

An Evaluation of Demand Response Baseline Performance for Onsite Generation and High Frequency Event Participants

Ethan Barquest, Verdant Associates

Collin Elliot, Verdant Associates

Jean Shelton, Verdant Associates

Jahon Amirebrahimi, Pacific Gas and Electric Company

ABSTRACT

Demand Response (DR) programs have employed a common set of methods for customer compensation and estimation of baselines. These broadly include various day matching methods, such as x-in-10 baseline and weather matched baseline. In most cases, adjustment factors are used to shift the reference loads so they better align with actual event day profiles in the pre-event hours. While not perfect, these baselines are considered industry standard practice for non-regression settlement of DR.

Until recently, California self-generation and storage customers were excluded from DR participation and events were dispatched only a few times during the event season. As a result, the settlement baselines in California were designed and refined under these conditions. However, newer pilot programs, like PG&E's Supply Side II Pilot and Excess Supply Pilot allow participation from previously excluded customer types and dispatch events at higher rates than traditional DR programs. This evolving DR landscape has implications for the use of the existing baseline methodologies.

This paper presents findings from an evaluation of DR baseline performance focusing on self-generation, solar, storage and EV customers as well as customers with high frequency events. Baseline methods included in the examination include traditional x-in-10 and weather matching baselines as well as solar irradiance matching, meter generation output, and look ahead baselines. Key insights include strategies for DR settlement, handling of adjustment factors during times of negative load and alternative baselines for programs with high event frequency.

Introduction

In February of 2017, The California Public Utilities Commission (CPUC) issued [Decision \(D.\) 17-12-003](#), approving Pacific Gas and Electric Company's (PG&E) request to continue administering two demand response (DR) Pilots pilot programs from 2018 through 2020 – the Supply Side II DR Pilot (SSP II) and the Excess Supply DR Pilot (XSP), together referred to as the Pilots. The key objectives of the SSP II and XSP Pilots are multi-faceted and are a direct result of PG&E's commitment to utilize and integrate DR resources into the California Independent System Operator (CAISO) market and provide other services, like distribution reliability. The Pilots were also a response to the increasing complexity of the grid, where net load shapes are impacted by the deep penetration of disruptive end-use technologies, including solar and electric vehicles, that heighten the need for demand response during non-traditional times and intervals. The SSP II and XSP were developed to test various demand response program structures. To be successful, the Pilots must provide customers with incentives to participate and successfully and cost-effectively achieve the load changes desired by the Pilots.

Overall the SSP II and XSP consisted of three distinct programs that operated between 2018 and 2020. The descriptions of these programs include:

- **Supply Side II Pilot (SSP II):** The SSP II was designed as a continuation and expansion of the Intermittent Renewable Management Pilot Phase 2 and the Supply Side DR Pilot which studied the feasibility of demand-side resources participating in the CAISO wholesale market as Proxy Demand Resources (PDR) and enabled participation in wholesale real time energy and non-spinning reserve markets. The SSP II expanded on SSP by enabling DR resources to participate in CAISO market-based services and in 2020 the SSP II called events designed to address local distribution reliability issues, or simulated issues, for the distribution grid.
- **Excess Supply Pilot (XSP):** The XSP was intended to explore how customers can shift their load to align better with renewable integration, increasing load during periods of excess supply and reducing load during periods of grid constraint. The XSP is an out-of-market program because at the time the mechanism for bidding load-increasing DR into the CAISO market had not been developed. Results from the XSP, however, have been used for proposing how baselines for load increasing bids could be calculated and potentially allow DR to provide load-increasing bids in the future. XSP participants can elect to respond to load increasing events or bi-directional events.
- **Excess Supply Electric Vehicle Charging Network Pilot (XSP-EVCN):** The XSP-EVCN component of the Pilots provided site-hosts with a Load Management Protocol for site-hosts that did not want to charge their EV customers at TOU rates. The XSP-EVCN was designed to provide site-hosts with incentives to increase the load on their charging networks during periods of excess supply and reduce the usage of their charging networks during periods of grid constraints. The XSP-EVCN Pilot was initiated in 2019. The number of site-hosts enrolled in the XSP-EVCN expanded substantially in 2020, but the COVID pandemic likely impacted EV charging during this period.

Ultimately, the goal of this evaluation (as it pertains to this paper) is to explore baseline methodologies used for demand response settlement, assess their performance for various customer types, end use, and event types, and provide insights for high frequency event participation.

Baseline Assessment Methodology

During the life of the SSP II/XSP, the Pilots relied on the CAISO's 10-in-10 baseline to estimate impacts for settlement. In addition to the assessing the Program's settlement baseline, this analysis also included a 5-in-10 baseline, two weather matching baselines, and finally a metering generator output (MGO) to be considered only for the XSP-EVCN Program. All these baselines are based on the CAISO's methods and their calculation is covered comprehensively in the CAISO Business Practice Manual¹, so this paper will limit discussion primarily to the details baseline approaches defined shown in

Table 1.

¹ https://bpmcm.caiso.com/BPM%20Document%20Library/Demand%20Response/BPM_for_Demand_Response_V3_clean.pdf
for the CAISO Business Practice Manual.

