Energy Storage and Distributed Energy Resources Stakeholder Initiative Phase 2 ("ESDER 2")

Second Revised Straw Proposal

September 19, 2016

Market & Infrastructure Policy

Table of Contents

Energy Storage and Distributed Energy Resource Stakeholder Initiative Phase 2 ("ESDER 2")

Second Revised Straw Proposal

1 Executive Summary

The central focus of the ISO's energy storage and distributed energy resources ("ESDER") initiative is to lower barriers and enhance the ability of transmission grid-connected energy storage and the many examples of distribution-connected resources (i.e., distributed energy resources or "DER") [1] to participate in the ISO market. The number and diversity of these resources are growing and they represent an increasingly important part of the resource mix. Integrating these resources is expected to help lower carbon emissions and add operational flexibility.

The ESDER initiative is rather unique in that it is an omnibus initiative covering several related but distinct topics. For the second phase of ESDER (i.e., "ESDER 2") these topic areas include non-generator resources, demand response, multiple-use applications, and station power for storage resources. Another unusual attribute of ESDER 2 is the use of multiple approaches to pursue and address each topic area. For instance, in the case of the demand response topic area two stakeholder-led working groups – the Baseline Analysis Working Group and the Load Consumption Working Group – were established to discuss and recommend stakeholder-

[1] Distributed energy resources are those resources on the distribution system on either the utility side or the customer side of the end-use customer meter, including rooftop solar, energy storage, plug-in electric vehicles, and demand response.

desired enhancements to proxy demand resource (PDR). The proposals produced by these two working groups are not ISO proposals, but are the work product of the respective working groups. More recently, a working group has been established within the non-generator resources topic area to explore use-limitations for storage resources. An entirely different approach is being used for the remaining two topic areas of ESDER 2 – multiple-use applications and station power for storage resources – wherein the ISO is continuing its efforts to address these two topic areas through its participation in the CPUC's energy storage proceeding.[2]

In this second revised straw proposal the ISO presents the latest status of its work with stakeholders in addressing the four topic areas of ESDER 2. Although proposals in the four topic areas are not yet developed sufficiently to present the ISO Board of Governors for approval, substantial progress has been made in each topic area to varying degrees.

2 Changes from revised straw proposal

The ISO received comments from stakeholders in all topics areas addressed in the July 21 revised straw proposal – NGR enhancements, demand response enhancements, multiple-use applications, and distinction between charging energy and station power.[3]

The following is a summary of the changes from the straw proposal in consideration of these comments.

NGR enhancements – In ESDER 2 the ISO is working with stakeholders to understand and consider NGR modeling enhancements that best reflect resource use limitations and use characteristics for NGR modeled resources. The ISO received input that resources modeled under NGR should be considered for use-limited resource status. The ISO also received stakeholder input on other use limitations, such as, annual charge and discharge limits, physical MW limits based on time of day, and daily limits on cycling, with the ability to change these throughput limitations on a daily basis. The ISO recommended in the July 21, 2016, revised straw proposal that a stakeholder working group be established to discuss and further understand how NGR resources might be qualified and treated as use-limited resources and whether this has merit. This working group is using the progress made in the ISO's Commitment Cost Enhancement 3 (CCE3) initiative as a foundation for how the ISO defines use limited

---

[2] CPUC Rulemaking 15-03-011.

[3] Stakeholder comments on the July 21 revised straw proposal were submitted by the California Department of Water Resources (CDWR), California Energy Storage Alliance (CESA), California Large Energy Consumers Association (CLECA), Calpine Corp., LS Power Development (LS Power), Pacific Gas & Electric (PG&E), SolarCity, and Southern California Edison (SCE).

resource qualification and subsequent treatment of those resources. The working group held its first meeting via web conference on September 13, 2016.

In the area of NGR modeling enhancements to better reflect performance based on state of charge (SOC), the ISO has had an opportunity to work further with stakeholders and battery manufacturers to better understand and determine SOC impact on ramping and megawatt throughput.

Demand response enhancements –The Load Consumption Working Group (LCWG) has added clarifications and simplifications to its PDR load consumption proposal. Specifically, performance measurement for load consumption will be based on a modification of existing PDR performance measurement practices. The LCWG is abandoning further development of an ISO wholesale market daily load shift product, but will retain the concept of "bi-directional" PDR. Finally, the LCWG has reconsidered energy settlement for PDR frequency regulation and will support this option, in addition to its prior support for a zero net energy regulation option.

The Baseline Analysis Working Group (BAWG) has narrowed its research and is pursuing changes and updates in the following three areas: (1) use of alternative traditional baseline methods to estimate the load impact of current demand response resources; (2) options for using control groups rather than traditional baselines to estimate the load impacts of demand response resources; and, (3) ways to accurately measure load impacts of resources that are frequently dispatched.

Multiple-use applications – Since the straw proposal, the ISO has continued its efforts to address multiple-use applications through its participation in the CPUC's energy storage proceeding.[4] The ISO and CPUC began a collaborative stakeholder process on this subject with a joint workshop held on May 2-3, 2016, at the CPUC to address multiple-use applications and station power. Many stakeholders made informative presentations at the workshop, and the CPUC and ISO received extensive written comments on May 13 and reply comments on May 20. Based on the workshop presentations and the submitted comments the ISO has not identified any issues or topics that should be addressed in a separate effort under ESDER 2. If further activities in the CPUC proceeding identify issues that require treatment in an ISO initiative or develop proposals appropriate for ISO consideration, refinement and possible adoption, the ISO will consider these in its stakeholder initiatives catalog and roadmap for 2017.

Distinction between charging energy and station power –Stakeholders continue to support the ISO's proposal on station power and thus the ISO retains its proposal in this paper. At this point

[4] CPUC Rulemaking 15-03-011.

interest in this topic generally has turned to issues that are jurisdictional to the state and the utilities' retail tariffs (e.g., permitted netting, station power metering configurations). Since its straw proposal, the ISO has proposed that it will seek approval to amend its tariff to remain consistent with the utilities' retail tariffs should they change so as to avoid over- or under-lapping charges and settlements between the wholesale and retail markets. The ISO discusses these issues below to help the ISO and stakeholders speculate on what such rules could look like. The ISO also reiterates that it agrees that additional guidance is needed on station power rules both for storage and for conventional resources. In addition to the papers produced through this initiative, the ISO will revise its BPMs at the conclusion of this initiative.

3 Background

The central focus of the ISO's ESDER initiative is to lower barriers and enhance the ability of transmission grid-connected energy storage and the many examples of distribution-connected resources (i.e., distributed energy resources or "DER") [5] to participate in the ISO market. The number and diversity of these resources are growing and they represent an increasingly important part of the resource mix. Integrating these resources is expected to help lower carbon emissions and add operational flexibility.

In 2015 the ISO conducted the first phase of ESDER ("ESDER 1")[6], which made progress in enhancing the ability of storage and DER to participate in ISO markets. The ISO worked with stakeholders to develop policy proposals, and those triggering the need for tariff change – enhancements to the non-generator resources model and enhancements to demand response performance measures – were approved by the ISO Board of Governors at its February 3-4, 2016 meeting and tariff changes filed with FERC on May 18, 2016.[7] On August 16, 2016, the FERC accepted the ISO's tariff revisions effective October 1, 2016, as requested.[8]

[5] Distributed energy resources are those resources on the distribution system on either the utility side or the customer side of the end-use customer meter, including rooftop solar, energy storage, plug-in electric vehicles, and demand response.

[6] More information about the first phase of the ESDER initiative may be found at: http://www.caiso.com/informed/Pages/StakeholderProcesses/EnergyStorage_DistributedEnergyResourcesphase1.aspx.

[7] The ESDER 1 tariff filing may be found at: http://www.caiso.com/Documents/May18_2016_TariffAmendment_ImplementEnergyStorageEnhancements_ER16-1735.pdf

[8] http://www.caiso.com/Documents/Aug16_2016_LetterOrderAcceptingTariffAmendment_EnergyStorage_DistributionEnergyResourceInitiative_ER16-1735.pdf

In 2016 the ISO began conducting the second phase of ESDER ("ESDER 2") to continue this important work and explore additional topics of interest to stakeholders.

In the March 22, 2016 issue paper, the ISO proposed that ESDER 2 comprise the following topic areas: further NGR model enhancements, further demand response enhancements, further work on multiple-use applications, clarify station power for energy storage, and review the allocation of transmission access charge to load served by DER.

In the May 24 straw proposal paper, the ISO refined the scope of topic areas being addressed in ESDER 2 and clarified its proposed direction on these topic areas based on stakeholder feedback (e.g., feedback received from both written comments and the joint workshop held with the CPUC).

In the July 21 revised straw proposal the ISO made further refinements to the topic areas in scope and made progress in framing the issues and developing proposals to address those issues.

In this second revised straw proposal, the ISO presents the latest status of its work with stakeholders in addressing the four topic areas of ESDER 2.

4 Second Revised Straw Proposal

4.1 NGR enhancements

During the July 28th stakeholder web conference and in the subsequent written stakeholder comments, the ISO received valuable inputs to help inform and direct the focus on areas for improving the non-generator resource (NGR) model. The ESDER 2 initiative identified two areas that the ISO is proposing to explore for NGR enhancement: (1) representing use limitations in the NGR model, and (2) representing throughput limitations based on a resource's state of charge. The ISO uses this second revised straw proposal paper to further clarify these areas of NGR enhancement and refine the proposals to focus on facilitating enhancements that provide the highest value to non-generator type resources.

4.1.1 Represent use limitations in the NGR model

Representing use limitations in the NGR model continues to be a high priority among stakeholders. Stakeholder comments focus on two areas of use limitations for NGR enhancements.

The first area seeks to allow NGR modeled storage resources to qualify as a use-limited resource. The ISO tariff defines a use-limited resources as "a resource that, due to design considerations, environmental restrictions on operations, cyclical requirements, such as the need to recharge or refill, or other non-economic reasons, is unable to operate continuously." This use-limited resource status is available for certain generating resources that are able to define commitment costs, such as start-up costs, minimum load costs, and minimum megawatt hour run time for market optimization and bid cost recovery. A resource can be flagged as use-limited in the ISO market if it meets the current definition, completes the application/registration process, and provides an annual use plan. However, stakeholders should be aware that the use-limited concept is in the midst of an evolution regarding the definition, application process, and market treatment of use-limited resources. For example, the CCE3 initiative modified the definition of use-limited resource.[9] While the CCE3 initiative defined rules for storage modeled as a proxy demand resource (PDR), it did not consider storage modeled as an NGR but deferred to the ESDER 2 initiative.

The second area of interest is in looking at annual charge and discharge limitations, physical MW limits based on time of day, depth of cycling, and daily limits on cycling, with the ability to change these throughput limitations on a daily basis. As was presented in the July 28th stakeholder call, the ISO believes that limitations for total charge and discharge, or depth and frequency of cycling, are best tracked and managed by the resource owner to meet their resource specific performance guarantees and operating profiles. The ISO's market systems are not designed to track cumulative NGR performance parameters on an individual resource level. In the area of MW limits, the ISO believes that this capability already exists by allowing resource owners and scheduling coordinators to submit operational profiles in the ISO Outage Management System (OMS).

The ISO recommended in the July 21 revised straw proposal that a working group be established to discuss and further understand how NGR resources might be qualified and treated under use-limited designation and whether there is merit in this approach. The working group held its first meeting via web conference on September 13, 2016.[10]

---

[9] The CCE3 proposal has been approved by the ISO Board of Governors. Submittal of tariff revisions to FERC to implement the proposal is pending.

[10] The agenda and presentation used during this meeting may be found on the initiative webpage at http://www.caiso.com/informed/Pages/StakeholderProcesses/EnergyStorage_DistributedEnergyResourcesPhase2.aspx

4.1.2 Evaluate model enhancements based on reduced MW throughput at high and low state of charge

With the help of stakeholders, the ISO is developing a better understanding of the issue of storage performance limitations and non-linear degradation based on SOC and depth of cycling. Stakeholders have suggested that one option may be to submit multiple bid stacks where the most recently available resource SOC would be the determining factor on which bid stack was used at real-time execution. However, it is the ISO's current view that this area deserves further observation once more storage resources enter the market and the storage industry develops experience in storage modeling and management before developing a specific proposal. With so few NGR resources operating in the ISO market, the ISO proposes to re-evaluate the NGR model capability for improvements once more resources are participating in the ISO market.

4.2 Demand response enhancements

The ISO recommended in the March 22 issue paper that stakeholder-led working groups form to discuss and recommend stakeholder-desired enhancements to proxy demand resource (PDR). Since then, two stakeholder-led working groups have formed and are actively vetting two particular enhancements. The Load Consumption Working Group (LCWG) is exploring the ability for PDR to consume load based on an ISO dispatch, including the ability for PDR to provide regulation service. The Baseline Analysis Working Group (BAWG) is considering additional baseline evaluation methods to assess the performance of PDR when application of the current approved 10 of 10 baseline methodology is not sufficiently accurate.

Both of these issues – enabling directed load consumption and instituting new performance evaluation methods – require a thorough vetting by stakeholders with special end-use customer and retail ratemaking expertise. Incorporated here for broader stakeholder review and input are the revised straw proposals of the respective working groups. These are not ISO proposals, but are the work product of the respective working groups.

4.2.1 Load Consumption Working Group (LCWG) proposal

In this second revised straw proposal, the LCWG provides some further development of the two remaining elements of load consumption and frequency regulation for PDRs.

Beyond the progress made from the initial straw proposal, updates are provided in the following areas:

- A general consensus as well as an opinion from the CAISO legal department that the wholesale and retail components of PDR consumption as discussed are properly separated.
- An illustration of how baseline measurement for load consumption can work.
- A listing of the remaining issues to be resolved either through this stakeholder process or in the development of business requirements prior to implementation.

To this end, the enhancements of load-increasing PDRs and of PDR Regulation provision are developed further herein. Note that some of the detailed working group discussion that was included in the initial straw proposal is eliminated since, while it informed subsequent working group discussion, it is no longer necessary to describe the current proposal.

4.2.1.1 **Load Consumption**

4.2.1.1.1 Opportunity

Market resources should be able to compete to provide value to the grid through price-signals. A key limitation with the PDR design results from its focus solely on demand reduction, rather than a focus on both reducing and increasing demand. Recognizing that oversupply of generation has already resulted in periods of low prices in the middle of the day, there are benefits from incenting additional demand during key periods from as many resources as possible. Growth in load consumption during periods of excess supply could also benefit California by reducing the need to curtail renewable generation.

