FREEMAN, SULLIVAN & CO.

A MEMBER OF THE FSC GROUP

2012 Impact Evaluation of Ontario Power Authority's Commercial & Industrial Demand Response Programs

Prepared for:
The Ontario Power Authority

August 2013

Prepared by:
Josh Bode, M.P.P.
Jeeheh Oh, B.S.
Aimee Savage, B.S.
Freeman, Sullivan & Co.

Freeman, Sullivan & Co.
101 Montgomery St., 15th Fl
San Francisco, CA 94104
fscgroup.com

Message from the Evaluator

I fully endorse the process used to generate the evaluation results for OPA's Commercial & Industrial Demand Response programs, as documented in this report.

Stephen S. George
Chief Executive Officer and Principal Consultant
The FSC Group

July 30, 2013

Acknowledgements

We would like to thank Nik Schruder and Jessei Kanagarajan at OPA, John Antonakos at the IESO, as well as Stephen George at The FSC Group, for their helpful comments, guidance and insight.

Table of Contents

1 Executive Summary

This report documents the 2012 evaluation of the Ontario Power Authority's (OPA) DR-2 and DR-3 programs. Considering there were no active participants or events for DR-1 in 2012, it is not included in the evaluation.

DR-2 is a contractual load shifting program in which participants specify the load shift window and amount of load shifting. The program does not compensate participants for changes in the overall level of electricity consumption that may be associated with economic fluctuations. Load reduction during peak periods must be accompanied by shifting to off peak periods.

DR-3 is a contractual demand response (DR) program that offers large commercial and industrial consumers reserve payments for being available to provide load reductions when called by the OPA. Participants are required to meet contractual demand reductions when called, or else payments are reduced or eliminated.[1] In addition, DR-3 participants notify OPA and the IESO of any short-term fluctuations in load reduction capability due to facility maintenance or down time. Nonperformance days lead to reductions in the participant payments. Unscheduled nonperformance – failure to deliver scheduled load reductions during events – leads to even larger payment reductions.

1.1 Summary of Evaluation Goals and Objectives

The DR-2 and DR-3 evaluations are designed to meet the following objectives:

- Estimate regional and providence wide load reduction capability (i.e., *ex ante* load reductions) for each program;
- Estimate regional and provincial load reductions and annual electricity savings delivered by each program in 2012 (i.e., *ex post* load impacts);
- Analyze how impacts vary for different customer segments including:
    - Direct participants versus aggregators;
    - Industry type;
    - Location; and
    - Customer size
- Determine if performance of contributors and direct participants has improved; and
- Analyze key drivers of performance, including planned nonperformance, settlement baseline error, event-day conditions and concentration of resources.

1.2 Summary of Impact Evaluation Results

Table 1-1 summarizes the evaluation results. DR-2 had two contributors in 2012 with 106 MW of contracted load. Enrolment for DR-2 did not change from 2011. DR-3 added 89 new contributors in 2012, accounting for 27 MW of summer contracted load reductions. By the end of 2012, there were 526 total contributors in the DR-3 program with an aggregate summer contracted load reduction of 408 MW. The mix of participants in DR-3 consolidated, particularly among aggregators. In 2011, DR-3 had three direct participants and five aggregators. In 2012, two of the aggregators exited OPA's program and sold their demand response resources to another aggregator.

[1] Compliance is determined by the settlement baseline method, which differs from the evaluation impacts.

