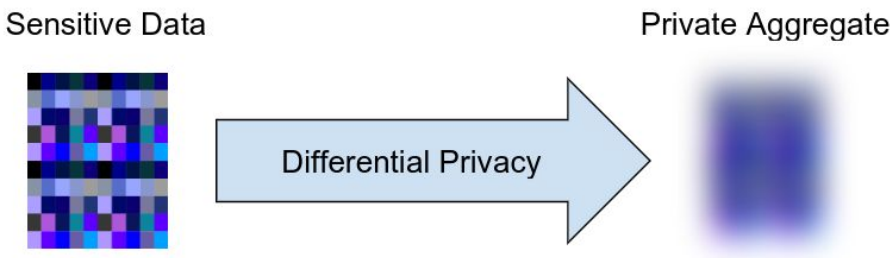
With an Energy Differential Privacy protected comparison group, Recurve is providing a transparent revenue-grade calculator of the **net hourly impact to demand** during this event window delivered by OhmConnect's customers in the MCE service territory.

Data Privacy Context

To make appropriate legal use of non-participant data for comparison groups and other use cases, the data must be obscured through privatization techniques that guarantee data for any particular customer cannot be revealed by any means. Such privacy restrictions are common. As the use of restricted data is vital for the continued development and improvement of the DR industry, it is essential to develop dedicated techniques to fully preserve data privacy while allowing for its productive use. The development and application of one such technique, differential privacy, is the principal focus of this report. Non-participant data is provided by our research partner MCE (California's first Community Choice Aggregator), while OhmConnect provides participant data; both datasets are privatized using differential privacy.

In the 15/100 approach, a common privatization tactic used in the industry, an aggregation of buildings must be composed of at least 100 buildings. No building can contribute to more than 15% of the total usage; if these conditions are met, the aggregation is considered publishable, and the data safe. Unfortunately, this class of anonymization technique, while simple to understand, suffers from a number of unintuitive weaknesses, which have been gradually revealed by the privacy research community over the last decade. These weaknesses are only growing more significant as more data are collected and computational power increases.

Differential privacy is an emerging solution to these problems. Differential privacy is a rigorous mathematical framework for obscuring data through the addition of "noise," i.e., random numbers, which allows for quantifiable guarantees regarding the data's safety. It was first developed in 2006 and is now commercially used by Google, Apple, and Uber as well as in the public sector by the US Census.



In practice, privatized data is represented not as a single number but rather as a confidence interval within which the true result falls to a certain degree of confidence. For example, if the true result is 1.7, the published report might declare that the true result lies within the interval (1.5, 1.8) at 95% confidence.

Sensitive data in this report has been protected through differential privacy. An attacker with perfect information can only increase their knowledge about whether an individual's data was even included in the dataset up to a threshold, parameterized by the variable ε (epsilon). Smaller ε values result in a stronger privacy guarantee, while larger ε values provide a weaker privacy guarantee.

Over the past two years, Recurve has investigated differential privacy methods and developed Energy Differential Privacy, an open-source library, *EEprivacy*, which provides convenient implementations of differential privacy functions tailored towards energy data privacy use cases.

With Energy Differential Privacy to enable data access with mathematically rigorous privacy methods, we are confident that the foundational measurement barriers inhibiting large scale deployment of demand response and other behind-the-meter resources can be solved.

Privacy Model

In this section, we describe the data, transformations, and privacy considerations for comparison group analysis. What sensitive data are we handling? How is it transformed into aggregate statistics? Which statistics will be published and to whom? How do these statistics tie back to privacy risk for individuals?

This overall model of the privacy problem will then inform the concrete anonymization procedure that follows.

Datasets

We consider one sensitive dataset of residential energy consumption: the Comparison Group Pool. The other dataset employed, the Treatment Group Pool, is not sensitive, as participants gave their consent for data to be used for Demand Response applications.