4.2.1.1.2 ISO Product Construct

This construct would require a provision for both consumption and curtailment PDR with bi-directional bidding where a single resource is able to offer both consumption and load reduction bids under the same resource ID, a functionality already included in the CAISO market for NGR, which allows simultaneous bi-directional bidding.[11] Thus, this same functionality could be applied to PDR without extensive market development. This bi-directional construct would be needed to support load consumption by demand response resources that also curtail as well as bi-directional frequency regulation.

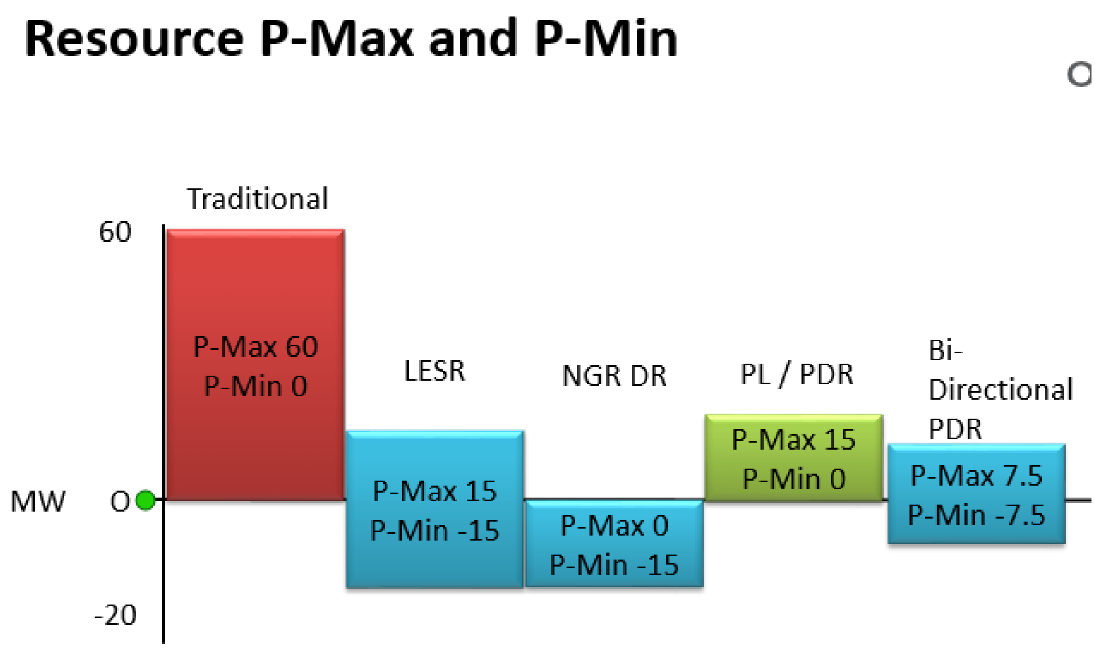
---

[11] Non Regulation Energy Management (Non REM) NGRs can submit both supply and demand bids under a single resource.

The construct could require that a bi-directional PDR establish a "mid-point"[12] to establish a demarcation between supply and consumption based on directional capability which could also require a "split" baseline for energy measurement if the resource chooses. The resource range is likely a parameter that would be set in the Resource Data Template (RDT), allowing it to be modified periodically rather than being a daily bidding element.

Traditional generators are defined within a range of zero as a minimum to a positive number as a maximum. When the ISO developed NGR for Limited Energy Storage Resources (LESR), it introduced the concept of resources with a range from negative to positive and at the same time contemplated that NGRs comprised of demand would have a range from a negative value to a maximum of zero. PDRs and participating loads have their load reduction capability "inverted" so they can be modeled and treated the same as traditional generation. The extension of the LESR to PDR would allow the statement of a range that would accommodate both additional consumption (negative) and reduction (positive) quantities. In the figure below, a PDR with 15 MW of dispatchable range could set half as additional consumption and half as reduction.

[12] While referring to this element as a mid-point, it would not need to be symmetrical since a PDR might have more capability in one direction than the other (e.g. drop more load for supply since it could include processes and house loads while additional consumption might be limited to adding processing loads).



The point of demarcation for NGR supply and demand bids is energy discharge for supply and energy consumption for demand. Currently PDRs are modeled to invert a reduction in load to appear to the market systems as positive generation based on their behavior and that performance is measured against "normal" consumption (baseline) to the consumption when dispatched (event). This construct can be maintained (and even exclusively if there were a use case for a "load increase only" resource) for increases in load as well. Current performance evaluation methodologies could be extended to PDR that includes load consumption and curtailment PDR.

Just as it is for load reductions, the PDR construct, whether it is applied to "traditional" demand or BTM storage, is an appealing model for instructing additional consumption since the model segregates the roles of scheduling the underlying load from the bidding of the load response capability in the wholesale market. Additionally, the model allows for the aggregation of customers' load response. To deviate from the PDR construct and not allow load consumption to be bid and dispatched by a third party into the wholesale market would either limit participation to the incumbent LSE or raise a set of issues that have not yet been resolved.

Therefore, the working group proposes that the CAISO modify its tariff and all relevant practices and procedures to allow PDR resources to place bids for both demand curtailment and demand consumption. The working group also believes that current performance evaluation methodologies available for demand reduction can be used for load consumption, albeit with the direction reversed, with one needed change: that PDRs be allowed to have a resource minimum value below zero, i.e. a negative value that indicates additional consumption.

4.2.1.1.3 Jurisdictional Issues

In developing this market enhancement, the legal authority by which the ISO, regulated by the Federal Energy Regulatory Commission (FERC), "directs" market behaviors such as load-consumption, even when the activity seems wholly unrelated to transmission or the sale of energy for resale (which generally are viewed as setting the parameter of the FERC's domain under the Federal Power Act), must not interfere with the right of the state to regulate retail rates. Additional consumption on a retail meter that results from a wholesale market dispatch will be recorded as retail consumption. The end-use consumer would pay retail prices for load consumed. The ISO would settle wholesale energy at the wholesale market clearing price, positive or negative. The bid to consume load will simply be a price the bidder is willing to pay or be paid for energy and will be settled in the wholesale market through a Scheduling Coordinator independently from the retail settlement. The bidder could, for example, structure a negative bid, which means the bidder expects to be paid for consumption of energy if negative bids are in the money and clear the market in certain intervals. There is no presumption of "capacity-like" payment to address the challenge of excess energy and over-supply in the forward planning horizon as there is no payment like "installed capacity" or resource adequacy capacity, which are not CAISO wholesale products. Such capacity is currently procured bi-laterally in California.

Subsequent to the revised straw proposal was circulated, CAISO Legal opined that it did not believe the ISO dispatching load consumption would be a practice FERC would reject. FERC has jurisdiction in two ways: 1) over reliability and 2) over practices that directly affect rates. Order 745 affirmed that a wholesale transaction doesn't have to be a "sale for resale," but a practice that *directly* affects wholesale rates. Load consumption follows this same logic, which was affirmed by the U.S. Supreme Court in EPSA vs. FERC.

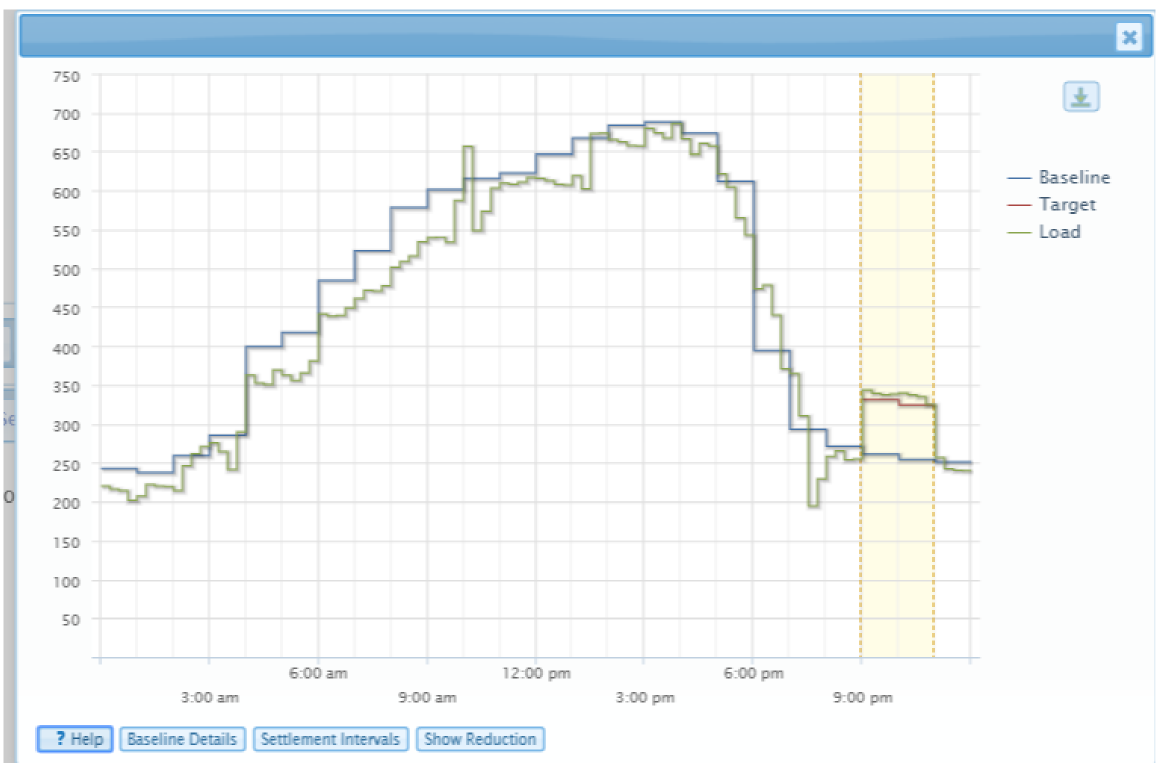
4.2.1.1.4 Enhancement Concept/Design

A requirement for this enhancement is to support performance evaluation methodologies for 'increasing load' (purposeful additional consumption) dispatches. This can build on the existing performance evaluation methodologies. For example, the methodology for load increases can be the methodology for load reductions reversed. These performance evaluation methodologies help differentiate and compensate wholesale behavior from retail behaviors and settlement, a key challenge with PDRs.

The payment for load consumption is in almost all ways just the inverse of demand reduction participation in wholesale markets. Any discussion of jurisdictional issues or some kind of settlement against the retail meter needs to specify why the treatment of load consumption is different than existing rules for demand reduction. A "negative" baseline has been implemented successfully in the PG&E Excess Supply Pilot (XSP) without modifications to the existing processes necessary to collect retail meter data, convert to SQMD and calculate performance.

For this and any aggregation of locations for the makeup of PDR, the operative assumption is that a customer location can only be associated with one aggregation and PDR at any single point in time. There have been previous discussions that at some point single locations could concurrently be associated with multiple aggregations/resources but it isn't clear if this opportunity would be developed in a timeline to support the elements of this straw proposal.

A calculation of a "negative" baseline for additional consumption is illustrated below measuring a load increase dispatch. The dispatch hours are HE 22 and HE 23 with the baseline for both hours 250kW and a dispatch quantity (Target) of 70kW. Performance is spot on with the actual load during the event at 320kW.



While it is straightforward to calculate a single direction baseline (either consumption as illustrated above or traditional curtailment) calculating performance using a 10 of 10 baseline with a same day adjustment would likely require some modifications and accommodations.

4.2.1.2 **PDR Frequency Regulation**

4.2.1.2.1 Opportunity

Extending frequency regulation participation to PDR would allow a set of DER deployed resources to bring their capability to a regulation market that is ripe for improvement. As more new technologies are being deployed behind the meter, tapping into storage and other resources that can rapidly respond to an automatic generation control (AGC) signal can serve to increase ISO control performance results. The fleet of regulation resources fell short of reasonable performance as evidenced by the year one pay for performance enhancements which resulted in a reduction of 50% performance to 25% performance before sanctioning a resource. The current ISO frequency regulation market provides a level of revenue through capacity and mileage payments that possibly support the additional technology costs of telemetry for a PDR that could participate. Moreover, allowing PDR resources to provide regulation may improve the competitiveness, depth, and liquidity of ISO markets, thereby improving efficiency.

4.2.1.2.2 Product Construct

Two different types of PDR Regulation are contemplated by the Working Group.

<u>PDR Regulation with No Energy Settlement</u>

Unlike conventional regulation services which may require sustained energy output across multiple dispatch intervals, some PDR resources might be better suited to provide dispatchable regulation services in a “zero-net energy” (ZNE) structure. Similar to REM, a ZNE dispatch could function by returning a regulating resource to its original energy set-point every so often, e.g. every 15-minutes. As a PDR, the ZNE set point would be the baseline load level or some equivalent scheduling set point. With a ZNE focus, and also to mitigate retail/wholesale rate complications, PDR ZNE regulation could have no energy settlement since energy deliveries would likely be netted to zero within a small period, implying regulation up and regulation down services could likely occur at similar consecutive 5-minute RTD prices. The PDR ZNE regulation service would respond to AGC signals. Performance would be measured through telemetry. This follows the notion of eliminating wholesale energy settlement since regulation should be tilted toward energy neutrality for bi-directional participation. No specific concessions to the existing requirements for the frequency regulation product would be required. The Working Group initially contemplated that PDRs would need to be at least 500 kW to participate and acquire certification through testing. The resource type construct would have to accommodate the bi-directional design of positive and negative ranges for PDRs as discussed in daily load shift section. There are reasonably defined rules for telemetry aggregation that are applicable to DERs. Direct telemetry assures visibility to the ISO and is the basis for determining accuracy and mileage independent of interval metering (point being little revenue would be lost w/o energy settlement).

PDR Regulation with Energy Settlement

For some PDR resources, the idea of hour-long regulation service holds appeal. In these cases, an energy settlement will be needed. For this type of PDR participation, a PDR resource would bid for and compete to provide regulation up or down, rather than just ZNE regulation. The PDR resource could then, when dispatched, expect to receive uninstructed energy settlements for movements up or down from an initial energy schedule.

Details around the use of MGO-adjusted or other baselines, performance measurements, and AGC responsiveness would need to be included in this enhancement.

4.2.1.2.3 Jurisdictional Issues

In the case of a ZNE PDR regulation provider, the elimination of wholesale energy settlement largely avoids the possibility of any jurisdictional issues raised about the two other products discussed in this straw proposal and simplifies wholesale settlement. When a behind-the-meter (BTM) storage device provides ZNE (bi-directional) regulation service, any energy charged/discharged that modifies the customer's load would be charged at the retail rate, i.e. there would be no wholesale energy settlement or compensation, only a regulation capacity payment. The regulation capacity bids (and subsequent payment) would have to be structured by the provider to cover any retail energy charges that might exist (including the round-trip efficiency of the storage device).