Table 1. Baselines Used in Assessment

Baseline	Description
10-in-10	Average of the most recent ten days from the previous 45 of the same type (weekday vs. weekend/holiday) where no event occurred. An adjustment based on pre-event hours is applied, capped at 0.8 and 1.2. No adjustment applied when data results in a negative adjustment ratio
5-in-10	Average of the top five days based on event hour consumption from the 10-in-10 baseline. An adjustment based on both pre-and post-event hours is applied, capped at 0.8 and 1.2.
Temperature Matching	Average of the four days from the previous 90 with the closest daily maximum temperature to the event day. An adjustment based on both pre-and post-event hours is applied, capped at 0.8 and 1.2.
Irradiance Matching	Average of the four days from the previous 90 with the closest daily maximum irradiance to the event day. An adjustment based on both pre-and post-event hours is applied, capped at 0.8 and 1.2. This was added to XSP, where potential triggers might include excess PV/solar export.
Meter Generator Output (MGO)	Average of the most recent days of the same type where no event occurred in the hour being averaged. Essentially the 10-in-10 baseline where the event exclusion is by hour, not day. No adjustment is applied. Baseline applies for XSP-EVCN only.

Other than the details above, there are two details not covered by other baseline calculation documentation. First, for the 5-in-10 and weather matching baselines, the CAISO caps to adjustments are 0.71 and 1.4. The CAISO caps are set primarily for residential customers, so to be consistent with the 10-in-10 baseline, for this analysis, all baselines with adjustments use the 0.8 and 1.2 limits. Second, one distinct characteristic of the customers in these Pilots is the potential for net negative load, which can result in a negative adjustment ratio. In these cases, no adjustment is applied to the baseline.

Proxy/Event Day Assessment Day Definition and Selection

The assessment of the different baseline methods is based on comparing the observed load and calculated baselines on days where no event occurred but with event-like conditions, which this analysis refers to as proxy event days. For most demand response programs with less frequent events coinciding with extreme temperature events, proxy event days are typically the remaining hot non-event days in the year. The SSP II and XSP Pilots had, in some cases, more than a hundred event days per year, depending on the customers, which called for a slightly different approach to ensure that we captured all the conditions during which events can occur. For this analysis, we selected five weekdays and three weekend/holidays from each year and quarter of the available interval data based on the assignment of a random number to the available non-event days. The evaluation explored the use of proxy event days based on maximum temperature and irradiance, however, there was no correlation between the event days and weather conditions, atypical of many DR programs.

Proxy Event Hours

The calculation of the adjustment depends on an event's start and, for the 5-in-10 baseline, end time. Given that proxy event days are not actual events, some presumed event hours needed to be assigned to the proxy events to calculate the adjusted baseline. To illustrate how sensitive the baseline performance can be to the event hour, Figure 1 shows the 10-in-10 baselines calculated for the same day where the event is in hour ending 11 and 14, respectively. While the unadjusted baseline is the same for the two figures, due to a high amount of morning load, the hour ending 11 event gets a substantial adjustment, which results in a poor baseline fit during the event hours. In contrast, the hour ending 14 event receives only a small downward adjustment and the adjusted baseline lines up very closely to the observed load. Without a varied representation of event hours, the analysis would fail to capture these types of details.

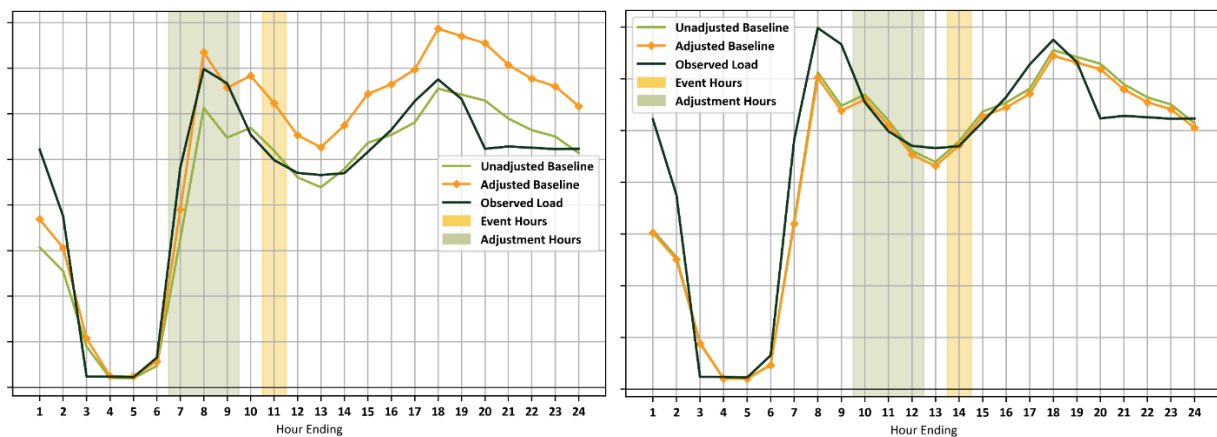


Figure 1. Proxy Event 10-in-10 Baseline with Event Hour Ending 11 (Left) and Hour Ending 14 (Right)

Given the clear importance of event timing, to ensure the event hours applied to proxy event days reflected realistic conditions, we relied on the actual event data to determine which hours to apply as event hours. The different permutations of event hours, though they varied by program and aggregator, are far too numerous to include comprehensively. However, the variability in event hours is presented in Figure 2 which shows the count of event hours by program, aggregator, and the hour ending for all program years.