Table 1-1: Summary of Impact Evaluation Results

Program Metric	DR-2	DR-3
Number of Participants / Contributors in Program	2/2	8/476
Total Contracted MW Enrolled in Initiative[2]	106	408
Ex ante Load Impact Estimate (MW)[3]	53.6	343.6
Ex post Energy Savings (GWh)[4]	73.9	9.2

Key findings from the DR-2 evaluation include the following:

- *The DR-2 program led to distinct changes in customer loads and reduced the volatility of use during peak hours.*
- *The demand reductions were not fully matched by shifting to off peak hours.* Overall, electricity consumption for DR-2 participants is approximately 30% lower than their consumption prior to enrolling in DR-2. Given their production capacity, the participants are unable to fully shift reduction during peak periods to off peak periods and fully comply with DR-2 program rules.
- *A significant portion of reductions occurred outside of customers' contracted windows.* In addition, customers did not match the scheduled reductions well. This is in part due to the process driven, step-like load patterns of participants.
- *Ex ante impact estimates use a balanced approach.* As a lower bound, FSC used estimates of load shifting that assumed any changes in overall electricity consumption levels cannot be attributed to DR-2. As an upper bound, it was assumed that the demand reductions and the lower consumption levels from DR-2 observed between 2009 and 2012 will be sustained.

Key findings from the DR-3 evaluation include the following:

- *Program performance improved*. The gross *ex ante* load impact estimate for DR-3 is 343.6 MW,[5] representing 84% of the peak summer contracted capacity (343.6 MW out of 408 MW).
- *The program diversified but still remains highly concentrated.*
- *The largest 20 contributors account for 60% of the contractual demand reduction – that is, less than 5% of contributors account for the majority of the load reductions.*
- *The current baseline method for DR-3 is biased and, in aggregate, overstates demand reductions by approximately 22%.* The exact effect on individual settlement accounts varies widely – some aggregators are underpaid while others are overpaid relative to their actual demand reduction. Since aggregators comply with the baseline rules, correcting the

[2] This value is analogous to the nameplate capacity. For DR-2 it represents the peak contractual demand reduction during the 7 AM to 7 PM window. Actual values vary by hour. For DR-3, the contracted MW reflects the aggregate of the peak summer load reduction capability specified for each contributor. This value is relatively constant. Some aggregators adjust their settlement account contractual obligations to vary by season or intentionally derate the value to manage risk. The contracted MW does not factor in those adjustments. They are factored into the *ex ante* load impact estimates.

[3] *Ex ante* impacts reflect the load reduction capability and factor in historical performance. For DR-2, *ex ante* impacts are based on the 2-3 PM period. For DR-3, they are based on performance for historical event hours from 2008 to 2011.

[4] DR-2 energy savings reflects 50% of the savings calculated for the year. It is a middle ground approach between assuming the savings are all due to DR-2 (108.5 GWh) and assuming that the load shift program does not produce any energy savings whatsoever. In prior years (2009 & 2010), the higher value was reported. DR-3 values reflect energy savings during all 24 hours of each day and event was called.

[5] This *ex ante* value factors in the amount of contracted load reduction, the share of contracted reduction scheduled on a day-ahead basis and the degree to which these customers have historically delivered scheduled demand reductions.

baseline should better align settlements with actual demand reductions and reduce over and underpayments.

1.3 Conclusions and Recommendations

Both DR-2 and DR-3 performance in 2012 was stable relative to 2011, but the performance for both programs can be improved by better aligning them with system value. A key question is whether and how OPA should modify its DR programs in light of Ontario's changing system needs.

FSC made two main recommendations for the DR-2:

- *FSC does not recommend continuing the DR-2 contracts with existing participants beyond the current contract period.* There are two main reasons for this recommendation. The estimate of the load shifting and energy saving relies on constructing baselines using data from 2009 and earlier. The baselines, and by connection the shifting and savings estimate, becomes less reliable as time elapses. The second reason is to determine whether or not the change in load shapes persists after the contracts expires. Since the program led to fundamental changes in production processes and schedule of participants, they may continue their consumption patterns.
- OPA should consider whether to continue a load shifting program in the future but with a revised design and different target contributors. Load shifting can provide significant benefits, particularly given the current state of Ontario's electricity system, which currently has surplus generation during off peak hours on shoulder months. A revised load shifting program should allow aggregation across customers and, ideally, focus on Class B customers, which are not exposed to wholesale market prices. A revised load shifting program also should strongly consider narrowing the load shifting period to hours with the most values or providing higher incentives for the most critical hours.