For regulation that includes a wholesale energy settlement, the establishment of performance measurements is required to separate and settle wholesale responses from 'regular' retail actions.

4.2.1.2.4 Working Group Discussion

For resources seeking to provide traditional regulation down/up services and exposed up to a full hour of dispatch in one direction (and not ZNE regulation), the costs of retail energy settlements may create barriers to participation. For instance, to provide 1 MW of PDR regulation down dispatch for a full hour, a resource could conceivably show an extra 1 MWh on its retail bill if the metering does not adjust for the regulation-directed energy. Regulation capacity and mileage payments are unlikely to cover such costs. For this reason, ZNE options are preferred. Solutions to hour-long regulation services from PDRs will likely require some form of either a) energy payments from the ISO and/or b) other solutions, maybe involving utility metering adjustments.

As part of this effort, accuracy considerations should inform the design. FERC Order 755 directed rules to compensate regulation resources for being faster and more accurate while also noting that regulation capacity procurement can be lower through the use of fast and accurate resources. As part of these PDR enhancements to provide regulation, the ISO should also apply the regulation accuracy adjustment to the regulation capacity payments made to providers so that the capacity of highly inaccurate resources is more appropriate valued.

Discussion subsequent to the initial straw proposal surfaced the notion that not all frequency response participation by demand response resources would necessarily be focused on bi-directional zero net energy participation as initially assumed. As such, the notion that exclusion of an energy settlement would be desirable for simplification of implementation has been revisited.

While there still may be specific resources that choose to participate as bi-directional resources offering both regulation up and regulation down service concurrently, there could be other resources that choose to only offer regulation down during one period in the day and regulation up during other periods. This approach to participation is a natural extension of self-directed bi-directional daily load shifting since it allows a participant to be a net load consumer during one period of the day and a net load reduction during a different period. As such, energy settlement becomes an important element of market participation.

The table below illuminates some of the key differences and impacts of energy settlement of the different types of participation.

Frequency Regulation Participation	Wholesale Energy Settlement Impact	Pros	Cons	Comment
Bi-directional	Net Zero Energy (no energy settlement)	Avoids any wholesale vs. retail settlement issues	Managing state of charge for customer applications becomes complicated If customer is on	Most closely aligned with NGR REM but would not be required to have a CAISO meter or be a full time market participant

			residential TOU then periods of charging and discharging over the course of regulation period has different energy values which is a risk	
Regulation Down Only	Net Buyer	Easier to manage state of charge and customer risk	Might raise concerns of double payment when discharging	Best fit for ramp out periods (consume more)
Regulation Up Only	Net Seller	Same	Might raise concerns of double charging for energy consuming when discharging	Best fit for ramp in periods (consume less)

This doesn't alter the expectation that both capacity and mileage payments would apply as it would for any other resource type participating in frequency regulation. But in situations where a demand response resource only chooses to offer either regulation up or regulation down during any given settlement hour, the impact of and result of energy settlement becomes a more significant part of wholesale market participation.

What is not clear at this point and needs further vetting with the CAISO is whether or not it would be feasible to exclude symmetrical bi-directional participation from energy settlement and allow single direction frequency regulation to include energy settlement. The working group recognizes that it could be challenging from an implementation standpoint to have separate settlement schema for a single resource type but still sees value in eliminating energy settlement from concurrent bi-directional frequency response if it is feasible.

Initial feedback from the CAISO is that energy settlement will happen as normal course of business; however, no energy settlement would be a change. The ISO would have to turn the energy settlement off for ZNE regulation resources. Regulation with energy settlement could work for MGO-direct metered devices. Applying energy settlement to traditional DR providing regulation is challenging to directly determine performance if using a baseline. With MGO, the device can be directly metered. This is not so with traditional loads that rely on a baseline to determine resource performance.

4.2.1.3 Open Issues

- CAISO vetting of feasibility of applying bi-directional bidding to accommodate consumption in addition to curtailment for PDR.
- Development of the process to determine PDR performance measurement for both curtailment and consumption within a single day using existing 10 in 10 baseline.
- CAISO confirmation that, under existing technical and certification requirements, PDRs would be able to provide Regulation Up and Regulation Down service(s).
- CAISO vetting of feasibility to allow election of bi-directional frequency regulation participation for PDRs without energy settlement.

4.2.2 Baseline Analysis Working Group (BAWG) proposal

4.2.2.1 Introduction

Currently, the proxy demand resource (PDR) and reliability demand response resource (RDRR) use a 10 of 10 baseline with a 20% same day adjustment to estimate the load impact achieved by the resource. While research has shown this baseline to be accurate for many medium and large commercial and industrial customers, research has also shown that this baseline is not accurate for all customer types. The purpose of the Baseline Analysis Working group (BAWG) is to identify additional settlement methods which, when offered in addition to the 10 of 10 baseline, will enable the load impacts from a wider variety of demand response resources to be accurately estimated.

The BAWG identified three major areas of research.

- The use of alternative traditional baseline methods to estimate the load impact of current demand response resources.
- The option of using control groups rather than traditional baselines to estimate the load impacts of demand response resources.
- Ways to accurately measure load impacts of resources that are frequently dispatched.

4.2.2.1.1 Traditional baselines methodologies for current demand response resources

The research objective has been to identify additional traditional baselines which accurately estimate the load impacts of existing demand response resources that are not accurately estimated by the current CAISO-approved 10 of 10 baseline. Research has shown that the 10 of 10 baseline underestimates the load impact from residential customers, so identifying baselines for residential customers was an important task. In order to address this issue, analysis was done using data from the air-conditioning cycling programs of all three utilities. The analysis estimated the effectiveness of the current 10 of 10 baseline and tested the effectiveness of alternative baseline methodologies. In addition, the effectiveness of the 10 of 10 baseline on estimating the load impacts of reliability programs such as the Base Interruptible Program (BIP) and Agricultural Pump Interruptible Program has not been rigorously tested and these customers currently do not rely on a 10 of 10 baseline for their retail compensation.

The working group also addressed the issue of how to determine which baseline should be applied to which resources. Offering more than one baseline option raises the issue of whether or not all baseline options should be available to all customer types. For example, if a particular baseline is more accurate for residential customers than it is for commercial customers, the baseline might only be made available to resources consisting of residential customers. The working group also identified other operational barriers that may arise due to offering more than one baseline option.

4.2.2.1.2 Control Groups

Control groups provide an alternative to traditional baseline methodologies for the estimate of load impacts. Control group methodologies use the energy use of a group of customers who do not participate in the demand response event to compare to that of those who do. There are two main types of control groups: 1) a randomized controlled trial (RCT) and, 2) a matched control group. In the RCT a subset of participants is randomly selected in advance and withheld from curtailment during the event period. A matched control group consists of non-participants with similar characteristics to participants. The working group studied control group settlement methodologies already in use by other independent system operators and determined if they can be implemented by the CAISO. Questions that were addressed in this area include:

1. What requirements would need to be put in place to ensure the energy use of the control group accurately reflects the energy use of the treatment group?
2. What requirements regarding samples sizes or precision should be established?
3. How will the control groups be identified operationally?

4. Is it feasible to allow control groups to vary by events/rotate?

5. How can control group methodologies be established that work for both utilities and third party demand response providers (DRPs)?

4.2.2.1.3 Frequent Dispatch

The current 10 of 10 PDR baseline methodology relies upon historical non-event day data in order to estimate a baseline. It may be challenging to find 10 previous non-event days for resources which are frequently dispatched during a period within a reasonable proximity of the event day. In particular, behind the meter storage which is not separately metered and participating in a PDR or RDRR product may participate frequently in the market. The working group explored how the load impact of frequently dispatched resources can be accurately estimated using only data from the premise. Cases in which meter generator output is available and used for settlement will be considered out of the scope of this working group because it has been addressed in the ESDER Phase 1 initiative. Research was conducted to examine how many days are necessary to establish an accurate baseline.

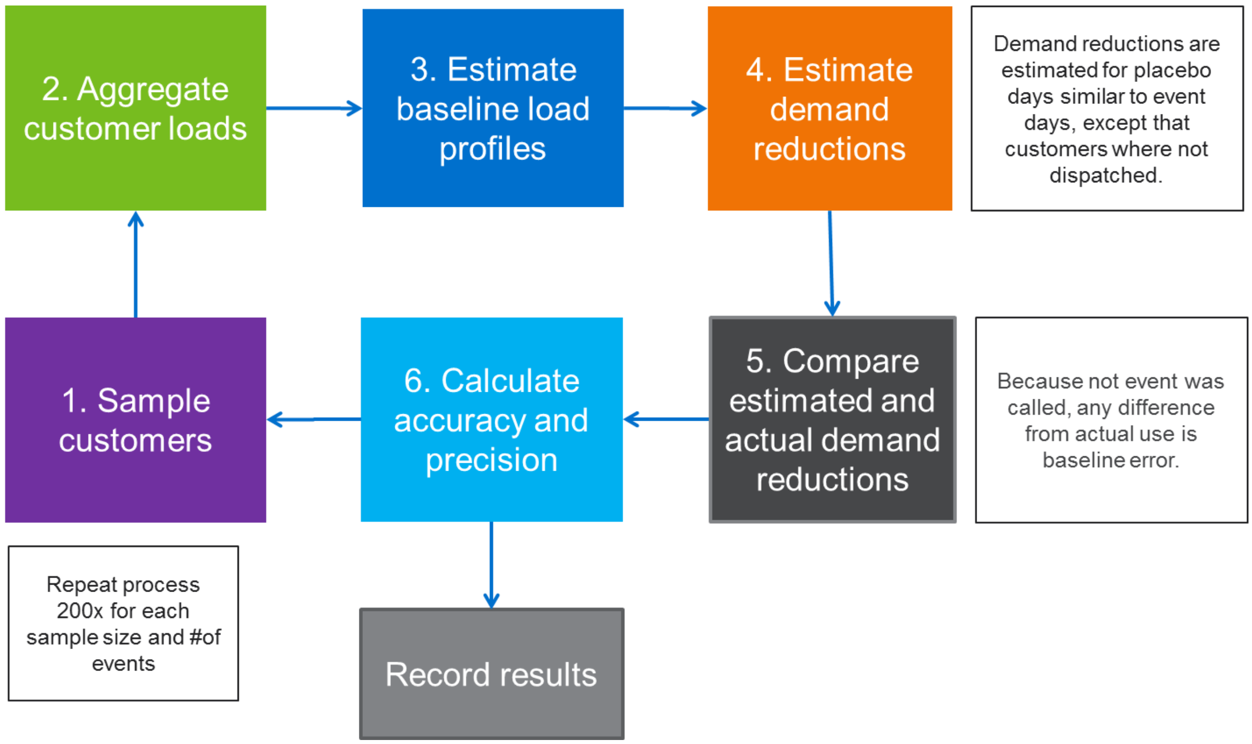
4.2.2.2 **Method for Assessing Baseline Accuracy**

To assess the accuracy of the estimated values, one needs to know the correct values. When the correct answers are known, it is possible to assess if each alternative settlement option correctly measures the demand reduction and, if not, by how much it deviates from the known values. Figure 1 summarizes the approach for assessing accuracy and precision. The basic approach is used to address all three primary areas of research.

The objective is to test different baselines with different samples of participants using actual data from participants in order to identify the most accurate analysis method. Baseline accuracy is assessed on placebo days, which are treated as event days. Because no event was called, any deviation between the baseline and actual loads is due to error.

The process is repeated hundreds of times, using slightly different samples – a procedure known as bootstrapping – to construct the distribution of baseline errors. In addition, the accuracy of the baselines is tested at granular geographic levels, such as subLAPs, to mimic market settlement. A key question is the degree to which more or less aggregation influences the accuracy and precision of the estimates. This is assessed by repeating the below process using different subsets of customers so the relationship between the amount of aggregation and baseline accuracy is quantified. Another important question is how high frequency dispatch, which limits baseline days, affects baseline accuracy. This is assessed by repeating the same process described below for different number of event days per year, thus producing a plot of accuracy and precision as a function of the number of events.

Figure 1: Method for Testing Baseline Accuracy



4.2.2.3 Metrics of Identifying Suitable Baselines

For both the accuracy of the baseline and the demand reduction, the BAWG identified the best baselines as those that are both accurate and precise. The figure below illustrates the difference between accuracy and precision. An ideal model is both accurate and precise (example #1). Baselines can be accurate but imprecise when errors are large but cancel each other out (#2). They can also exhibit false precision when the results are very similar for individual events but are biased (#3). The worst baselines are both imprecise and inaccurate, i.e. the individual event results vary substantially and they are also biased.

Figure 2: Precision versus Accuracy

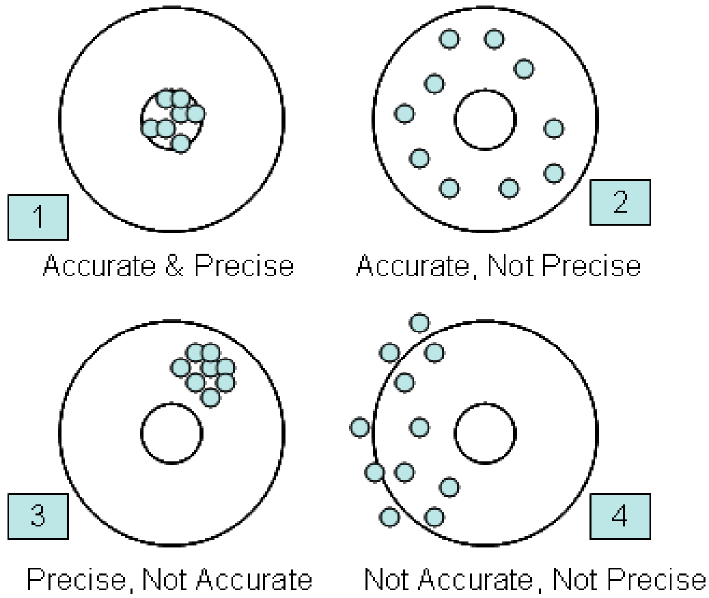


Table 1 summarizes metrics for accuracy (bias) and precision (goodness - of - fit) that were produced to assess the different baseline alternatives. Bias metrics measure the tendency of different approaches to over or under predict (accuracy or lack of bias) and are measured over multiple days. The BAWG used the mean percent error since it describes the relative magnitude and direction of the bias. A negative value indicates a tendency to under-predict and a positive value indicates a tendency to over-predict. This tendency is best measured using multiple days. Baselines that exhibit substantial bias were eliminated from consideration.