To build CalTRACK hourly models and counterfactuals for all meters in the treatment and comparison pools, hourly data for the months leading up to and subsequent to the demand response event on Aug. 14, 2020 were utilized. The comparison pool includes 62,174 non-solar electric meters that are randomly sampled from MCE's residential customer base of approximately 500,000, with 4,948 eventually selected as the comparison group. The treatment group is composed of 961 non-solar buildings selected from OhmConnect customers in MCE service territory.

Besides directly observed energy consumption, the dataset also included "predicted" energy consumption, which is weather normalized energy consumption predicted by the *eemeter*.

Site	**Measure**	**Hour of Day**	**Energy Consumption (kWh)**
ae871abc82	observed	0	1.0
ae871abc82	predicted	1	1.1
ae871abc82	observed	0	1.0
...			

Dataset structure

All of these data are classified as sensitive.

Outputs

Two population-level outputs are desired:

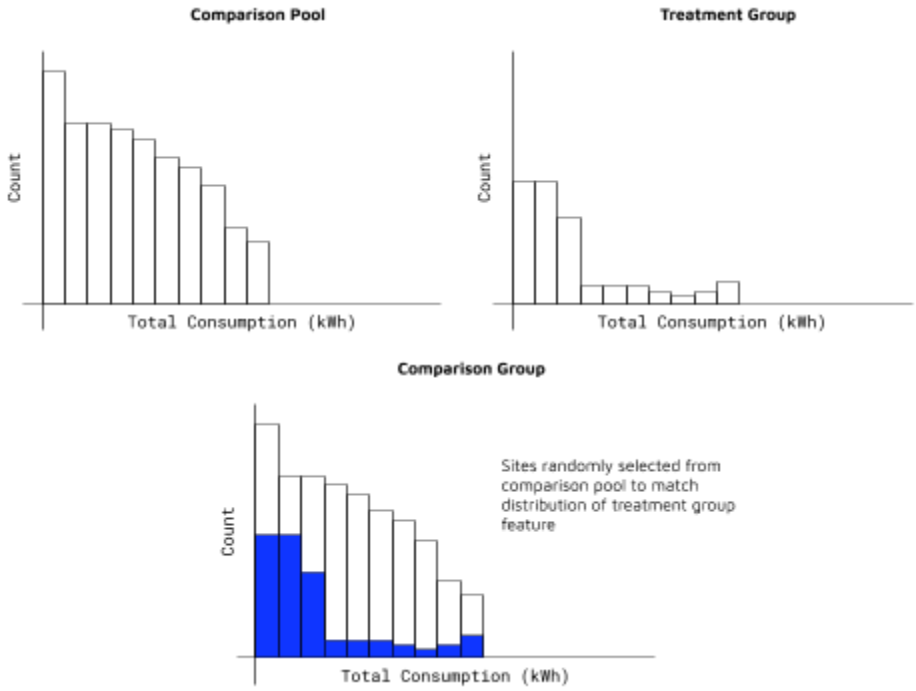
- Average Load Shape
- Percent Savings

However, before these outputs can be constructed, a comparison group must be selected from the comparison pool.

A simplified version of this procedure is described for privacy analysis.

First, a set of one or more numerical features derived from energy consumption (e.g., total energy consumption) is computed for each meter in the treatment group and comparison pool during a time that excludes the DR event of interest, producing a treatment dataset and comparison dataset.

The treatment dataset is grouped into a set of bins, and the proportion of treatment meters in each bin is computed. Next, those bins are applied to the comparison dataset. A set of meters is sampled from the comparison pool so that the proportion of comparison meters in each bin is equal to the proportion of treatment meters in each bin. The resulting set of sampled meters is called the "comparison group" and should have energy consumption patterns similar to those of the treatment group. The procedure is diagrammed below:



Schematic of simplified comparison group matching procedure

From there, the desired outputs are computed for each of the two groups and compared to one another to quantify energy savings.