For DR-3, FSC provides several recommendations:

- *Improve the data tracking and collection process.* The current process is highly manual and introduces significant risk for error. It also makes comprehensive analysis difficult and prolongs the settlement period with DR-3 participants. The data files are kept separately for each customer and month on separate spreadsheets that are not consistently labeled. As part of the evaluation, FSC was unable to obtain data for contributors that accounted for 8% of the contracted resources despite repeated efforts.
- *Ensure that settlement accounts reflect aggregation.* The empirical evidence shows that the accuracy of settlement estimates degrades severely with settlement accounts that fail to aggregate. FSC recommends that no single contributor should be more than 20% of load in the settlement account. There may be exceptions for load that are proven to be highly predictable using out-of-sample testing. Currently, DR-3 participants are required to have separate settlement accounts by IESO zone. While the ability to dispatch locally is important, it is not necessary to settle by IESO zone for each event. Doing so limits the benefit of aggregation and introduces a substantial amount of payment error – many accounts are either over or under compensated due to the inherent noise associated with disaggregated baseline results.
- *Require settlement accounts to reduce at least 10% of historical 12 PM to 9 PM summer weekday demand.* The empirical evidence shows that settlement estimates are unreliable when the percent reductions are small. Therefore, by requiring a minimum percent reduction for settlement account, it is easier to distinguish actual demand reduction from inherent variability in loads – background noise.
- *Conduct tests to assess if DR-3 can be used to specify specific amounts of DR resources for specific hours under the current rules*. The current practice of calling either all customers on the 200 hour option and all customers on the 100- and 200-hour options for the same hours, does not provide the flexibility necessary for system operations.

- *Consider developing a different DR product for large customers (average monthly peak usage greater than 5 MW) that better utilizes demand reductions for operations and does not rely on settlement baselines (which are inherently less accurate for individual customers).*

2 Introduction

This report contains *ex post* and *ex ante* load reduction estimates for the Ontario Power Authority's DR-2 and DR-3 programs. Both programs target commercial and industrial customers; combined, they represent the Demand Response C&I portfolio. These programs have substantial differences in their design.

DR-2 is a load shifting program in which participants contract to shift specific amounts of load from peak hours. Participants have the ability to specify the amount of load and hours of load shifting on a monthly basis. The program only compensates customers for shifting load from peak hours and does not provide payment for overall reductions in energy consumption that may be associated with economic conditions.

DR-3 is a contractual demand response (DR) program that offers large commercial and industrial consumers reserve payments for being available to provide load reductions when called by the OPA. Participants are required to meet a 95% reliability requirement when they are called, or else financial penalties apply. For settlement purposes, compliance with the contracted load reductions is determined through day-matching baseline methods and excludes planned nonperformance.

In 2008 and 2009, the DR-1 program was also evaluated. Considering there were no active participants or events in DR-1 in 2010 or 2011, it is not included in this year's evaluation. DR-1 is a voluntary DR program offered by the OPA to pilot the concept of DR among large industrial and commercial customers. It is a demand bidding program in which participants are paid for reducing load when Independent Electricity System Operator (IESO) market prices exceed a strike price proposed by participants. Most of the former DR-1 participants have transitioned to DR-2 or DR-3. A revised, voluntary DR-1 program was launched in 2011 as part of the OPA's Conservation Demand Management (CDM) Portfolio; however, there were no active participants in 2011 or 2012.

Table 2-1 provides a comparison of the key features of each program and the recent program experience. DR-2 reached a steady state in 2010 and the program did not experience growth. DR-3 added 50 new contributors in 2012, accounting for 25 MW of contracted load reductions. By the end of 2012, there were 526 total contributors in the DR-3 program with an aggregate contracted summer capacity of 408 MW.