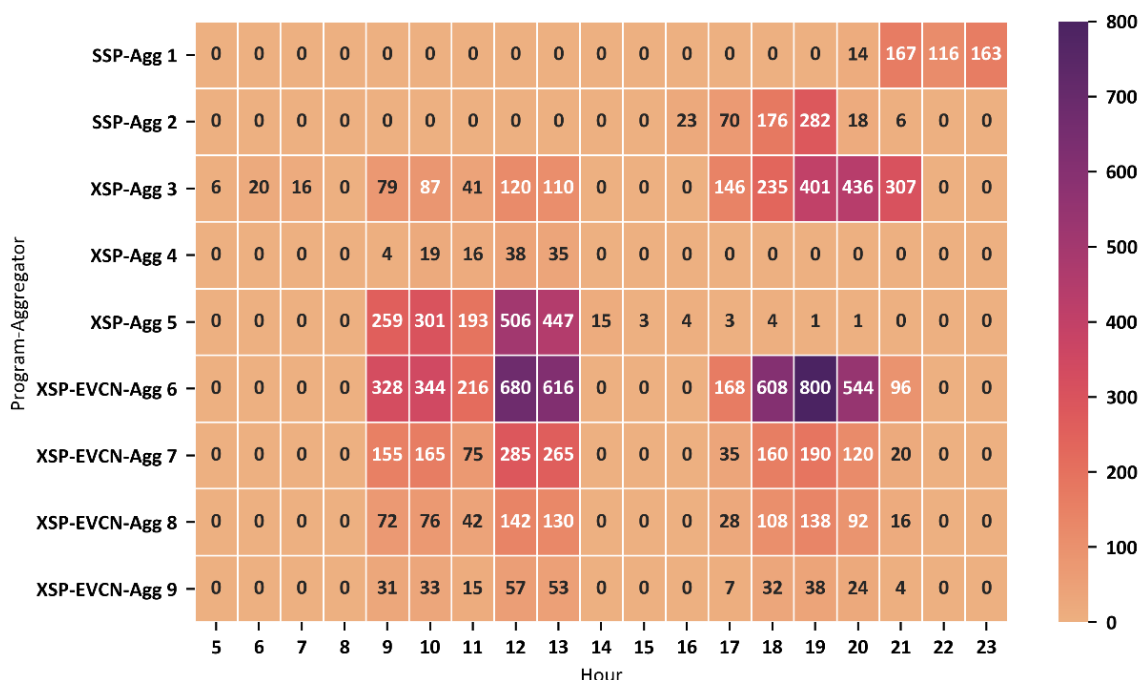


Figure 2. Event Hour Counts by Program and Aggregator

Despite the many event hour permutations, only the start and end hours are relevant to the calculation of the baselines. As a result, we used only the unique permutation of start and end times by program, aggregator, customer, and day type, excluding any that occurred fewer than five times. While this level of detail and rigor regarding the event hours is not typical and a smaller set of representative event hours might have sufficed, the only cost was computing time. Given the atypical variation in event hours, it means the analysis will provide a more complete view of baseline performance.

Performance Metrics

The assessment of these different settlement baselines is concerned primarily with accuracy and precision. Accuracy represents how closely on average the calculated baseline matches the observed load. A component of measuring accuracy is bias, which indicates the extent to which the calculated baseline over or underestimates the load. In contrast, precision indicates how reliably a baseline is close to the observed load. It is possible to have a baseline that is on average highly accurate with very poor precision, such as when a method both under and over predicts by substantial amounts with regularity. Likewise, it is possible to have a method that is very precise but highly inaccurate, such as when a model over or underestimates the load with high consistency. Of course, a baseline can also be neither accurate nor precise.

The primary metrics for accuracy and precision in this analysis are Normalized Mean Bias Error (NMBE) and Normalized Mean Absolute Error (NMAE), respectively. Other assessments of baselines have often used the Mean Percent Error (MPE) as the metric to assess accuracy and the Mean Absolute Percent Error (MAPE) and Coefficient of Variation of the Root Mean Square Error (CVRMSE) as the metrics for precision. For this analysis, while we calculated these metrics the focus of the analysis is on NMBE and NMAE.

Table 2 presents descriptions and the equations for the metrics calculated for accuracy and precision. One thing to note is that for the standard calculation of NMBE and NMAE, the formulas go against a convention seen in some contexts (e.g., ASHRAE²), where the error is calculated as the baseline minus the observed. This runs contrary to the more typical conventions of calculating MPE and MAPE, so for the sake of consistent interpretation of the NMBE and MPE, where negative values indicate overestimation of the baseline, we calculated the error as the observed load minus the calculated baseline for all metrics.