Using the GRIDmeter, the stratified sampling is conducted on a multidimensional basis using parameters expected to be sensitive to the intervention under investigation. In this case, Recurve used three usage-derived features as the basis of stratified sampling: Summer MWh usage, the percentage of usage from cooling (determined from the CalTRACK 2.0 Daily model run through the OpenEEmeter, and the percentage of customer usage during the summer peak period (June - September; 4 - 9 pm).

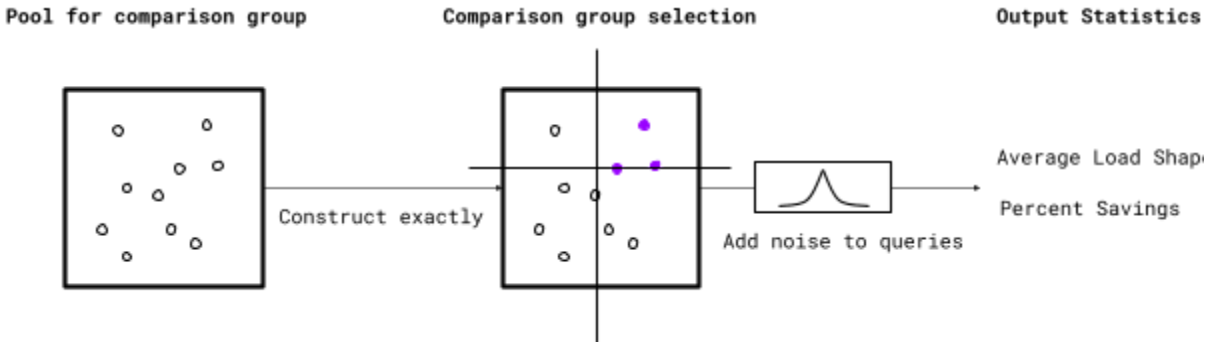
The automated sampling approach generates candidate comparison groups based on every possible stratification binning arrangement. The quality of the match between the treatment group and each candidate comparison group is then assessed across the full distribution of treatment group customers. This is done by breaking the treatment group into dozens of subgroups on the basis of total consumption. The average summer weekly load shape is computed for each subgroup. This procedure is also done for each candidate comparison group. The sum of squares (sum chi-squared statistic) is then calculated across all treatment and candidate comparison subgroup weekly load-shaped profiles. The final comparison group is selected as the group that minimizes this chi-squared summation. This process is fully automated via the GRIDmeter code.

Energy Differential Privacy Approach

The methods described here adopt the global model of differential privacy, adding noise to the final statistic value before output. That is, after calculating the statistic exactly, a random number is drawn and added to the statistic.

For comparison group statistics, data flows as follows:

- Starting from a pool of meters
- Select a stratified comparison group sample that best matches the treatment group
- Compute statistic from the comparison group
- Add noise to the exact statistic
- Return the noisy statistic



Comparison group statistic workflow

This model realizes a differential privacy guarantee for participants in the comparison group, as even if an attacker can manipulate the comparison group to guarantee a site's inclusion, a calibrated amount of uncertainty will remain for this site.

Procedure

This section documents the step-by-step anonymization procedure to realize a differential privacy guarantee of comparison group statistics.