Table 2-1: Comparison of OPA Commercial and Industrial Demand Response Programs

Program Feature	DR-2	DR-3
Program Description Summary	A load shifting program in which participants contract to shift a specific amount of load reduction from peak to off peak hours. Participants have the ability to specify the amount of load and hours of load shifting on a monthly basis.	A contractual DR program that offers aggregators or direct participants availability payments for being available to provide load reductions when called by the OPA.
Amount of Contracted Resources	Varies by hour and is, at most, 106 MW of load reduction.	408 MW of contracted summer capacity at end of 2012.
Number and Concentration of Customers	Three very large participants, with 119 MW of contracted load shifting (153 MW during transition).	526 contributors. The top 20 contributors account for 60% of the contracted load reduction.
Event Frequency	DR-2 is a permanent load shift program operating on all non-holiday weekdays.	A total of five events were called in 2012.
2012 Program Growth	DR-2 did not experience growth in 2012.	DR-3 experienced a small amount of growth during 2012.
Nonperformance Rules	Participants subject to set-offs for nonperformance at the hourly, daily and monthly levels. Set-off charges are typically equivalent to payments for performance. Set-offs may be mitigated through advanced warning by participants.	Participants incur penalties if they provide less than 85% of their scheduled MW amount relative to the baseline.
Length of Program Experience	Program has been active since 2009.	Program has been active since 2008.
Settlement Baselines	Baselines based on pre-contract period and adjusted for changes in consumption.	Highest 15 out of 20 past days, excluding nonperformance days.

2.1 Evaluation Objectives and Key Research Questions

This report focuses on the evaluation of 2012 *ex post* and *ex ante* load reductions. The 2012 program evaluation of these programs is designed to meet multiple objectives, including:

- Estimate regional and providence wide load reduction capability (i.e., *ex ante* load reductions) for each program;
- Estimate regional and provincial load reductions and annual electricity savings delivered by each program in 2012 (i.e., *ex post* load impacts);
- Analyze how impacts vary for different customer segments including:
    - Direct participants versus aggregators
    - Industry type
    - Location
    - Customer size

- Determine if performance of contributors and direct participants has improved; and
- Analyze key drivers of performance, including planned nonperformance, settlement baseline error, event-day conditions and concentration of resources.

2.2 How to Use Load Reduction Results

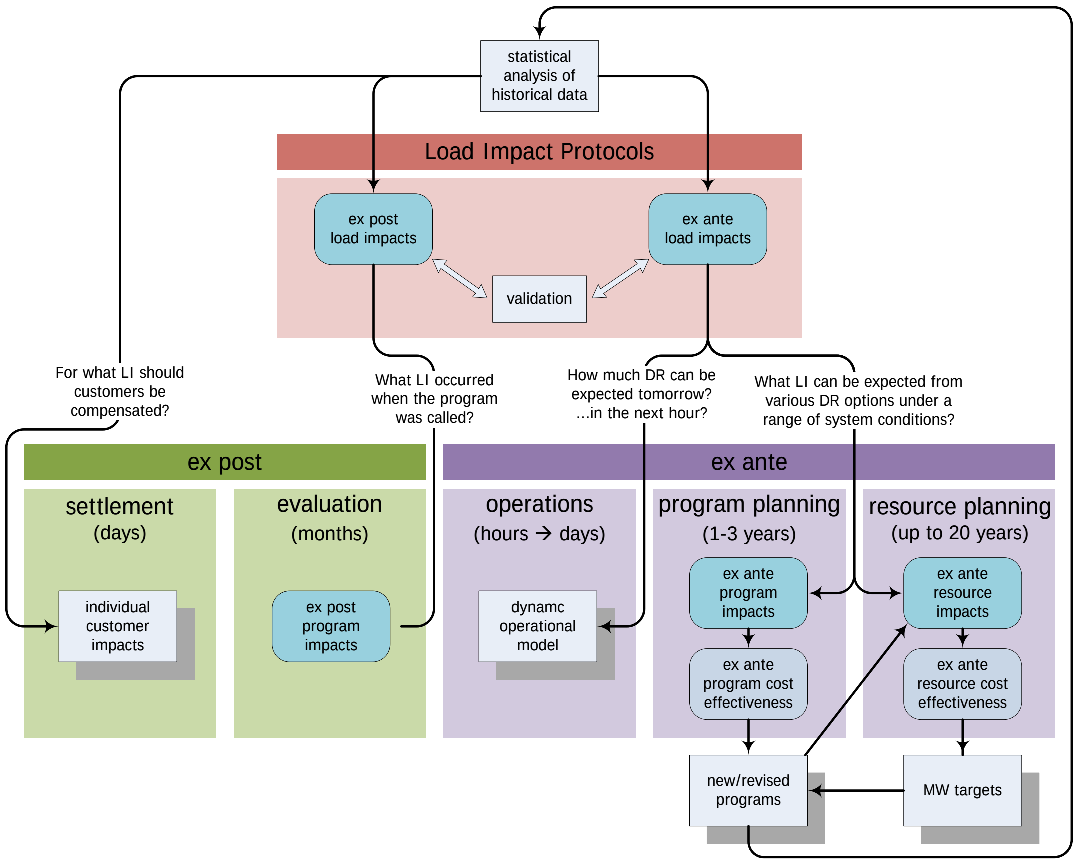
There are various ways in which to use DR load reduction estimates. The estimation methods and output requirements may vary depending on how the estimates are used. The estimates presented here include *ex post* analysis for events called in 2012 and *ex ante* estimates of the load reduction capability. The analysis also provides useful input for improvements in program design and operation.