Table 2. Descriptions and Equations for Performance Metrics

Metric Type	Metric	Description	Equation
Accuracy/Bias	Normalized Mean Bias Error (NMBE)	Represents the normalized average bias in the calculated baselines.	$NMBE = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)}{\bar{y}}$
Precision	Normalized Mean Absolute Error (NMAE)	Represents that average of the normalized absolute error in the calculated baselines.	$NMAE = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)}{\bar{y}}$
	Coefficient of Variation of the Root Mean Squared Errors CV(RMSE)	Represents the normalized average of the squared errors between the observed load and calculated baselines.	$CV(RMSE) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{y}}$

Overall Baseline Assessment Results

The presented set of results is intended to answer the question of which of the available baselines methods is best for determining program impacts. This baseline assessment generated a large quantity of data, with tens of thousands of individual baselines calculated for the different programs, aggregators, and customers across a broad variety of different proxy events dates and event hours. Of the many ways to assess these data, only a fraction can be covered in this paper, so the approach is to examine baselines at the highest level of aggregation.

For this paper, the highest level of meaningful aggregation is a summary of the primary accuracy and precision metrics by program and year. These summaries of baseline performance are presented for SSP II, XSP, and XSP-EVCN in Table 3, Table 4, and Table 5, respectively. The results are based on the randomly selected proxy event days because analysis showed that there is no clear relationship between temperature or solar irradiance and events. While this lack of relationship is not typical of demand response programs, given the frequency of events and the full year availability of the Pilots, it is also not particularly surprising.

Note that the results are presented for both unadjusted and adjusted baselines. Reporting the unadjusted baseline is uncommon as it is usually an intermediary step in producing the final adjusted baseline. However, there were circumstances for customers in both SSP II and XSP in which reporting it is necessary. SSP II and XSP customers with on-site generation can be net exporters of energy, thus resulting in the calculation of a negative adjustment ratio. In this case, the unadjusted baseline should be used for estimating impacts. Such cases would occur

² ASHRAE is the American Society of Heating, Refrigerating and Air-Conditioning Engineers.

frequently depending on the customers and the timing of the events. For example, with SSP II in 2019, it occurred in nearly 12% of the events. For some customers that figure is higher.

An illustration of the importance of negative adjustment ratios is shown in Figure 3, which presents a 10-in-10 baseline for a proxy event starting at hour ending 12. For this event, the average of the observed load is negative during the adjustment hours, while the average of the unadjusted baseline is positive, resulting in the negative adjustment ratio. Based on the rules for the baseline, this negative adjustment will be capped at 0.8. However, when multiplied by the unadjusted baseline, it has the effect of making the adjusted baseline move farther away from the observed load, which is opposite to its intent. A similar issue results whenever an unadjusted baseline crosses zero, regardless of whether the adjustment has a negative ratio. In this case, with a positive adjustment, the positive baseline values would move up, and the negative would be adjusted further down. There are instances of this among the events, but analysis showed that in these cases, the load is so close to zero that the effects are not of great concern.

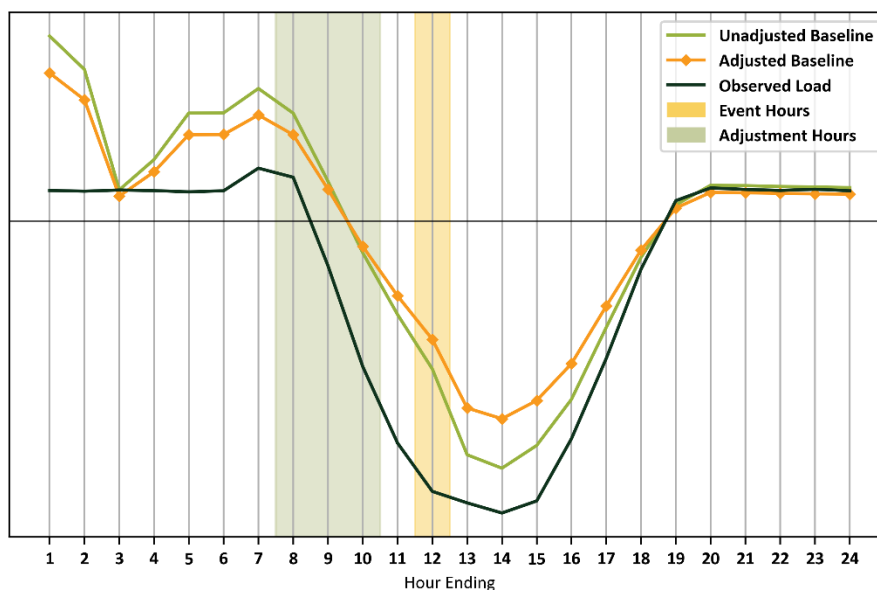


Figure 3. Example of Proxy Event with Negative Adjustment Ratio

SSP II Overall Summary

For the SSP II Program results in Table 3, with respect to accuracy, there is little evidence to differentiate the performance of the baseline methods based on the adjusted baselines. The program's 10-in-10 baseline is the most accurate in 2019 and 2020, with only the irradiance matching showing slightly more accuracy in 2018. With two exceptions, the differences among the other baselines are all small. The first is the 5-in-10 baseline, which is markedly less accurate in every year. The second is the 10-in-10 in 2020, where the other baselines were generally less accurate, resulting in the accuracy of the 10-in-10 standing out. Another factor to consider when comparing the accuracy of the 10-in-10 in 2020 to the accuracy of the baselines in other years is that in 2020 one aggregator discontinued their participation in the Pilot and another reconfigured its participation to reflect the location of sites on the distribution system. These changes undermine any direct comparison of the performance metrics in 2020 with other years. It is also worth noting that in most cases, the NMBE is negative for

both the adjusted and unadjusted baselines, indicating that the baselines are slightly overestimated.