Precision metrics describe the magnitude of errors for individual events days and are always positive. The closer they are to zero, the more precise the results. The primary metric for precision was CVRMSE, or normalized root mean squared error. Among baselines which exhibit little or no bias, more precise metrics will be favored. Last, but not least, multiple baselines can prove to be both relatively accurate and precise. In which case, the BAWG has submitted its recommendation based on practical considerations such ease of implementation or potential for gaming.

Table 1: Accuracy and Precision Metrics Used to Identify Best Performing Baselines

Type of Metric	Metric	Description	Mathematical Expression

Accuracy (Bias)	Mean Percent Error (MPE)	Indicates the percentage by which the measurement, on average, over or underestimates the true demand reduction.	$MPE = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}}$
Precision (Goodness-of-Fit)	Mean Absolute Percentage Error (MAPE)	Measures the relative magnitude of errors across event days, regardless of positive or negative direction.	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
	CV(RMSE)	This metric normalizes the RMSE by dividing it by the average of the actual demand reduction.	$CV(RMSE) = \frac{RMSE}{\bar{y}}$

4.2.2.4 Baselines Included for Testing

There are a variety of approaches for measuring the magnitude of demand reduction with different degrees of complexity, data sources, and metering requirements. In addition, each method can be varied based on differences in the number of eligible days used to develop baselines, the type of days used to develop baselines, caps on the magnitude of adjustments, use of different sample sizes, and the granularity of estimates. At a high level, however, the baselines under consideration by the BAWG can be classified under three broad categories:

- Day Matching — Day-matching baselines estimate what electricity use would have been in the absence of curtailment by relying on electricity use in the days leading up to the event. It does not include information from a control group. A subset of non-event days in close proximity to the event day are identified and averaged to produce baselines. A total of 13 day matching baselines are being tested.
- Weather Matching — The process for weather matching baselines is similar to day-matching except that the baseline load profile is selected from non-event days with similar temperature conditions and then calibrated with an in-day adjustment. In general, weather matching tends to include a wider range of eligible baseline days, which are narrowed to the ones with weather conditions closest to those observed during events. A total of 7 weather matching baselines are being tested.
- Control Groups — An ideal control group has nearly identical load patterns in aggregate and experiences the same weather patterns and conditions. The only difference is that on some days, one group has loads curtailed while the control group does not. The control group is used to establish the baseline of what load patterns would have been absent the curtailment event. This approach is the primary method for settlement of residential AC cycling and thermostat programs by Texas' system operator, ERCOT. There are three basis ways to establish control valid control groups: random assignment

of customers; random assignment of clusters (for one-way devices that are not directly addressable) and matching. For the purpose of the BAWG, the focus is on random assignment of customers.

For all baseline methods, the analysis tested unadjusted baselines and the use of same day adjustments with caps of 20%, 30%, 40%, 50% and unlimited caps. Same day adjustments assume that any difference between baselines and loads in the hours leading up to the event are due to estimation and calibrate the baseline based on hours leading up to the event, with a buffer between the calibration period and the actual event. In total, 120 different baseline rules are being tested (21 baseline methods x 6 level of same adjustments).

Table 2 provides additional details about the baselines being tested. These baselines were identified by reviewing the best performing baselines for past studies, inside and outside of California, for residential, industrial, and commercial loads.

Table 2: Baselines to be Tested and Compared

Control group	Day Matching	Weather Matching
1. Comparison of means	2. Average 3 of last 3 eligible days 3. Use 3 of last 3 eligible days; more recent days receive higher weight 4. Average the top 3 of the last 5 eligible days 5. Use top 3 of the last 5 eligible days; more recent days receive higher weight 6. Average 3 of last 5 eligible days and adjust upward by 5% for all customers 7. Average top 4 of the last 5 eligible days 8. Average top 5 of the last 5 eligible days 9. Average top 3 of the last 10 eligible days 10. Average top 5 of the last 10 eligible days 11. Average 10 of the last 10 eligible days 12. Average top 3 of the last 20 eligible days 13. Average top 5 of the last 20 eligible days 14. Average top 10 of the last 20 eligible days	15. Average 3 days with most similar weather during the last three months 16. Average 4 days with most similar weather during the last three months 17. Average 5 days with similar weather during the last three months 18. Average top 3 of last 14 eligible days (including weekends); discard days that don't have similar weather based on temperature-humidity index (THI) 19. Assign days with high temperatures exceeding 80°F to 3 bins based on maximum temperature; baseline equals the average peak-period load on non-event days in a similar bin 20. Assign days with high temperatures exceeding 80°F to 3 bins based on CDD for the day; baseline equals the average peak-period load on non-event days in a similar bin 21. Assign days with high temperatures exceeding 80F to 3 bins based on the total CDH for the day; baseline equals the average peak-period load on non-event days in a similar bin

4.2.2.5 Baseline Recommendations

Table 3 shows the best performing baselines for each program. In all cases, a control group option was made available to DRPs due to their superior performance. However, certain day and weather matching baselines are also recommended in the case where DRPs do not want to withhold participants to form a control group, preventing them from being dispatched.

Table 3: Recommended Baselines for CAISO Settlement

Customer Segment	Baselines Recommended	Notes
Residential	Control Group – unadjusted (best performing for AC Cycling programs) 4 day weather match by max temperature	Since not all aggregators may be able to develop a control group, the BAWG recommends a non-control group option also be made available. The next best performing baseline was weather matching by daily max temperature and average the closest four days with most similar temperature The BAWG may also propose a simpler day-matching baseline for DRPs unable to perform weather-matching.
Non-residential	5/5 unadjusted (best performing for Agricultural Pumping) 10/10 +/- 20% adjustment (best performing for Baseline Interruptible programs) Control Group unlimited adjustment (best performing for AC Cycling programs)	Prefer a simple 5/5 baseline rather than a baseline with 10% adjustment as both perform well. Additional tests on agricultural customer data determined that the 5/5 baseline also performed well in non-drought years. The 10/10 +/- 20% adjustment is the current settlement baseline

4.2.2.5.1 Baseline Calculation Process

This section outlines the process of calculating baselines for within-subjects (such as day or weather matching baseline) methods. Baselines for randomized control trials are simply the average aggregate load profile of the control group on the event day, whether randomly assigned or developed through matching (more detail about control groups is in section 4.2.2.6).

In general, the process involves:

1. Identifying eligible baseline days that occurred prior to an event

2. Averaging participant load on the event day and on each eligible baseline day

3. Selecting the candidate baseline days out of the pool of eligible days according to the baseline method

4. Averaging customer load across the candidate days to generate the baseline.

5. Calculate the same-day adjustment, if necessary. Apply the ratio of event-day average kWh in the pre-event period to that same period of the average baseline day to the overall baseline day. If the ratio of the event to baseline day is greater than the adjustment cap, apply only the adjustment cap to the baseline. Steps for calculating the same-day adjustment can be found on page 35.

The baseline methods from step three above are described in more detail below.

5/5 Day Matching Baseline: Candidate days are non-holiday weekdays. Event days are also excluded from candidate days if customers for whom the baselines are calculated are dually enrolled in that program. The aggregate load on the five previous candidate days prior to the event are averaged to create the baseline.

10/10 Day Matching Baseline: Candidate days are non-holiday weekdays. Event days are also excluded from candidate days if customers for whom the baselines are calculated are dually enrolled in that program. The average aggregate load on the ten previous candidate days prior to the event are averaged to create the baseline.

4 Day Weather Match Baseline by Max Temperature: Candidate days are non-holiday weekdays. Event days are also excluded from candidate days if customers for whom the baselines are calculated are dually enrolled in that program. Eligible days must be within 90 days of the event day. Weather for that group of customers is determined by their zip code. Zip codes are mapped to the NOAA weather stations laid out in the attached document[13].

ZIP to Weather
Station Mapping

The temperature used to find the days that make up the baseline must be the customer-count weighted average of temperature experienced by the participant group. That is, if a resource is comprised of 500 customers served by weather station A and 2000 customers served by weather station B, the average temperature profile for each hour h, would be:

$$T_h^{avg} = (500/2500) * T_h^A + (2000/2500) * \ T_h^B$$

The maximum average temperature for each day is used to identify the days that will comprise the baseline. For each of the eligible days, measure the absolute difference between the maximum temperature of that day and the maximum temperature of the event day, and pick the 4 days that have the least difference between that day's maximum temperature and that of the event day. Average the daily average aggregate load across those four days to create the baseline.

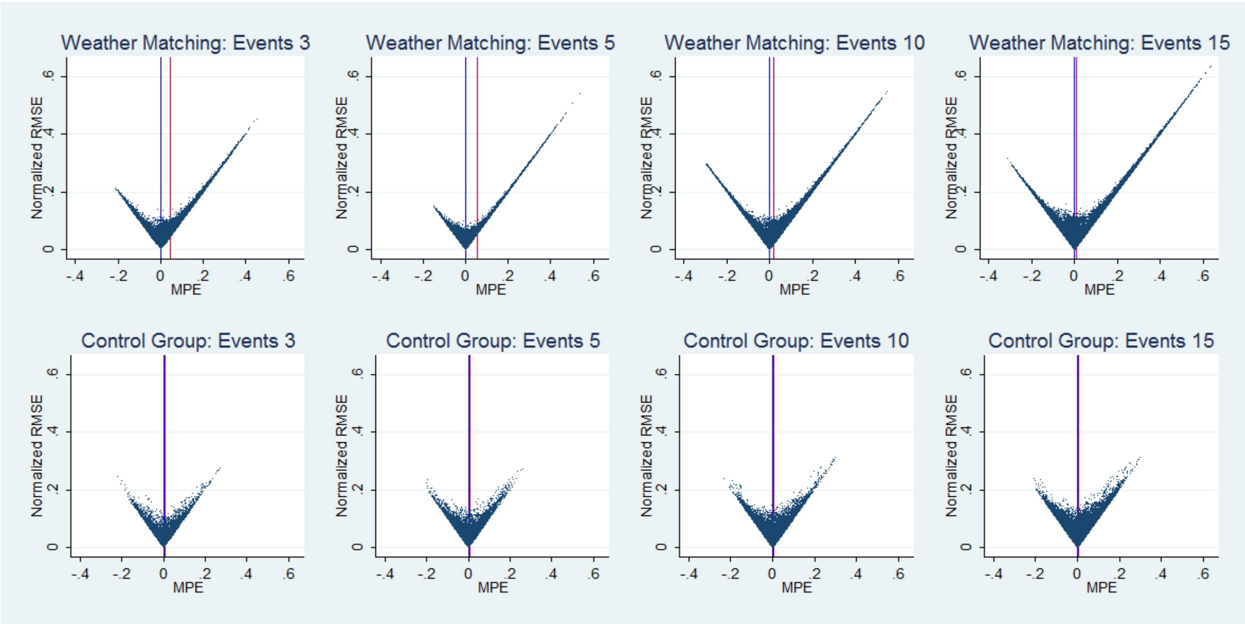
4.2.2.5.2 Issues Surrounding Frequent Dispatch of Resources

In general, the baseline type chosen is the primary driver of overall bias and precision for each baseline, rather than the number of event days that were called. That is, a good baseline will perform better than a bad baseline regardless of the number of event days called. It is therefore more important to select baselines that have low bias and high precision on average even if frequent dispatch limits the accuracy of baselines on individual events in simulation. Shown in Figure 3 is the dispersion of individual event and simulation runs for two baseline methods across different frequencies of event dispatch. Within subjects methods, like the weather matching baseline shown below, are more sensitive to frequent dispatch as they explicitly rely on past eligible days to create a baseline. As more of the recent past days are simulated as event days, the pool of candidate baseline days become less similar to the event day, leading to higher bias for those individual events. However, as more events are called, the average bias across

[13] An online source for weather data can be found here: http://www.ncdc.noaa.gov/qclcd/QCLCD?prior=N

all event days decreases, leading to more accurate settlement on average. In the figure below, the blue line represents an MPE of 0, or no bias. The red line is the average bias of all simulated event runs. As the number of dispatched resources increases, the average bias of all simulations moves closer to zero for within subjects methods. For control group methods, frequent dispatch does not affect the bias or dispersion of results. In those cases, the baselines are developed from a control group's usage on the same day the event is being held.

Figure 3: Dispersion of Bias and Precision Metrics for Frequent Dispatch



4.2.2.6 Implementation of Control Group Settlement Methodology

Randomized control groups consistently outperformed day and weather matching baselines for residential and commercial AC cycling programs during testing. With large enough sample sizes, between 200 and 400 participants, they were more than twice as precise as day or weather matching baselines. For this reason, the BAWG recommends that control groups be one of the settlement options for both residential and non-residential customers.

Control groups involve using a set of customers who did not experience events to establish a baseline. A control group should be made of customers who are statistically indistinguishable from the participant group on non-event days to act as a comparison on event days, instead of relying on participants' past performance. There are many ways to develop a control group; however, the two that the BAWG has considered are randomized control groups (RCTs) and matched control groups. While both methods are valid alternative settlement methodologies, issues surrounding the development of matched control groups (e.g. data security, legality, and cost) were out of scope for the BAWG and are not discussed in this document. The following section, therefore, addresses the process of developing and validating a randomized control trial. In a RCT, a subset of participants enrolled in the DR resource is withheld from receiving the treatment or participating in the event. This subset is randomly assigned. Random assignment with a sufficient sample size ensures that the control group is statistically indistinguishable from the treatment group. This then means that any difference in load profiles on event days can be attributed to the effect of treatment, and that any difference between the two groups on non-event days should be negligible.

However, before a control settlement methodology can be employed it is necessary to demonstrate that the energy use of the control group is an accurate predictor of the energy use of the participants. Three high level requirements for demonstrating the validity of a control group are shown below. Instructions for demonstrating control group equivalence follow, with applied examples in the appendix to this document. Once a suitably accurate and precise baseline has been developed, it can be adjusted using same-day adjustments as described at the end of this section. However, it is the unadjusted baseline that must meet the accuracy, precision and sample size criteria.