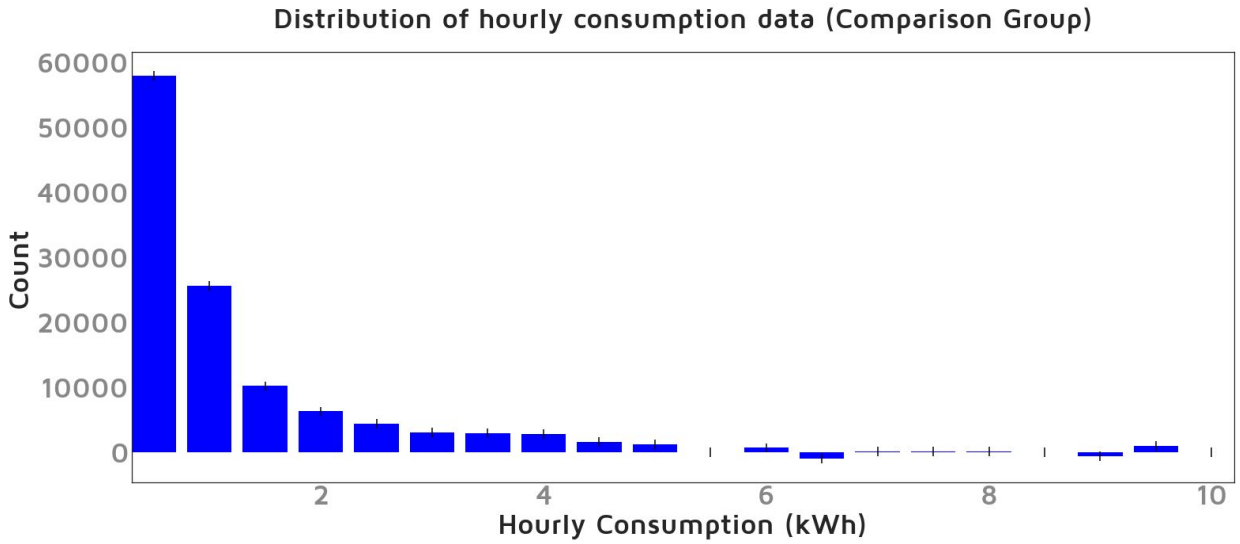
Choosing Clamping Bounds

For the private queries that follow, clamping bounds must be chosen to bound sensitivity. That is, a lower and upper bound for hourly consumption data must be specified upfront. The tighter the clamping bounds that are set, the more efficiently the privacy budget will be used. However, too tight of clamping bounds will introduce unpredictable error into downstream calculations.

For this dataset, we use the Sparse Vector Technique (SVT) to choose a clamping bound. The background on this approach is documented in *EEprivacy*: Choosing Clamping Bounds with the Sparse Vector Technique.

SVT was run against a series of clamped sum queries. Each query computed the difference between two clamped sum queries of total energy consumption with clamping bounds differing by 1 kWh. This results in a sensitivity of 1 kWh for the SVT queries. The target threshold was an error of less than 0.01 kWh average across 24 hours (or 0.24 kWh total across 24 hours).

To choose a privacy parameter for SVT, the distribution of hourly consumption values was found with a histogram query at ε =0.1.



The histogram query was used to generate a synthetic dataset to find the required SVT ε. It was found that an $\boldsymbol{\varepsilon}$ **=0.2** was required for accurate SVT (999/1000 trials agreeing on a clamping bound).

An SVT query for a clamping bound returned **6.0 kWh** to realize an error below a threshold of 1,652 kWh (0.01 kWh hourly).

Comparison Group Average Load Shape (Gaussian Mechanism)

A private load shape was constructed for the comparison group using the vector-valued Gaussian Mechanism.

Exact averages were computed for each hour for both predicted and observed consumption values, then Gaussian noise was added. Since the exact count of sites is non-private, it does

not contribute to the sensitivity. The sensitivity for each element of the private mean vector is identical:

$$\Delta = (U-L)/N$$

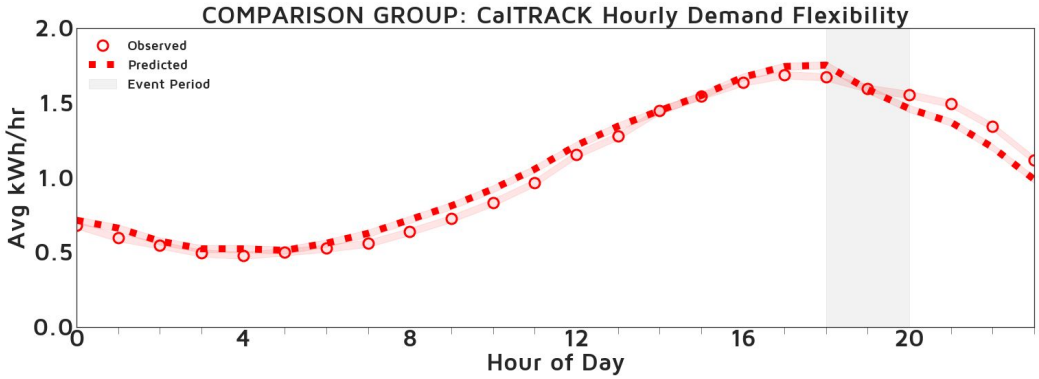
where *U* is the upper clamping bound (6 kWh), *L* is the lower clamping bound (zero), and N is the number of sites.

In our privacy model, the differential privacy guarantee is applied across all hours under analysis. Therefore, a vector-valued Gaussian Mechanism query was executed for 48 data points (24 each of observed and predicted consumption).

The *EEprivacy* library's *epsilon_for_confidence_interval* function was employed to choose an ε value to result in a 95% confidence interval of ±0.1 kWh.

The $\boldsymbol{\delta}$ parameter was set to $1/N^2 = 4.08 * 10^{-8}$

The comparison group (N=4948) required ε =4.0.



Percent Load Change (Laplace Mechanism)

Percent Load Change is defined as the following:

$$PercentLoadChange = \frac{\mu_{Predicted} - \mu_{Observed}}{\mu_{Predicted}} = \frac{\Sigma Predicted - \Sigma Observed}{\Sigma Predicted}$$

Where $\mu_{Observed}$ is the average of observed consumption values, while $\mu_{Predicted}$ is the average of weather-normalized consumption values (computed using the *EEmeter*), these averages are calculated across the event period (6 PM-9 PM).

Two sum queries are issued in total, one for each sum (predicted, observed) for the comparison group. The Laplace Mechanism was used for the sum query with a sensitivity of 18 kWh (6 kWh x 3 hours).

The required ε was determined stochastically to achieve approximately 1% (one percentage point) error in the final percent load change.

For the comparison group ε =1.25 per sum query.

The final results for the three hour event were:

Total Predicted Comparison	23602.2 kWh
Total Observed Comparison	24023.2 kWh
Total Predicted Treatment	4439.6 kWh
Total Observed Treatment	3662.2 kWh
Percent Load Change Comparison	-1.78±0.51%
Percent Load Change Treatment	17.5%
Difference of Load Changes	19.3%±0.51%

Implementation Notes

The algorithm design performed above utilized the open-source EEprivacy library for determining confidence bounds and choosing ε values.

Google's differential privacy Go library was used to determine the Analytical Gaussian sigma values and securely add noise. This secure noise source was employed primarily to mitigate the floating-point attack[1] against differential privacy.

Total Privacy Impact

We can tally up total privacy impact by Basic Composition Theorem, adding the ε for each output statistic, then accounting for privacy amplification by "Secrecy of the Sample" (since the population under study is a sample of a larger population, and it can be expected that

[1] Mironov, I., 2012. On significance of the least significant bits for differential privacy, in: Proceedings of the 2012 ACM Conference on Computer and Communications Security - CCS '12. Presented at the the 2012 ACM conference, ACM Press, Raleigh, North Carolina, USA, p. 650. https://doi.org/10.1145/2382196.2382264

the sample used for analysis will remain secret, then the privacy impact can be reduced by the fraction of the total population that this sample represents).

The total privacy impact for this scenario is summarized below:

	Comparison Group ε
Consumption Histogram	0.1
SVT for Clamping Bound	0.2
Comparison Group Load Shape	4.0
Total Predicted Comparison	1.25
Total Observed Comparison	1.25
Total Privacy Impact (Pre-amplification)	6.8
Amplification factor[2]	0.124
Total Privacy Impact	**0.843**

Privacy Impact of OhmConnect Comparison Group Use Case

Interpreting Epsilon

The large size of the comparison group, as well as the even larger size of the comparison pool, resulted in quite small overall ε values for comparison group participants. But what does "quite small" mean, exactly?