Focusing again on the adjusted baselines, the precision metrics for SSP II show that the Pilot's 10-in-10 baseline performs well enough to be the most precise baseline using the NMAE in 2018 and 2020, though it is slightly outperformed by both weather matching baselines in 2019. Using the CV(RMSE) as the metric, the two weather matching baselines are the most precise in 2018 and 2019, but in 2020 are the least precise. This year-to-year inconsistency is suggestive of a great deal of volatility in these baselines, likely due to the underlying customers, though identifying the specific factors is difficult. Regardless, with few exceptions, the differences are all small and there is no compelling evidence in these data that the 10-in-10 baseline is not adequate for the SSP II Program.

Table 3. Overall SSP II Results by Year

Year	Baseline	Unadjusted Baselines			Adjusted Baselines		
		Accuracy	Precision		Accuracy	Precision	
		NMBE	NMAE	CV(RMSE)	NMBE	NMAE	CV(RMSE)
2018	10-in-10	-0.0069	0.0544	0.0782	-0.0064	0.0322	0.0432
2018	5-in-10	-0.0256	0.0932	0.1899	0.0147	0.0606	0.1554
2018	Irradiance	0.0006	0.0603	0.0856	-0.0046	0.0317	0.0427
2018	Temperature	0.0043	0.0551	0.0789	-0.0073	0.0329	0.0445
2019	10-in-10	-0.0094	0.0852	0.1697	-0.0054	0.0514	0.0788
2019	5-in-10	-0.0681	0.1054	0.1888	-0.0192	0.0528	0.0851
2019	Irradiance	-0.0028	0.0734	0.1440	-0.0077	0.0478	0.0691
2019	Temperature	-0.0063	0.0761	0.1496	-0.0086	0.0473	0.0719
2020	10-in-10	-0.0142	0.0871	0.1214	-0.0057	0.0378	0.0533
2020	5-in-10	-0.0856	0.1072	0.1509	-0.0272	0.0493	0.0694
2020	Irradiance	-0.0209	0.1086	0.1625	-0.0171	0.0541	0.0743
2020	Temperature	-0.0228	0.1007	0.1527	-0.0141	0.0557	0.0758

Looking at the unadjusted baselines, in 2018 and 2019, the weather matching baselines perform better (particularly irradiance) compared to the 10-in-10. For example, in 2018, the irradiance matching baseline had an NMBE of 0.001 compared to -0.007 for the 10-in-10. Whether it is truly a substantive difference, the advantage disappears in 2020, however, and it is worth noting that in 2020, none of the SSP II customers exhibited net export behavior.

The advantages in accuracy seen in the unadjusted weather matching baselines are far less with respect to precision. The 10-in-10 baseline performs slightly less well in a few cases, but the differences are marginal. The 2020 results again show marked contrast to the other years, with the 10-in-10 exhibiting better performance. Overall, these numbers still suggest that the 10-in-10 baseline performs well, but there is some evidence that in cases where many customers exhibit net export or have high PV production, there might be value in exploring the use of the irradiance matching baselines.

XSP Overall Summary

For the XSP (Table 4), irradiance matching consistently outperforms the accuracy of the 10-in-10 for both unadjusted and adjusted baselines. The temperature matching also has an advantage in some instances (including over the irradiance matching in some cases), but it is not as consistent. Compared to SSP II, the baselines are markedly less accurate, but the difference is starker when looking at the precision – or imprecision, in this case – of the baselines. When comparing precision metrics, the 10-in-10 performs much more closely to the other baselines and in some cases slightly outperforms them. There is a clear difference between the 2018, where the NMAE values are roughly half those from other years, which is likely related to the customers in those years.

Table 4. Overall XSP Results by Year

Year	Baseline	Unadjusted Baselines			Adjusted Baselines		
		Accuracy	Precision		Accuracy	Precision	
		NMBE	NMAE	CV(RMSE)	NMBE	NMAE	CV(RMSE)
2018	10-in-10	-0.0530	0.3082	0.4802	-0.0240	0.2586	0.4074
2018	5-in-10	-0.2073	0.3588	0.5881	-0.1372	0.2953	0.4809
2018	Irradiance	-0.0305	0.3113	0.4866	-0.0113	0.2676	0.4228
2018	Temperature	-0.0624	0.3279	0.4956	-0.0289	0.2757	0.4337
2019	10-in-10	-0.1208	0.5691	1.0745	-0.0876	0.4716	0.8297
2019	5-in-10	-0.4259	0.6865	1.2665	-0.3185	0.5682	0.9951
2019	Irradiance	-0.0774	0.5494	1.0364	-0.0481	0.4583	0.7818
2019	Temperature	-0.0419	0.6898	1.6421	0.0040	0.6056	1.3979
2020	10-in-10	-0.1102	0.5961	1.0334	-0.0386	0.4842	0.8642
2020	5-in-10	-0.3926	0.6675	1.1051	-0.2700	0.5225	0.8808
2020	Irradiance	-0.0982	0.6119	1.0540	0.0027	0.4732	0.8445
2020	Temperature	-0.0744	0.7178	1.3063	0.0214	0.5729	1.1173