Figure 4: Control Group Requirements

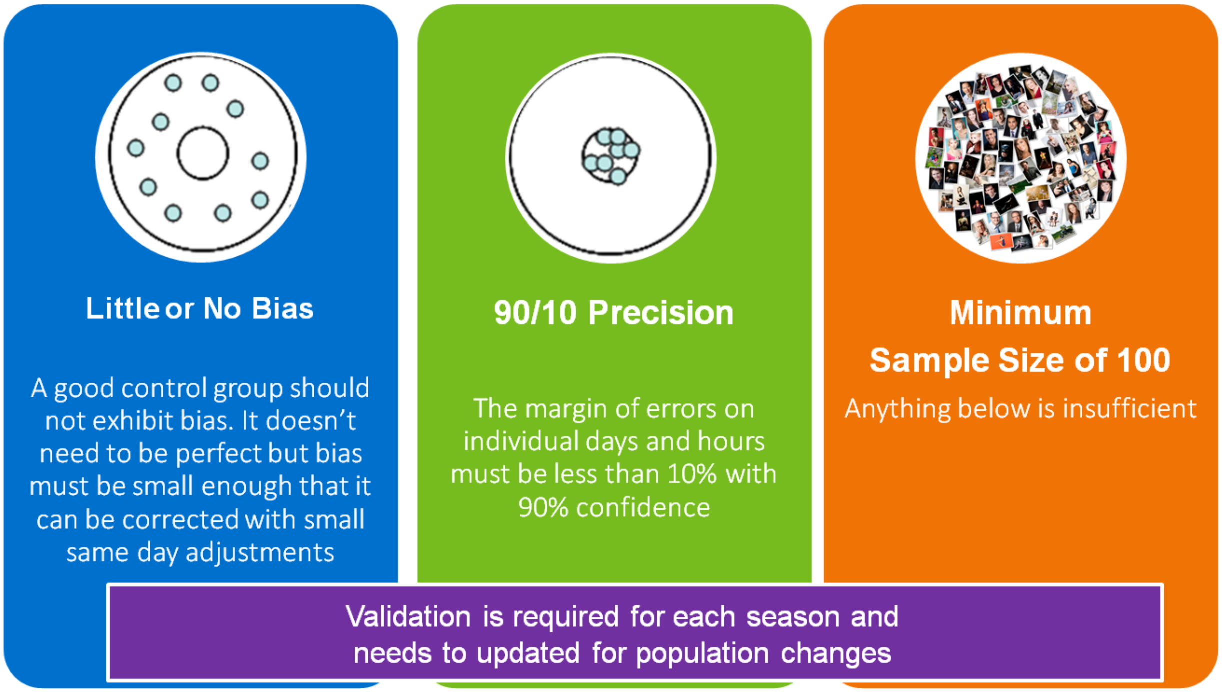


Figure 4 demonstrates the three key principles for the development and validation of control groups. They must exhibit little or no bias, must be sufficiently precise, and be large enough to represent the treatment population.

4.2.2.6.1 Statistical Checks Necessary to Demonstrate Control Group Validity

DRPs will need to demonstrate that the control group reflects the electricity use patterns of customers curtailed (validation). The process for demonstrating equivalence is outlined below. It is the responsibility of the DRP to develop the control group and demonstrate equivalence. The control group(s) developed are subject to audit by the CAISO.

1. The DRP Identifies a control pool of at least 100 customers to be selected via statistical matching or randomly withheld from the participant population. A single control group may be used for multiple subLAP settlement groups; however, equivalence, using the procedure outlined below, must be demonstrated for each of the treatment groups against the control group. For example, if there are five subLAPs, five equivalence checks must be completed to show that the control customers are equivalent to treatment customers in subLAPs A, B, C, D and E. Use of a different control group for each subLAP is also

permitted and will be necessary if there are significant differences in weather sensitivity or other characteristics among treatment groups in different subLAPs. In those cases, equivalence must be demonstrated only between the treatment group and the control group for which it is acting as control.

2. For each resource ID, pull hourly data from the previous applicable RA season. The RA seasons are currently defined as summer from April to October and winter from November to March. So if a resource is expected to be bid into the market in July, use data from the prior year's summer RA season. Note that the RA season definitions may change in the future and it is the DRP's responsibility to update their validation periods accordingly. The BAWG may also chose to allow the verification window to instead be a 6-month rolling period prior to the month of the control group's formation.

3. Average the hourly load profile for all treatment group customers and all control group customers by day hour and season. The RA seasons are currently defined as summer from April to October and winter from November to March.

4. Flag and remove days during which the resource is unavailable. This may include weekends, holidays, and outage days. In addition, exclude event days that the customers in the resource could have participated in. If customers are dually participating in utility load modifying programs, event days of the load modifying resource may also be excluded.

5. Arrange the data in the appropriate format. For most statistical packages and Excel, regressions are easiest to perform when data is in a long format by date and hour and wide by treatment status. Note that the datasets should be separate for each RA season and treatment/control group pairing to be tested.

6. Regress average treatment hourly load against average control hourly load during event hours with no constant. This can be done in a statistical package like R or Stata, or within an Excel file or other spreadsheet application. The functional form of this model should be

$$y_{i,h}^{T} = \beta y_{i,h}^{C} + \varepsilon_{i,h}$$

   Where $y_{i,h}^{T}$ is the average kW across all treatment customers for the non-event day i and hour h, and $y_{i,h}^{C}$ is the average kW across all control customers for that same hour and day. The coefficient, β, represents the bias that exists in the control group; that is, the percent difference between the average treatment kW and the average control kW across all days and event hours. A coefficient of 1.05 means that the treatment group demand is on average 5% higher than that of the control group. Similarly, a coefficient of 0.86 means that the control group load is 86% that of the treatment group.

7. To demonstrate lack of bias, the coefficient β should be between 0.95 and 1.05, minimizing the unadjusted absolute bias from the treatment group.

8. To demonstrate that the control group has sufficient precision, the value of the normalized root mean squared error at the 90% confidence level should be less than 10%. The normalized root mean squared error, or CVRMSE, is calculated according to

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i,h}(y_{i,h}^{C} - y_{i,h}^{T})^2}{n}}}{(1/n)\sum_{i,h} y_{i,h}^{T}}$$

In this equation, the squared difference between treatment and control for each event hour and day is summed over all event hours and days, and then divided by the total number of event hours and days (n). The square root of that value is divided by the average treatment load across all event hours and days to normalize the error. Under the assumption that the CVRMSE is normally distributed, the 90% confidence level for this statistic is 1.645 times the CVRMSE. For example, if the CVRMSE is 0.86%, the 90% confidence level for the statistic is 1.414%.

4.2.2.6.2 Using Matched Control Groups to Generate a Baseline

Use of a matched control group would allow DRPs to dispatch their entire participant group during an event, while a separate group of non-participants would act as a control. Alternatively, participants that include customers both inside and outside a subLAP could act as a control group.

The BAWG is open to the possibility of a matched control group baseline option. It is the preferred option for SCE. However, PG&E, SCE, and SDG&E were concerned about customer data security, the allocation of cost to fund this option, and potential legal issues associated with having utilities involved in identifying a matched control group on behalf of other DRPs. While matched control groups are subject to the same validation criteria as randomized control groups, the use of non-participants to develop a control group is of considerable interest to DRPs that wish to dispatch their entire enrolled population during an event. However, no recommendation has been developed that would allow DRPs access to non-participant data to develop the matched control group.

However, a few agreements were reached.

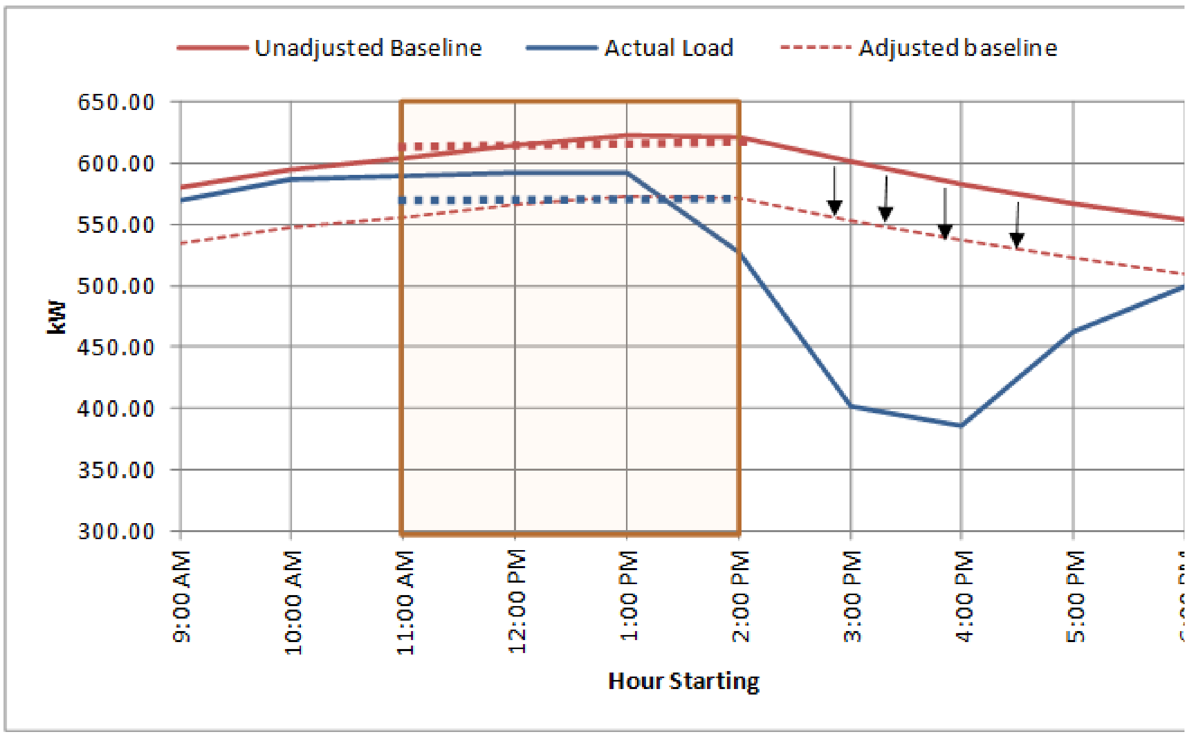
- DRPs with access to non-participant interval data may have the option to utilize matched control groups. The BAWG may choose to withhold the ability to create a matched control group if the access to non-participant data is not available to all parties. *These matched control groups are subject to the same validation requirements as the randomly assigned control groups, as outlined above.*
- The issue of access to non-participant data is broader than its use for settlement baselines and needs be worked out at the CPUC.
- The matched control group can be updated on a monthly basis but needs to be designated in advance. It cannot be changed once it is set for the month and cannot be changed after the fact.
- The matched control group assignment is subject to audit. The purpose of audits is to assure that baselines were properly calculated and control groups met precision and validation criteria. Audits may include delivery of customer interval data with the goal of recreating bias and precision metrics assessed in the validation process.

4.2.2.6.3 Same-Day Adjustments to Calibrate Control Groups

Baseline estimates of electricity use during an event period can be adjusted up or down based on electricity use patterns during the hours leading up to an event. This procedure is known as *same-day adjustment*. If, during pre-event hours, the baseline is less than the actual load, it is adjusted upwards. Similarly, if the pre-event baseline is above the actual load before the event, it is adjusted downwards. To adjust the load, the initial baseline value is multiplied by the ratio between the unadjusted baseline and the actual load during pre-event hours. In other words, the baseline is calibrated to match actual usage patterns in the hours leading up to the event. Note that the same-day adjustment procedure implicitly assumes that differences between the baseline and actual loads during hours leading up to an event are due to predictive error, and *not* due to customer behavior such as shifting of production to pre-event hours or implementing demand reductions early.

Figure 5 illustrates the baseline adjustment process. In the example, the event starts at 3 PM. The first three of the four hours leading up to the event, from 11 AM to 2 PM, are used to calculate the adjustment. The blue line represents the actual load for the day. The red line reflects the calculated baseline prior to the application of same-day adjustments. In this example, in the hours leading up to the event, the unadjusted baseline is higher than the actual load. The baseline adjustment process assumes this difference is due to error. To correct for this difference, the baseline is calibrated downward by roughly 8%, as reflected by the red dotted line.

Figure 5: Example of Baseline Same-day Adjustment



$$Adjustment = \frac{Avg.\,kW\ during\ adjusment\ period}{Avg.\,unadjusted\ baseline\ over\ adjustment\ period} = \frac{571}{619} = 92.2\%$$

If the difference between the unadjusted baseline and the actual load is truly due to baseline estimation error, the adjustment process reduces those errors. Same-day adjustments are often capped because adjustment can introduce the potential for manipulation of pre-event loads to bias baselines. The concern is that participants may be able to "game" the system by increasing their electricity use during the adjustment period, leading to baselines that are too high and that overestimate actual demand reductions. Capping the magnitude of the adjustment limits the potential for this kind of abuse.

To calculate a same-day adjustment once the unadjusted baseline has been calculated, the following steps are performed:

1. Calculate the average treatment load in the three hours prior to the event prior to a one-hour buffer period before the event starts. For example, if an event started at 1pm, the adjustment window would be 9am to 12pm. Calculate the average control group load during the same window using the event baseline (the average load of the control group customers).

2. The ratio of treatment kW during the adjustment window to that of the control group during that same window is the percentage adjustment. The BAWG may also permit using both pre- and post-event load data to develop this ratio.

3. Cap the ratio if using a cap. For example, if the adjustment ratio is 112% but the cap on adjustments is 10%, then the adjustment ratio will now be 110%. If no cap is being used, the adjustment ratio remains 112%.

4. Apply the adjustment ratio to all baseline hours for the control group on the event day. Each average kW of the baseline across the 24 hours of the event day is multiplied by the capped or uncapped adjustment ratio from step 3.

5. Average customer load impacts are then estimated by the difference between the adjusted baseline and the observed treatment load.

4.2.2.7 Baseline Process Discussion

The following additional process discussion points were addressed in meetings of the full working group.

- **Levels of Acceptable Aggregation of Control Groups across SubLAPs:** Aggregation of control groups is permissible across different subLAPs; however, the same performance on intra-subLAP equivalence checks must be demonstrated. While sourcing a control group from a region with similar weather and customer mix conditions is not explicitly mandated, considerations for these attributes that affect load may help in developing an appropriate control group.

- **Accommodations needed for rotating or updating control groups regularly:** The assignment to treatment and control groups can be updated on a monthly basis; however, this assignment must be completed prior to any events. Validation of new control groups must also be completed prior to any events in concurrence with any new control group development. The assignment cannot be changed once set for the month and cannot be changed after the fact.
- **Audit Criteria and Frequency:** Control group equivalence is subject to audit. In the case where the California ISO deems it necessary, DRPs will be required to securely provide the control and treatment group's interval data to recreate the regression coefficient and CVRMSE to ensure they meet the criteria laid out in this document.
- **Allowing custom or alternate baselines:** CAISO does not support any recommendation for new or custom baselines.
- **Who will estimate the baselines:** The BAWG recommends that DRPs estimate the baselines and provide them to CAISO. CAISO will have an annual process where the DRPs attest to the accuracy of the baselines and may also audit the accuracy of the baselines on an as-needed basis.
- **Managing baselines for customer transitions:** Further work in this area is needed. The registration process for new PDRs needs to be fully understood by the BAWG participants to ensure that the proper recommendation is developed. A suspension period for customers transitioning to a new settlement group may be necessary to ensure there are sufficient past candidate days to develop a baseline. A method of tracking past event days for customers who transition is also required.