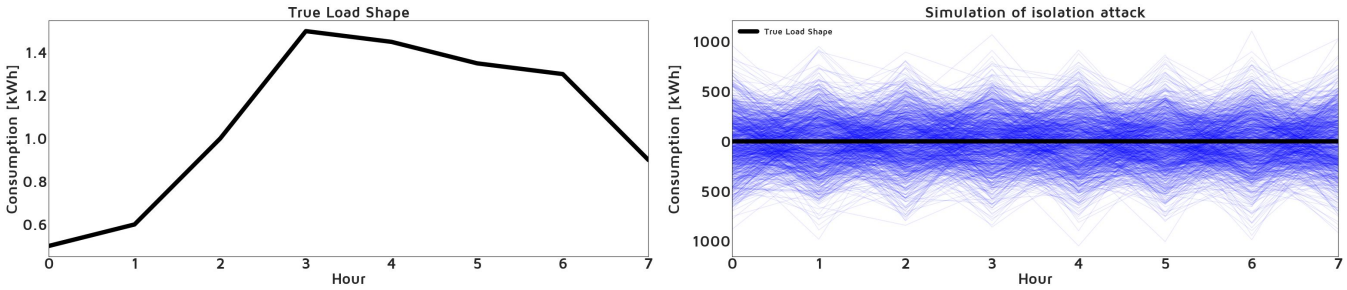
By way of comparison, Google's COVID-19 mobility data subjects users to a daily ε =1.76, while Facebook's COVID-19 Movement Range report subjects users to daily ε =2.0. Other applications, like LinkedIn and the US Census, set ε values in the range of 4 to 9.

These applications all consider significantly more sensitive datasets than the comparison groups use case outlined here. For example, the location data in Google's mobility report can directly lead to privacy harm for individuals. Energy data, on the other hand, does not directly lead to privacy harm. To illustrate this point, we simulate an isolation attack against the private load shape data.

[2] Differential privacy and the secrecy of the sample, 2009. Oddly Shaped Pegs. URL https://adamdsmith.wordpress.com/2009/09/02/sample-secrecy/ (accessed 10.19.20).

In an isolation attack, an attacker is assumed to know the energy consumption for all buildings in a dataset except one. Using this knowledge, the attacker subtracts out all but a single site's energy consumption, revealing an individual's data.

The figures below illustrate the best guess that an attacker would have of an individual's energy consumption for a query of ε =0.843 (the overall ε for OhmConnect comparison group participants) *if they had access to every other site's energy data in the treatment group.*



That is, even being able to isolate a single site, the noise added by the Gaussian Mechanism limits an attacker to a confidence interval of around ±1,000 kWh for hourly consumption. A house consuming this much energy would have to be literally melting tons of steel in a backyard furnace.

Interpreting the ε guarantee is an ongoing area of research with no final answer yet. So far, it appears that useful analytics tasks are possible at ε with strong privacy guarantees.

Conclusions

The method described in this document is the first integration of population comparison group methods with differential privacy. These results demonstrate that usable outputs are possible to achieve with strong privacy guarantees.

- We showed that it is possible to enable population control groups while rigorously protecting customer data utilizing differential privacy.
- It is possible to publicly share data on portfolio load shapes utilizing differential privacy methods on treated customer data.

The OhmConnect example described above shows that it is possible to overcome historical roadblocks created by outdated privacy rules to use population data to measure demand impacts while maintaining the privacy of all individuals at a level comparable to what Google uses to protect location data.

Energy Differential Privacy and the results of this essential and timely measurement of Ohmconnect's VPP are examples of how we can utilize our investments in smart metering infrastructure and data to solve important public and private problems. There are many use cases for the population data that can be unlocked using these methods, including learning more in our white paper.