It bears repeating that relative to SSP II, none of these baselines are very accurate and they are all far less precise. The customers in the XSP program have volatile loads due to a variety of factors (i.e., varying industries, storage/generation technologies, etc.) that limit the effectiveness of these and other baseline methods.

XSP-EVCN Overall Summary

Before discussing the results for XSP-EVCN in Table 5, it is important to note the following points:

- Based on the volatility of the load, the program did not apply an adjustment to the baselines used to calculate impacts. The adjusted baselines are included in these results to see how this might affect accuracy and precision.
- There is not an adjustment for the MGO baseline, so the metrics are the same.
- In comparing 2019 and 2020, it is important to keep in mind that:

- The number of customers more than doubled in 2020, so differences could be due to a variety of associated factors.
- While COVID-19 likely influenced customers in all programs, its effects are potentially more acute for these customers due to the greatly reduced driving or altered driving patterns.

Table 5. Overall XSP-EVCN Results by Year

Year	Baseline	Unadjusted Baselines			Adjusted Baselines		
		Accuracy	Precision		Accuracy	Precision	
		NMBE	NMAE	CV(RMSE)	NMBE	NMAE	CV(RMSE)
2019	10-in-10	-0.0596	0.5914	1.1597	-0.0321	0.5266	1.1549
2019	5-in-10	-0.4442	0.7791	1.4413	-0.3658	0.6884	1.4357
2019	MGO	-0.0825	0.5910	1.1417	-0.0825	0.5910	1.1417
2019	Irradiance	-0.0370	0.5792	1.1408	-0.0110	0.5131	1.1192
2019	Temperature	-0.0111	0.5592	1.0963	-0.0002	0.4908	1.0571
2020	10-in-10	-0.0040	0.6769	1.4216	0.0324	0.6052	1.5528
2020	5-in-10	-0.3924	0.8542	1.7161	-0.2878	0.7304	1.8425
2020	MGO	0.0044	0.6654	1.3971	0.0044	0.6654	1.3971
2020	Irradiance	-0.0408	0.7127	1.4825	-0.0014	0.6275	1.5860
2020	Temperature	-0.0069	0.7135	1.5120	0.0278	0.6319	1.5974

As seen in Table 5, none of the unadjusted baselines are very accurate and all have far less precision than even the baselines for the regular XSP customers. Having acknowledged that none of these baselines are “good,” there are a couple observations worthy of discussion. First, all the baselines are positively biased, which makes sense given how rarely the customers show any load from vehicle charging. On average there should be some charging load in the baseline days selected, but more often than not, the observed load was zero, which results in overestimation. A related finding is that in most cases, the adjustments did result in improved accuracy, and while the precision was scantily changed, it might be worth revisiting the use of the adjustments for future programs of this nature.

Second, the MGO baseline was added for two reasons. First, because charging stations represent one source of load, they are akin to the data from a sub-metered generation source, for which the MGO baseline was created. Second, because of the event frequency and how the MGO excludes data based on the event hour, and not the event day, the MGO has the potential to limit the use of “reserve days” in the baseline calculation. While the aspiration of adding the MGO was to identify a more useful baseline for these customers, the results are mixed. In 2019, the MGO was the second least accurate of the unadjusted baselines, though it was the second most precise. In 2020, however, it tied for the most accurate with the 10-in-10 and was the most precise baseline, whether adjusted or unadjusted. Despite these inconsistencies in the overall performance, the MGO is at least as good as the current 10-in-10 and under more regular charging patterns is likely a preferable baseline.

Look-Ahead Baseline for High Frequency Event Participant

As discussed previously, all three programs often had a high number of events, in some cases more than one hundred in a year. Furthermore, rarely were the events evenly distributed over the year, so for some periods of time customers had several weeks in a row where every weekday was an event day. For the 10-in-10 baseline, which is based on the most recent ten non-event days from the most recent 45-days prior to the event, this frequent and uneven distribution of events can present a problem. The baseline rules dictate that if there are not a minimum of five non-event days available, the difference is made up by incorporating “reserve days,” which are comprised of previous events. For a host of reasons, the inclusion of previous event days in the calculation of the baseline is not ideal, but in general it means that the baseline used to estimate impacts represents, at least in part, an event.

The sparseness of non-event days for these customers calls for reconsideration of the baseline’s restriction to days prior to the event. In cases where non-event days are limited, might it be preferable to allow the selection of days to “look ahead” as well? A non-event day in the future should be preferable to an event day in the past. To address this question, we conducted an analysis on a special set of proxy event days that occurred during periods of high event density. The key questions for this analysis were 1) how adversely do limited input days affect baseline performance, and 2) does permitting some number of look-ahead days mitigate these effects?