4.2.2.8 Appendix

4.2.2.8.1 Applied Example of Validation Required – Using Excel

Shown below are examples of how to demonstrate equivalence between treatment and control groups in Excel. As described above, the steps to performing this calculation are:

1. Identify a control pool of at least 100 customers to be selected via statistical matching or randomly withheld from the participant population. Create a dataset that has the form shown in Figure 6 with control and participant's hourly usage by date from hours ending 1 through 24.

Figure 6: Base Dataset

Participant ID	Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
1	C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
1	C	Winter	1/1/2015	0.72	1.81	0.88	1.97	1.39	1.79		1.49	1.40
1	C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
1	C	Winter	1/3/2015	1.76	0.61	1.99	0.77	1.27	1.27		1.85	1.85
1	C	Winter	1/4/2015	1.60	0.66	1.55	1.08	1.86	1.57		0.68	0.83
1	C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
1	C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
2	T	Winter	12/31/2014	1.11	0.97	1.39	0.58	1.36	1.30		1.54	0.79
2	T	Winter	1/1/2015	0.65	1.04	1.38	1.31	0.81	1.68		0.80	1.47
2	T	Winter	1/2/2015	0.97	1.44	1.31	1.19	1.89	1.74		0.59	1.44
2	T	Winter	1/3/2015	1.16	1.59	1.70	1.25	1.11	1.63		0.79	0.97
2	T	Winter	1/4/2015	0.72	1.98	1.24	1.52	1.91	1.99		0.57	1.85
2	T	Winter	1/5/2015	0.56	1.20	1.19	1.34	1.33	0.50		1.23	1.38
2	T	Winter	1/6/2015	0.99	0.99	0.60	1.32	0.61	1.23		0.93	1.27
3	T	Winter	12/31/2014	1.59	1.81	0.58	1.69	1.49	1.15		0.55	1.81
3	T	Winter	1/1/2015	1.11	1.67	0.71	1.00	0.95	1.39		1.86	1.50
3	T	Winter	1/2/2015	1.71	1.54	1.26	1.40	1.67	1.52		1.90	1.67
3	T	Winter	1/3/2015	1.54	1.11	1.03	1.45	1.10	0.85		1.81	2.00
3	T	Winter	1/4/2015	1.13	0.67	1.25	0.83	1.96	1.58		0.78	0.64
3	T	Winter	1/5/2015	0.96	1.06	1.35	0.89	1.72	1.01		0.54	1.95
3	T	Winter	1/6/2015	0.99	1.35	1.32	0.75	0.82	1.16		1.08	1.11

2. Average the hourly load profile for all treatment group customers and all control group customers by day and hour.

Figure 7: Average Daily Treatment and Control Usage

Ineligible Day	Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
	C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
Holiday	C	Winter	1/1/2015	0.72	1.81	0.88	1.97	1.39	1.79		1.49	1.40
	C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
Weekend	C	Winter	1/3/2015	1.76	0.61	1.99	0.77	1.27	1.27		1.85	1.85
Weekend	C	Winter	1/4/2015	1.60	0.66	1.55	1.08	1.86	1.57		0.68	0.83
	C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
	C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
	T	Winter	12/31/2014	1.35	1.39	0.98	1.14	1.42	1.23		1.05	1.30
Holiday	T	Winter	1/1/2015	0.88	1.36	1.04	1.15	0.88	1.53		1.33	1.49
	T	Winter	1/2/2015	1.34	1.49	1.28	1.29	1.78	1.63		1.25	1.56
Weekend	T	Winter	1/3/2015	1.35	1.35	1.36	1.35	1.10	1.24		1.30	1.49
Weekend	T	Winter	1/4/2015	0.92	1.33	1.25	1.18	1.93	1.79		0.68	1.24
	T	Winter	1/5/2015	0.76	1.13	1.27	1.11	1.52	0.76		0.88	1.66
	T	Winter	1/6/2015	0.99	1.17	0.96	1.04	0.72	1.19		1.01	1.19

3. Flag and remove days in which the resource is not available and event days that the customers in the resource could have participated in.

Figure 8: Average Daily Treatment and Control Usage on Eligible Days

Treat	RA Season	Date	kWh1	kWh2	kWh3	kWh4	kWh5	kWh6	…	kWh23	kWh24
C	Winter	12/31/2014	2.00	1.11	1.91	1.29	0.78	1.25		0.97	1.44
C	Winter	1/2/2015	0.85	0.59	1.67	0.64	0.67	1.04		2.00	1.42
C	Winter	1/5/2015	1.59	1.32	0.53	1.32	1.44	0.88		1.12	1.18
C	Winter	1/6/2015	1.45	1.63	1.47	1.50	1.66	0.98		1.90	0.66
T	Winter	12/31/2014	1.35	1.39	0.98	1.14	1.42	1.23		1.05	1.30
T	Winter	1/2/2015	1.34	1.49	1.28	1.29	1.78	1.63		1.25	1.56
T	Winter	1/5/2015	0.76	1.13	1.27	1.11	1.52	0.76		0.88	1.66
T	Winter	1/6/2015	0.99	1.17	0.96	1.04	0.72	1.19		1.01	1.19

4. Arrange the data in the appropriate format.

Figure 9: Reshaped Average Daily Treatment and Control Usage on Eligible Days

Date	Hour	kWh_Treat	kWh_Control
12/31/2014	1	1.35	2.00
	2	1.39	1.11
	3	0.98	1.91
	4	1.14	1.29
	5	1.42	0.78
	6	1.23	1.25
	…		
	23	1.05	0.97
	24	1.30	1.44
1/2/2015	1	1.34	0.85
	2	1.49	0.59
	3	1.28	1.67
	4	1.29	0.64
	5	1.78	0.67
	6	1.63	1.04
	…		
	23	1.25	2.00
	24	1.56	1.42
1/5/2015	1	0.76	1.59
	2	1.13	1.32
	3	1.27	0.53
	4	1.11	1.32
	5	1.52	1.44
	6	0.76	0.88
	…		
	23	0.88	1.12
	24	1.66	1.18
1/6/2015	1	0.99	1.45
	2	1.17	1.63
	3	0.96	1.47
	4	1.04	1.50
	5	0.72	1.66
	6	1.19	0.98
	…		
	23	1.01	1.90
	24	1.19	0.66

5. Regress average treatment hourly load against average control hourly load during event hours with no constant by filling in the attached template and updating formulas in cells H20 and H24 to include the full range of the data added to columns B through E.

Randomization
Validation Template.x

Figure 10: Regression and Validation Template

	A	B	C	D	E	F
1				Treatment	Control	Error
2		Date	Hour	kWh	kWh	Squared
3		12/31/2014	1	1.35	2.00	0.42250
4		12/31/2014	2	1.39	1.11	0.07840
5		12/31/2014	3	0.98	1.91	0.85008
6		12/31/2014	4	1.14	1.29	0.02449
7		12/31/2014	5	1.42	0.78	0.42055
8		12/31/2014	6	1.23	1.25	0.00046
9		12/31/2014	...			0.00000
10		12/31/2014	23	1.05	0.97	0.00562
11		12/31/2014	24	1.30	1.44	0.01960
12		1/2/2015	1	1.34	0.85	0.24010
13		1/2/2015	2	1.49	0.59	0.81000
14		1/2/2015	3	1.28	1.67	0.15016
15		1/2/2015	4	1.29	0.64	0.43296
16		1/2/2015	5	1.78	0.67	1.22545
17		1/2/2015	6	1.63	1.04	0.34928
18		1/2/2015	...			0.00000
19		1/2/2015	23	1.25	2.00	0.57003
20		1/2/2015	24	1.56	1.42	0.01823
21		1/5/2015	1	0.76	1.59	0.68558
22		1/5/2015	2	1.13	1.32	0.03648
23		1/5/2015	3	1.27	0.53	0.54834
24		1/5/2015	4	1.11	1.32	0.04182
25		1/5/2015	5	1.52	1.44	0.00601
26		1/5/2015	6	0.76	0.88	0.01525
27		1/5/2015	...			0.00000
28		1/5/2015	23	0.88	1.12	0.05452
29		1/5/2015	24	1.66	1.18	0.23136
30		1/6/2015	1	0.99	1.45	0.20794
31		1/6/2015	2	1.17	1.63	0.20931
32		1/6/2015	3	0.96	1.47	0.26317
33		1/6/2015	4	1.04	1.50	0.21716
34		1/6/2015	5	0.72	1.66	0.89114
35		1/6/2015	6	1.19	0.98	0.04623
36		1/6/2015	...			0.00000
37		1/6/2015	23	1.01	1.90	0.79477
38		1/6/2015	24	1.19	0.66	0.28037

1. Populate the values to the right with eligible (n
winter (perform these calculations in separate ta

2. Update the formulas in cells H20 and H24 (the E
500, for example, ensure that the formulas in H20

3. Make a scatterplot with control kWh as the X-a

4. Right click on the scatterplot data in the graph
options circled to the right, then click 'OK'
a. Linear Regression Type
b. Set Intercept = 0
c. Display Equation on chart

BETA
0.999271146

Must be between 0.95 and 1.05

CVRMSE
4.84%

Margin of Error
with 90% Confidence
8.0%

Must be less than 10%

6. The statistics of interest are in cells H20, H24, and H29.

4.2.2.8.2 Applied Example of Validation Required – Using Stata

Example code that performs the control group validation can be found here:

Stata Code to Validate Equivalence.do

The command to perform this regression is: *reg kWh_treat kWh_control, noconstant.* If using Stata, the validation statistics can be calculated easily using the two commands underlined in green. The coefficient β is the value circled in orange. The 90% limit on the CVRMSE can be calculated using the output (circled in blue) from the same two commands as shown in Figure 11.

Figure 11: Stata Commands to Calculate Equivalence Statistics

```
. reg kwh_treat kwh_control, noconstant

      Source |       SS       df       MS              Number of obs =    5568
-------------+------------------------------           F(  1,  5567) =       .
       Model |  3792.8973     1  3792.8973             Prob > F      =  0.0000
    Residual |  10.197965  5567  .00183186             R-squared     =  0.9973
-------------+------------------------------           Adj R-squared =  0.9973
       Total | 3803.09527  5568  .683027167            Root MSE      =   .0428

------------------------------------------------------------------------------
   kwh_treat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 kwh_control |    1.00539   .0006987  1438.93   0.000     1.004021     1.00676
------------------------------------------------------------------------------


. sum kwh_treat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   kwh_treat |      5568    .7518921    .3430839   .1965188   3.313407


. di in red "the RMSE is " e(rmse)
the RMSE is .04280023

. di in red "the average treatment kWh is " r(mean)
the average treatment kWh is .75189212

. di in red "the 90% confidence limit of the CVRMSE is " 1.645 * (e(rmse)/r(mean)) * 100 "%"
the 90% confidence limit of the CVRMSE is 9.3638946%

. di in red "it can also be manually entered like this: " 1.645 * (.04280023/.75189212) * 100 "%"
it can also be manually entered like this: 9.3638936%
```

4.3 Multiple-use applications

Multiple-use applications are those where an energy resource or facility provides services to and receives compensation from more than one entity. DER could potentially provide and be compensated for many services to customers, the distribution system and the wholesale markets as new markets and services evolve across the energy supply chain.

4.3.1 Progress made in ESDER 1

In ESDER 1, the ISO addressed two broad categories or types of multiple-use applications: (1) DER providing reliability services to the distribution grid and services to the wholesale market; and (2) DER providing services such as demand management to end-use customers while participating in the wholesale market. ESDER 1 limited its treatment of these multiple-use applications to circumstances where the resource either is not providing resource adequacy (RA) capacity or can set aside a portion of its installed capacity not providing RA capacity. The criterion "not providing RA capacity" was defined to apply on a monthly basis for purposes of the initiative; i.e., the capacity in question should not be included in a load-serving entity's RA plan for the given month.

In the case of DER providing services to the distribution system and participating in the wholesale market (the first category of multiple use applications examined in ESDER Phase 1), the ISO posed three questions and developed a proposed approach to each.

First, if DER is procured by the distribution utility to provide a grid service and bids into the ISO market, how should conflicting real-time needs of the distribution utility and the ISO be managed? The ISO proposed that it would settle a DER dispatch as other generating resources are settled – i.e., that if the DER deviates from an ISO dispatch instruction to provide service to the distribution system or for another reason, its deviation will be settled as uninstructed imbalance energy. Rather than establish a priority among conflicting needs, the ISO proposed to leave it to the resource owner or operator to decide how to respond in light of the settlement consequences for deviating from an ISO dispatch instruction.

Second, for any market interval in which the DER follows an ISO dispatch instruction that aligns with the service the same DER is providing to the distribution utility, is there a double payment concern that must be addressed? The ISO proposed not to implement any provisions to address potential double payment situations where a DER is compensated by the distribution utility and is also settled through the ISO market for responding to an ISO dispatch. Instead, the ISO indicated that although it may reconsider this position, it did not believe the issue is ripe for resolution because distribution-level services have not yet been defined. The ISO's position is that double payment concerns from both the distribution utility for distribution-level services and the ISO for wholesale market participation must be based on an understanding of the specific distribution-level services involved and how they are procured, utilized and compensated by the distribution utility. These questions are being considered in CPUC proceedings[14] and may or may not be ripe for consideration by the ISO in ESDER 2.

Third, the ISO considered whether there should be limitations on the provision of distribution-level services by a multi-pricing node DER aggregation or the sub-resources of a single-pricing node or multi-pricing node DER aggregation that is an ISO market participating resource? If so, what limitations are appropriate? The ISO proposed not to impose any such limitations. This is because under the ISO's proposed DER aggregation framework[15], the ISO will require no specific performance by sub-resources that comprise either a multi-pricing node or single-pricing node DER aggregation. The ISO's requirement is that when the ISO issues a dispatch instruction to a DER aggregation, the net response at each constituent pricing node be in the direction of the dispatch and the net response across constituent pricing nodes be in proportion to the DER aggregation's distribution factors. As long as the DER aggregation complies with this requirement, the operational behavior of individual sub-resources will not be subject to ISO requirements. An individual sub-resource could respond to the needs of the distribution system as long as the DER provider who operates the DER aggregation delivers the net response at the associated pricing node that is in the same direction as the dispatch instruction and aligns with the distribution factors for the DER aggregation.