Finding the Balance Between Privacy and Accuracy

Finding the right balance point between making data both private and useful, is an ongoing area of research. As privacy protection is increased, greater levels of noise are added to data. However, noise also introduces error, which can decrease the value of measured resources in the market.

The choice of ε is an ongoing area of R&D in the energy sector. The question is not simply a technical one, but a negotiation between the mathematical, legal, political, and social aspects of data privacy.

For this exercise, we erred on the side of caution, picking epsilon values representing a high degree of privacy protection (lower epsilon equates to greater privacy protection), setting the comparison group at ε =0.843. This resulted in a savings of 19.3%±0.5%.

Our initial simulation of privacy threats against energy usage data, and especially derivatives like savings or demand flexibility (load shape impact), suggest that differential privacy guarantees dramatically reduce privacy risk. With continued stakeholder input and reflective of the societal value vs. relative risk to customer privacy, as well as improved privacy engineering, we believe there is significant room to improve precision and still maintain very appropriate and conservative privacy protections.

Future Work

This work is just the start. Additional privacy engineering could make more efficient use of the privacy budget. One area for improvement would be to re-use the Load Shape queries for the Percent Savings calculations; the outputs needed to calculate all of these quantities could be issued in a single vector-valued GaussianMechanism query.

Another area for future privacy engineering concerns the Laplace Mechanism composition bounds. The Laplace Mechanism sum queries do not tightly bound the sensitivity of the Percent Load Change statistic.

Additional threat modeling could scope down the privacy impact from this worst-case scenario. The US Census, for example, finds empirical privacy loss to be a factor 10X less than worst-case privacy loss in almost all cases[3].

There is also a need for a policy debate about the balance of privacy, societal value, and true privacy loss. The data used in this analysis is a derived value called hourly resource curve or savings, which is a derived value that contains minimal personal information and a seemingly low risk in real terms to privacy.

[3] Petti, S., Flaxman, A., 2019. Differential privacy in the 2020 US census: what will it do? Quantifying the accuracy/privacy tradeoff. Gates Open Res 3, 1722. https://doi.org/10.12688/gatesopenres.13089.1

Appendix A: Measurement Methods

FLEXmeter: Measuring net impact to load-shape of DR events

The methodology relies on three open-source tools developed at Recurve:

- **EEMeter**, an implementation of the **CalTRACK** methodology
- **GRIDmeter**, a stratified sampling method for comparison group construction
- **Energy Differential Privacy**, differential privacy techniques for energy data

Set up a comparison group

- Obtain a population of participant (treated) and non-participant (non-treated) customers. Typically, non-participants will be a large set from the utility customer pool of which limited metadata is available, and data privacy restrictions are higher.
- Extract typical load shapes for each participant and non-participant.
- Using GRIDmeter, identify a subset of non-participants who have load-shape characteristics similar to those of the participating customers. This is the comparison group.

Compute baselines and counterfactuals

- Using EEMeter, fit the CalTRACK hourly model to each customer in the treatment and comparison groups, with the DR event period blacked out, to learn each building's temperature-dependent behavior and occupancy schedule.
- During the DR event, apply the measured temperature and time-of-week to the fitted model and predict the energy use. This is the meter's counterfactual energy usage during the event.

Compute population-level savings and correct with a comparison group

- Sum observed and counterfactual energy usage hour-by-hour to create population-level observed and counterfactual load shapes for both treatment and comparison groups.
- Compute percentage impact to load-shape hour-by-hour (i.e., percentage savings), for both treatment and comparison groups, using observed and counterfactual.
- Compute net population-adjusted percent impact to load shape hour-by-hour by subtracting comparison group percent impact from treatment group percent impact.

Privatize

- Compute an appropriate quantity of random noise to be added to the final percent impact numbers using Energy Differential Privacy techniques in EEprivacy.
- Publish a set of confidence intervals that contain the true results within the desired accuracy, e.g., 95%. The source data is guaranteed to remain private.