For the selection of proxy event days, we calculated a 30-day rolling sum of weekday event days to identify those non-event days that were preceded by many events. Any non-event day that had 20 or more events in the preceding 45 days was selected as a proxy event. The objective was not to find days that had entirely event days preceding them, but a mix that would allow for comparison of baselines under a variety of conditions.

As with the baseline performance analysis, the proxy event days were again paired with various sets of presumptive event hours based on the actual events. Next the 10-in-10 baselines were calculated using four look-ahead configurations where 45 look-behind days were combined with look-ahead options of zero, five, ten, and fifteen days. In these scenarios, zero is the current practice, which is the basis of comparison for the analysis. One of the reasons for look-behind restriction is that it allows baselines to be calculated in a timely manner, which is why this analysis maintained this priority and used a constant 45 rather than include scenarios with fewer look-behind days.

The comparison of the different look-ahead scenarios relied on the same set of metrics used to assess baseline performance. Figure 4 and Figure 5 present box plots of the accuracy and precision metrics, respectively, for unadjusted and adjusted baselines based on the four look-ahead configurations. Without going into an exhaustive discussion of box plots, the key features to know are that the horizontal green line represents the median and the box represents the interquartile range. The extended lines (“whiskers”) and the individual round markers indicate extreme values, so if they extend far from the box, there are extreme values skewing the distribution.

To assess these box plots, there are two main things to examine to determine the effect of the look-ahead days. First, does the median value on the box plot move closer to zero, which would indicate overall improved accuracy? Second, does the box representing the interquartile range become smaller, which would indicate a narrower distribution of errors in the calculated baselines.

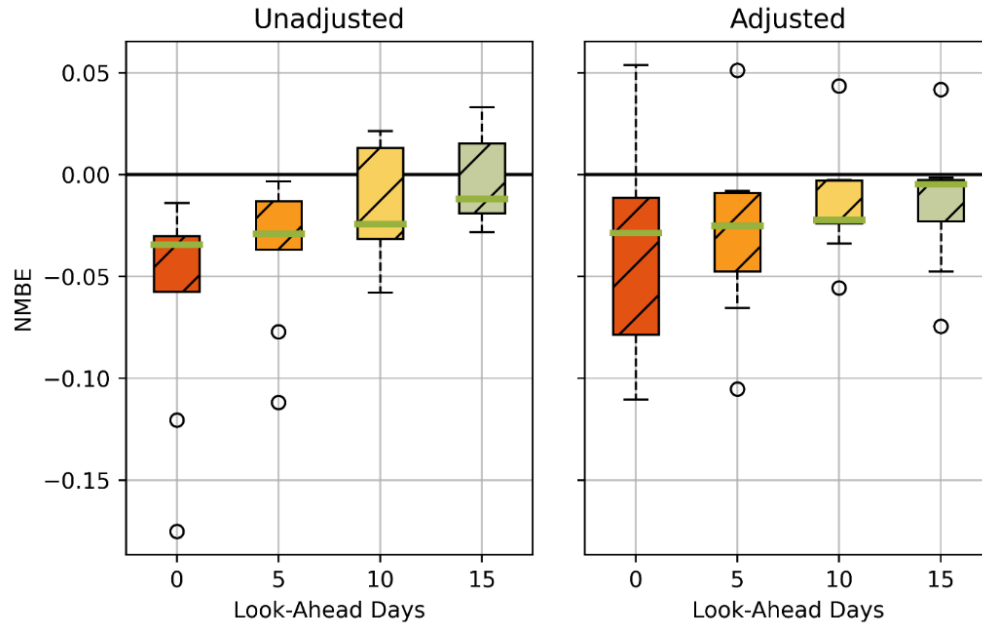


Figure 4. Boxplot of Accuracy (NMBE) for Look-Ahead Baselines

With respect to accuracy, Figure 4 shows that for both unadjusted and adjusted baselines, the results are unequivocal that allowing look-ahead days improves baseline performance. While the interquartile ranges are sometimes wider for the unadjusted baselines with look-ahead days, they progressively approach zero, even overlapping it. While the differences are not always large, the median NMBE moves closer to zero with each increase in look-ahead days for both unadjusted and adjusted baselines.