With DER that provide services to end-use customers and participate in the wholesale market (the second category of multiple use applications examined in ESDER 1), the ISO determined that no additional new provisions were needed beyond the provisions developed in ESDER 1 for PDR/RDRR involving behind-the-meter generation devices. To accommodate the proliferation of behind-the-meter generation devices involved in demand response, the ISO developed an alternative performance evaluation methodology that directly meters the behind-the-meter generation device to measure the demand response provided by the device separate from the facility load.[16] The

[14] See in particular the CPUC Distribution Resources Plan (DRP) proceeding (R.14-08-013) and the Integration of Distributed Energy Resources (IDER) proceeding (R.14-10-003).

[15] See the ISO's filing with the Federal Energy Regulatory Commission at this link: http://www.caiso.com/Documents/Mar4_2016_TariffAmendment_DistributedEnergyResourceProvider_ER16-1085.pdf

[16] This alternative performance evaluation methodology was approved by the Federal Energy Regulatory Commission in August 2016.
http://www.caiso.com/Documents/Aug16_2016_LetterOrderAcceptingTariffAmendment_EnergyStorage_DistributionEnergyResourceInitiative_ER16-1735.pdf

demand response performance is the demand reduction resulting from the output of the behind-the-meter generation device for the dispatch interval. Under the ISO's proposal, the resource's response is evaluated based on the physical meter generator output for the dispatch interval and reduced by an estimate of the typical energy output of the device used for retail load-modifying purposes and benefits. This adjustment appropriately removes an estimated quantity of energy delivered by the device to the facility for its retail load-modifying purposes, i.e., energy not produced in response to an ISO dispatch. The adjustment is intended to mitigate issues of wholesale and retail service overlap and the potential for double compensation present in this multiple use application scenario. The adjustment is calculated by taking an average of the energy delivered by the generation device during a prescribed number of prior non-event hours.

4.3.2 Effort in ESDER 2

In ESDER 2 the ISO has continued its efforts to address multiple-use applications through its participation in the CPUC's energy storage proceeding.[17] The ISO and CPUC began a collaborative stakeholder process on this subject with a joint workshop held on May 2-3, 2016 at the CPUC to address station power (see section 4.4) and multiple-use applications. Many stakeholders made informative presentations at the workshop, and the CPUC and ISO received extensive written comments on May 13 and reply comments on May 20. Based on the workshop presentations and the submitted comments the ISO has not identified any issues or topics that should be addressed in a separate effort under ESDER 2. If further activities in the CPUC proceeding identify issues that require treatment in an ISO initiative or develop proposals appropriate for ISO consideration, refinement and possible adoption, the ISO will consider those issues in its stakeholder initiatives catalog and roadmap for 2017.

4.3.1 Additional background from the ESDER 2 issue paper

The viable revenue streams available to energy storage resources will drive the number and variety of energy storage use-cases and configurations that will appear in the evolving DER marketplace. Revenue or "value streams" reflect the energy and capacity services energy storage resources can or will be able to provide and be compensated for as new markets and energy services evolve across the energy supply chain.

[17] CPUC Rulemaking 15-03-011.

Rocky Mountain Institute ("RMI") published a study on the economics of battery storage to address what services exist or may exist that will drive multi-use applications and the value proposition for energy storage. The study identified 13 services that energy storage can provide to three distinct stakeholder segments or areas of the supply chain, summarized in the table below.[18]

STAKEHOLDER GROUPS	SERVICES
ISO/RTO SERVICES	• Energy Arbitrage • Frequency Regulation • Spin / Non-Spin Reserves • Voltage Support • Black Start
UTILITY SERVICES	• Resource Adequacy • Distribution Deferral • Transmission Congestion Relief • Transmission Deferral
CUSTOMER SERVICES	• Time-of-Use Bill Management • Increased PV Self-Consumption • Demand Charge Reduction • Back-up Power

The list can be augmented in the future by distribution-level operational services being considered in the CPUC's Distribution Resources Plan proceeding, services such as local voltage support and power quality that would be additional utility services in the above table. Definition of distribution-level services that can be provided by storage and other DER is also being considered in the More Than Smart working group, which is an ongoing venue for stakeholders interested in the growth of DER and their impacts to discuss related planning and implementation issues.

Although some are not yet fully specified and ready to be turned into revenue streams, the list reflects existing and potential future revenue opportunities storage and other

[18] Rocky Mountain Institute Economics of Battery Storage study may be found here: http://www.rmi.org/Electricity

DERs can participate in if they have the right characteristics and, importantly, are interconnected where needed. In particular, a key insight of the RMI study is that it matters where the resource is interconnected, because it affects services and value streams the device can provide across the energy supply chain.

RMI points out that if a resource is interconnected to the ISO/RTO operated transmission system, it can offer only the ISO/RTO services, i.e., five of the thirteen services. However, if interconnected on the distribution system, in front of the customer meter, it can offer all four utility services, plus all five ISO/RTO services. Finally, a resource located behind the customer meter can offer all 13 services, four customer services and the other nine utility and ISO/RTO services. A resource's potential value and service offerings increase when it interconnects further out at the edge of the grid. This means we should expect to see use cases and configurations involving storage devices behind the customer meter designed to provide services directly to the customers where they are located and to the distribution and transmission systems. Because most of the distribution-level services identified in concept have not yet been specified in sufficient detail for implementation, we should expect configurations that serve end-use customers and participate in the ISO/RTO markets to dominate the multi-use arena in the near term.

Multi-use scenarios reflect distributed energy resource owners offering combinations of these thirteen (or perhaps more) services to the three identified stakeholders: the ISO, UDC, and end-use customer. As an industry, we need to define each service, its rules, performance requirements, measurement, etc., so the incremental value each service provides is fairly paid to each resource that provides the service while safeguarding against fraud, manipulation, and unearned revenue.

For instance, interconnecting a device at the edge of the grid enables the resource owner to capture multiple value streams, between the customer and ISO/RTO. Two problematic multi-use scenarios emerge, including variations on these scenarios, which include offering services mutually exclusive, and selling the same energy or capacity twice without adding incremental value.

Mutually Exclusive Capacity and Energy

The offering of capacity and energy services can be mutually exclusive. An example from the ISO market is that a successful bidder in the ancillary services market cannot resell the energy behind the ancillary services capacity award. For a spinning or non-spinning reserve award, the energy must be bid into the ISO market and must remain available so the ISO can dispatch it if and when needed in a contingency. The ISO has a means to monitor such activity and employs a no-pay settlement rule to subtract the ancillary services capacity payment if it finds that the energy behind an ancillary services capacity award was unavailable.

Another example of this mutual exclusivity between energy and capacity is when the capacity of a storage resource located behind a customer's meter is sold as resource adequacy capacity to an LSE, making that resource's capacity subject to a must-offer obligation. Because a storage resource has limited energy production capability, conflict can raise if the same capacity is also used to manage its host customer's demand charges and perform retail rate arbitrage. Because resource adequacy capacity comes with a must offer obligation, the energy is dedicated to the ISO, but if the resource exhausts its charge before the ISO needs to dispatch it, it will have violated its resource adequacy obligation to the ISO.

Selling the Same Energy Twice

The sale and export of energy sourced in the distribution system and sold into the bulk power system via a Wholesale Distribution Access Tariff ("WDAT") is an approved and acceptable means of providing energy services. The WDAT enables the safe and reliable interconnection of a distribution connected resource to sell its energy into the wholesale market. Other scenarios may exist that require no WDAT, but still allow resources behind the meter to export energy onto the grid, such as with Net Energy Metering ("NEM"). What must be avoided is a resource getting paid two or more times for the same energy delivered, capturing unearned value by simultaneously selling and banking the same energy.

Suppose a resource owner sells energy to the ISO/RTO from a large solar resource behind its facility meter, while the facility is enrolled under a utility's NEM tariff. The owner of the resource sets the resource up for participation in the ISO market and bids the excess energy from the resource into the wholesale market. Simultaneously, the owner "banks" the excess energy from the resource under the NEM tariff to be withdrawn and consumed by the facility at a different time. In this simple example, the resource owner would receive a double value or compensation: paid once by the ISO for wholesale energy and a second time for the value of energy withdrawn and consumed at a later time via the NEM tariff, receiving two value streams for the same energy.

In its opening comments in Track 2 of the energy storage proceeding, the ISO recommended the following to the CPUC:

1. Refine and assess the list of energy and capacity services: Start with the 13 services identified by RMI and the distribution-level services being considered in the DRP proceeding, and then refine the list in ways meaningful to the CPUC and the market structures in California. Each service type can then be evaluated against different use-cases to test for new rules, incompatibilities, and requirements, ensuring every identified service delivers incremental value when bundled with other energy and capacity services under a multi-use scenario.
2. Identify energy and capacity services already compensated: The CPUC should identify what incentives, tariffs, and rates exist that already compensate for certain energy and capacity services as identified in the RMI study and refined in this proceeding. If a multi-use scenario emerges where one or more of these services are already compensated, then such multi-use applications should be modified or rejected to account for the services already compensated.
3. Establish guiding principles: The ISO recommends CPUC staff work with interested parties to develop a set of principles that can test the validity of different multi-use scenarios. Does each service in a multi-use scenario provide incremental value, or is the same energy or capacity service being sold twice with no added benefit. Questions like these can be turned into guiding principles and are instructive for evaluating myriad different multi-use scenarios that will emerge.

4.4 Distinction between charging energy and station power

4.4.1 Background

Under this topic the ISO is working to resolve the distinction between wholesale charging energy and station power. The ISO is examining this topic area through its continued collaboration with the CPUC in Track 2 of the CPUC’s energy storage proceeding (CPUC Rulemaking 15-03-011) rather than exclusively through ESDER 2.

The ISO tariff defines station power as “energy for operating electric equipment, or portions thereof, located on the Generating Unit site owned by the same entity that owns the Generating Unit, which electrical equipment is used exclusively for the production of Energy and any useful thermal energy associated with the production of

Energy by the Generating Unit; and for the incidental heating, lighting, air conditioning and office equipment needs of buildings, or portions thereof, that are owned by the same entity that owns the Generating Unit; located on the Generating Unit site; and used exclusively in connection with the production of Energy and any useful thermal energy associated with the production of Energy by the Generating Unit."[19]

The ISO tariff explicitly states that station power includes, for example, the energy associated with motoring a hydroelectric generating unit to keep the unit synchronized at zero real power output to provide regulation or spinning reserve.[20] Importantly, because the ISO tariff allows for netting of consumption against output within a five-minute interval, station power under the ISO tariff is only measured as the amount of consumption that exceeds output within a five-minute interval.[21]

As part of the ISO's new resource implementation process, the ISO verifies that new resources have a load serving entity in place to meet station power needs prior to commercial operation. Similarly, an energy storage facility owner should consult with its load serving entity to determine how retail charges may apply to its station power consumption.

The ISO recognizes the need to further evaluate methods to distinguish between wholesale charging energy and station power and address such issues as the merits and drawbacks of treating battery temperature regulation as wholesale charging or station power; possible metering and battery configurations that would enable distinguishing among traditional station power uses, charging, and battery regulation; and any other areas where additional clarifications or enhancements to ISO rules are warranted. Revising the definition of station power to allow for energy consumed to regulate battery temperature could require revision to the ISO tariff's definition of station power, which would require FERC approval. The Federal Power Act requires equal treatment of similarly situated customers, so there would have to be a compelling difference between, for example, energy consumed to regulate battery temperature and energy consumed to start a combustion generator in order to consider one wholesale and the other retail.

---

[19] Appendix A to the ISO tariff.

[20] Station power does not include any energy used to power synchronous condensers; used for pumping at a pumped storage facility; provided during a black start procedure; or to serve loads outside the ISO BAA.

[21] See Sections 10.1.3, 10.2.9.2, and 10.3.2.2 of the ISO tariff.

The ISO also recognizes that its efforts in re-defining station power from a wholesale perspective could be unproductive if a different determination is made from the retail perspective by the CPUC.[22] The same energy could incur both wholesale and retail charges, resuscitating the years of litigation that preceded the current station power framework.[23] The ISO recognizes that its determinations regarding station power should be consistent with the CPUC's, and vice versa.

4.4.2 Current proposal

Stakeholders have generally supported the ISO's core proposal here: to modify the CAISO tariff definition of station power to exclude energy used to charge batteries for later resale. This charging load would include "efficiency losses," which are energy drawn from the grid to charge the battery for later resale, but ultimately lost because of the physics of the battery.

Stakeholders and the ISO also agree that these conclusions apply to in-front-of-the-meter applications only, and that behind-the-meter storage devices should continue to draw energy for both charging and station power at retail rates.

Finally, commenters seek additional clarification on the application of station power rules to energy storage resources. The ISO agrees that additional clarification is needed. In addition to the papers in this initiative, the ISO will revise its BPMs at the conclusion of this initiative to provide more guidance on station power generally and as applied to energy storage resources.

4.4.3 Potential enhancements contingent upon retail revisions

Station power currently is considered consumption, and not energy for resale. As such, FERC and the ISO do not have jurisdiction to effect many of the reforms sought by stakeholders, such as permitted netting.[24] For example, if the ISO were to amend its tariff to allow energy storage resources to net station power from output both when the resource is discharging *and charging* (thus treating negative generation as positive generation for settlement purposes), it would not have any effect on the energy storage

[22] See, e.g., *Southern California Edison Co. v. FERC*, 603 F.3d 996, 1002 (D.C. Cir. 2010)

[23] See, e.g., *id.*; *Calpine Corp. v. FERC*, 702 F.3d 41 (2012); *Duke Energy Moss Landing LLC v. CAISO*, 134 FERC ¶ 61,151 (2011).

[24] *Duke Energy Moss Landing LLC v. CAISO*, 134 FERC ¶ 61,151 at P 2 (2011) ("state-jurisdictional retail sales are properly the subject of state tariffs").

resource's retail rates and would instead result in confusion and litigation, just as it did when FERC mandated different netting periods than the California IOU retail tariffs.[25] As the court said there, "It is, of course, true that under differing netting periods FERC can conclude that no transmission for station power took place in a month in which California would recognize retail sales of that power, but that is hardly a conflict."[26] In other words, the ISO cannot mandate universal or sole netting rules—they can be different and thus overlap with retail netting rules.

Metering is another example: The retail tariffs have jurisdiction to impose specific metering requirements to distinguish station power from charging energy, although the ISO has jurisdiction to impose specific requirements to distinguish charging energy from station power.

The ISO will therefore continue to work with stakeholders in the CPUC energy storage proceeding on those issues where the retail tariffs should be revised first. If and when they are, the ISO will amend its tariff for consistency.

Based on stakeholder comments in ESDER 2 and the CPUC energy storage proceeding, the ISO speculates that the most critical and most likely revisions to the retail tariffs (and then the ISO tariff) will address the two examples described above: netting and metering. The ISO therefore takes this opportunity to propose its views on these topics so that stakeholders can comment. It is, however, important to keep in mind that while the proposals in section 4.4.2 above could be implemented immediately after Board approval, the proposals below would need to be addressed in the CPUC proceeding and the retail tariffs first.

Permitted Netting

In its comments, LS Power does not draw a distinction between energy storage and conventional generation as resource classes, but argues that the station power rules currently apply to the two classes differently because energy storage must charge throughout the day: "The key difference from traditional generation is that energy storage can provide Negative Generation, which does not align with the current constructs of Permitted Netting," which was created "with the assumption that all resources had positive generation ranges only."[27] LS Power argues that station power

---

[25] *Southern California Edison Co. v. FERC*, 603 F.3d 996, 1000 (D.C. Cir. 2010).

[26] *Id*. at 1002.

[27] LS Power R.15-03-011 Workshop Comments at 3.

for energy storage "should be treated the same way as it is for conventional power plants to the greatest extent possible. As such, situations where the energy storage system is either offline or online and discharging are exactly analogous to conventional power plants, and should be treated the same way." LS Power argues that a simple way to ensure fair treatment would be to amend the netting rules to allow station power load to be netted against both discharging *and charging*, in essence treating negative generation as if it were positive generation.[28]

LS Power demonstrates how the current station power and netting rules may result in an energy storage resource and a conventional generator having very different station power and energy supply settlements on a given day. LS Power evaluated an ISO node in the Bay Area on May 20, 2016, a day that exhibited the pronounced "duck curve" for which storage resources are sought to mitigate.[29] It then simulated the performances of a 100 MW natural gas peaker plant and 100 MW battery system, where both are configured to sell at any price greater than $50/MWh, and the battery system has bid into the market that it will buy/charge at any price below $15/MWh.[30] The gas peaker has a start-up spike in its station power use to 2.5 MW. Both the gas peaker and battery system are assumed to have 2 MW of various necessary auxiliary loads during operation, and this 2 MW is inclusive of the idle loads. The retail rate for station power is assumed to be $0.15 / kWh.[31] The following tables show the resources' output and auxiliary loads throughout the day:

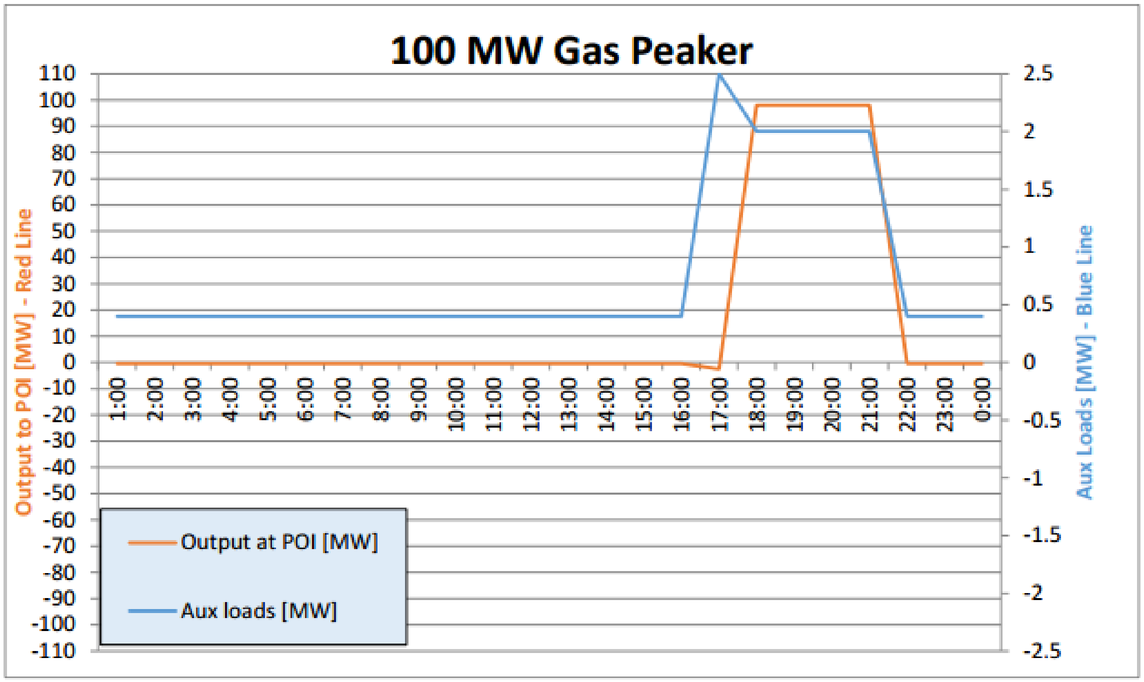
---

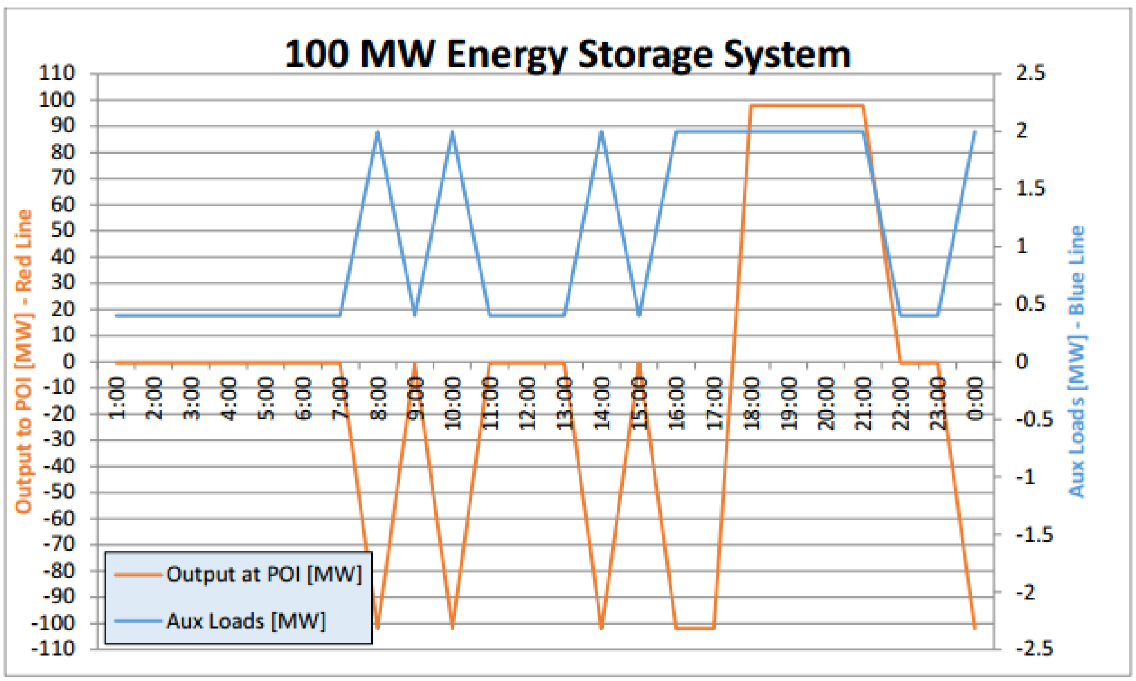
[28] *Id.* at 4.

[29] *Id*. at 14-20; For more information on the duck curve, *see* https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf.

[30] The battery system would be rated for 400 MWh of storage (nameplate power for 4 hours), has a round-trip efficiency of 80%, and starts the day at 0% SOC. Both the gas peaker and BESS are assumed to have idle loads of ~0.4% of their nameplate output, similar to the PV example previously. The gas peaker has a start-up spike in its station power use to 2.5 MW. Both the gas peaker and BESS are assumed to have 2 MW of various necessary auxiliary loads during operation as described above, and this 2 MW is inclusive of the idle loads. The retail rate for station power in this example is $0.15 / kWh.

[31] LS Power ESDER 2 Revised Straw Proposal Comments at 14-15.


100 MW Gas Peaker
Output to POI [MW] - Red Line
Aux Loads [MW] - Blue Line
Output at POI [MW]
Aux loads [MW]
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1:00 2:00 3:00 4:00 5:00 6:00 7:00 8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00 0:00


100 MW Energy Storage System
Output to POI [MW] - Red Line
Aux Loads [MW] - Blue Line
Output at POI [MW]
Aux Loads [MW]
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1:00 2:00 3:00 4:00 5:00 6:00 7:00 8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00 0:00

These are the results for the two resources:

	Gas Peaker	Battery System
Revenue	$51,921.12	$51,921.12
Wholesale Station Power	($1,038.42)	($1,038.42)
Retail Station Power	($1,515.00)	($2,640.00)
Total Station Power[32]	($2,533.42)	($3,678.42)
Net[33]	$49,367.70	$48,242.78

The battery system ultimately pays more for station power because the battery system's auxiliary loads rise not only when it supplies energy to the grid (like the gas peaker), but also when it must charge. LS Power notes that because the battery system is unable to net its station power consumption against any supply at these times, the battery system must procure energy for its station power at a retail rate.

Other stakeholders such as CESA reach the same conclusion as LS Power.[34] While the ISO's views on these issues are evolving, the ISO currently agrees that charging or negative generation should be treated as positive generation, which would allow

[32] This number could vary depending on the metering configuration. LS Power's simulation notes that these figures derive from a single-meter configuration, but that a dual-meter configuration with a calculated idle loss retail charge would result in a total station power charge of $3,840.

[33] The purpose of the data is to demonstrate how station power costs affect cost-effectiveness, and as such, this table omits any hypothetical fuel costs for the gas peaker. The wholesale charging energy for the battery system in this hypothetical would have been $7,825.74.

[34] This includes other stakeholders who comment in the CPUC proceeding but not on the ISO's Revised Straw Proposal. *See, e.g.*, NRG Workshop Comments in R.15-03-11.

storage devices to "net" their station power consumption against charging. Insofar as a resource is withdrawing energy or injecting energy subject to a ISO dispatch at a greater capacity than its consumption, that consumption should be able to be netted against the response to the ISO dispatch, just as it is for conventional generators. Doing so properly recognizes that resources of the future will be "bi-directional," in that they can provide grid services through both consumption and generation.[35] Recognizing bi-directional market services and properly incentivizing them also will benefit the grid by helping to mitigate the issues resulting from the duck curve. It also encourages resources to provide ancillary services to the ISO markets, which promotes competition, lowers prices, and provides greater reliability to the grid.

On the other hand, the ISO disagrees with some in the storage community who argue that energy storage resources should be subject to a wholesale rate when they are idle, bidding, or awaiting a ISO dispatch. Energy storage resources should not be able to net their consumption when idle or when charging or discharging less than their on-site consumption in a given settlement interval.[36] Amending the permitted netting rules to allow such netting would not promote the benefits described above, and instead would instead incentivize energy storage resources to remain idle.

Settlement and Metering

To date, the ISO, the CPUC, and stakeholders' principal focus has been defining station power and identifying rules that can ensure energy storage resources are on a level playing field with conventional generators. Metering has largely been a peripheral issue, but it will become a critical issue based on how the CPUC decides to define station power and what rate the CPUC then applies to station power. If the CPUC decides to

---

[35] As discussed above, the CAISO's ESDER 2 initiative is currently exploring the ability for proxy demand resources to be dispatched to both curtail and increase load and provide regulation, for example.

[36] To be sure, where an energy storage resource responds to dispatches both positive and negative in a given interval, those responses would be cumulative and not netted for settlement purposes, such that the total response would be netted against the resource's station power consumption. An energy storage resource using the CAISO's Regulation Energy Management might, in a given interval, make many rapid shifts between positive and negative generation. Assume this resource consumes 2 MWh in a given interval for station power purposes, and that it in the first third of the interval it withdraws 1 MWh from the grid, in the second third it injects 1 MWh onto the grid, and in the third interval it again withdraws 1 MWh from the grid. Its net dispatch performance would be 3 MWh (not -1 MWh), such that its response would be greater than its on-site consumption. As such, it would be a net "supplier" to the grid and could procure subject to a wholesale rate.

maintain existing rules without amending the permitted netting rules or the station power definitions, it will be necessary for the ISO, the LSE, and the storage resource to be able to determine for settlement purposes what energy was withdrawn to charge the storage resource, and what energy was drawn from the grid for station power. Conventional generators only draw from the grid energy for consumption, and therefore can easily use a single-meter configuration. The same cannot be said of energy storage resources, especially if some of their station power load (e.g., thermal regulation of the battery) is so ingrained into the storage resource that a separate meter in front of the storage resource may not be able to adequately separate station power load from charging. Moreover, trying to separate station power load from efficiency losses is equally difficult.

Energy storage resources and LSEs could rely on predetermined station power and/or efficiency loss calculations from which to subtract or add values for settlement purposes, but these calculations can be difficult to determine and may not be consistently accurate in real-time based upon the performance of the storage resource.

Most critically, energy storage resources are a nascent technology and it is often the case that these resources present wholly new, unique configurations to the ISO and the LSE. For these reasons, the ISO believes that neither it nor the CPUC should mandate a one-size-fits all metering rule for energy storage resources at this time. Instead, the ISO believes that the only requirement should be for energy storage resources and LSEs to find a mutually agreeable metering configuration that will provide cost-efficient and sufficiently accurate meter data such. Some energy storage resources may require or desire additional meters for optimal accuracy, some may require fewer meters, and some may only require a single-meter with predetermined deductions and additions. Moreover, metering issues may become much simpler to address if the CPUC amends its netting rules to allow storage resources to treat charging as generation, thus obviating the need to distinguish battery load from station power load. But in any case, the ISO believes that presently the energy storage community, the conventional generation community, the LSEs, and the ISO lack the experience with energy storage configurations to recommend a one-size-fits-all metering configuration.

5 Stakeholder process schedule

The following table outlines the schedule for the policy development portion of ESDER 2. This schedule does not include implementation steps including development and filing of tariff amendments, changing business process manuals, and making and implementing changes to market system software and models.

Stakeholder Process Schedule		
Step	**Date**	**Activity**
Issue Paper	March 22	Post issue paper
	April 4	Stakeholder web conference
	April 18	Stakeholder comments due
Straw Proposal	May 24	Post straw proposal
	May 31	Stakeholder web conference
	June 9	Stakeholder comments due
Revised Straw Proposal	July 21	Post revised straw proposal
	July 28	Stakeholder web conference
	August 11	Stakeholder comments due
Second revised straw proposal	September 19	Post second revised straw proposal proposal
	September 27	Stakeholder web conference
	October 11	Stakeholder comments due
Additional papers as needed	TBD	TBD
Board approval	TBD	ISO Board meeting