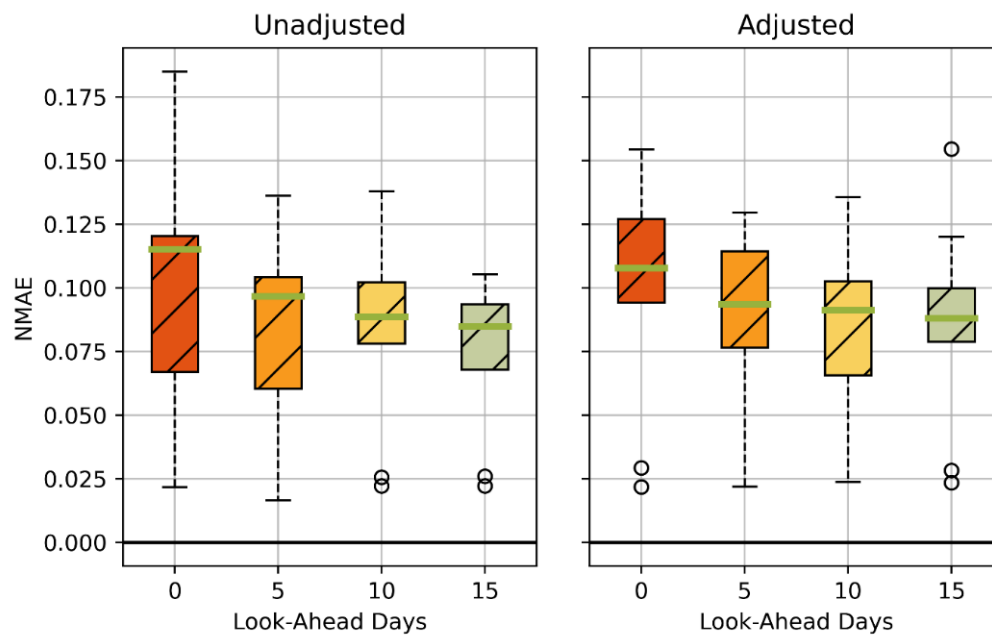


Figure 5. Boxplot of Precision (NMAE) for Look-Ahead Baselines

For precision, Figure 5 shows that incorporating look-ahead days improves the baseline performance. Here again one is looking to see if the median moves closer to zero, and for each addition of five look-ahead days there is improvement. The largest improvement comes from the initial five look-ahead days, which means that even if it is impractical to use 10 or 15 days, there is still benefit in modifying the baseline by adding 5 look-ahead days.

The event frequency observed in these programs is not typical for many DR programs, but for situations where the calculation of baselines is compromised by the presence of event days, this analysis suggests that there is a solution that would improve the performance of the baselines, and therefore the reliability of the impact estimates. Unless there is some compelling reason why postponing baseline calculation five, ten, or 15 days is untenable, the inclusion of look-ahead baseline days is worth consideration of future programs where there is the potential for a high number of events.

Summary

Utilities and ISOs will continue to see increases in DR participation of alternative DERs, such as battery storage, behind the meter self-generation and electric vehicles. Additionally, there are increased opportunities, especially in California, for DR aggregators to bid directly into wholesale energy markets such as CAISO that may result in more frequent dispatches of DR resources. As a result, there is a need to explore alternative settlement baselines to accommodate accurate estimates of load impacts for these resources. In this paper we explored traditional baseline calculations used for DR as well as new approaches such as solar irradiance baselines and look-ahead baselines. While there is evidence that traditional baselines still provide valid estimates of baselines, there is also evidence that improvements can be made in the baseline calculations for customers and aggregators with significant solar PV production and for high frequency event participants.

Findings

The key takeaways and findings include:

- The overall baseline performance summaries by program and year shows that there is scant evidence of any major shortcomings in the 10-in-10 baseline relative to the alternative methods. This is a positive finding because it means that the Pilots used using adequate baseline for settlement.
- The 5-in-10 baseline is not a good overall baseline for these programs, with substantial overestimation of the load during event hours. The key factor at play is that the 5-in-10 is a baseline conceived for programs where high temperatures are event triggers and the customers exhibit weather-sensitivity. For these Pilots, the numerous events are not associated with extreme heat and not all customers have large amounts of cooling. For examples where the 5-in-10 performs best, one has to look at the proxy event days selected based on maximum temperature and focus on specific customers, and even in these cases, the adjustments typically mitigate any advantage to the 5-in-10.

- A less clear conclusion is for the potential for the weather matching baselines, particularly when based on irradiance. These baselines perform similarly to the 10-in-10 and have the advantage of using 90 days of historical data compared to 45 days, which can prevent the need to rely on previous event days in cases of high event frequency. The irradiance baseline was included based on the theory that PV generation adds an additional element to load. For customers where generation is a large component of the net load, trying to match the day's generation could help improve the baseline. For the XSP, it was also hypothesized that the high generation could be a trigger for events. The causality of this is hard to prove, but the performance metrics do indicate that these baselines have potential.
- The inclusion of look-ahead days in the pool of settlement baseline reserve days clearly improves baseline performance for DR participants with high frequency events. While one potential recommendation would be to have fewer events to improve existing baseline performance, the look-ahead baselines should be considered for programs that aim to dispatch resources frequently.

References

- California Independent System Operator (CAISO). 2020. *Business Practice Manual for Demand Response*. Folsom, CA.
https://bpmcm.caiso.com/BPM%20Document%20Library/Demand%20Response/BPM_for_Demand_Response_V3_clean.pdf
- California Public Utilities Commission. 2017. *Decision Adopting Demand Response Activities and Budgets for 2018 through 2022*. Decision 12-12-03.
<https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M202/K275/202275258.PDF>
- Verdant Associates. 2021. *PG&E Supply Side II (SSP) and Excess Supply Pilot (XSP) Program Evaluations: 2018, 2019 and 2020*. Unpublished Report.