Figure 2-1, taken from OPA's Load Impact Protocols, provides an overview of the various ways in which load reduction estimates can be used and the interrelationship between *ex post* and *ex ante* analysis. *Ex post* load reductions determine what happened over a historical period, based on the conditions that were in effect during that time. Because historical performance is tied to past conditions such as weather, price levels and dispatch strategy, *ex post* load reductions may not reflect the full option value of a DR resource. For example, during several event-days in 2012, the full load reduction capability of the DR-3 program was not utilized because not all customer groups (e.g., 100-hour customers vs. 200-hour customers), were called for each event. As such, it would be inappropriate to use *ex post* load reductions for long-term planning, because planning should be based on the full option value of the resource and its impact on overall resource needs.

Ex ante load reductions are forward looking and are designed to reflect the load reduction capability of a DR resource under a standard set of conditions that match the market and system conditions that drive the need for additional capacity. Typically, *ex ante* load reductions are based on regression models developed from DR program data and customer behavior (that is, *ex post* analysis), but the models allow for adjustments to be made in the estimates to reflect the appropriate *ex ante* event conditions. After a program matures and the participant mix becomes stable, *ex ante* load reductions can also be used to estimate future expected load reductions. *Ex ante* load reductions are an important input to DR cost effectiveness analysis, both for program planning (comparing different DR program designs) and resource planning (comparing archetypal DR options against other conservation and supply resources).

Both DR-2 and DR-3 have prespecified contractual obligations and mostly consist of weather insensitive industrial customers. As a result, the focus of the *ex ante* impacts is on the share of contractual load reductions that is realized after accounting for both preannounced and unannounced nonperformance.

Figure 2-1: Context and Uses For Demand Response Evaluations



The regression model estimates can also contribute to program and day-to-day operations. They can form the basis for developing dispatch models to predict, in near real time, the impact of DR programs as they are activated, provided the participant mix does not change substantially.

Although not required by the OPA Load Impact Protocols, this evaluation also provides useful analysis of the reliability of DR as a resource, both in the context of operations and long-term planning. System operators and planners have different evaluation needs. For system operators, differences between scheduled and delivered load reductions are critical for scheduling resources to balance electricity supply and demand on a daily basis. On the other hand, for planning purposes, the focus is on the share of the contracted load reductions that are realized. Estimates for planning need to factor in both preannounced and unannounced nonperformance since both affect the ability to meet extreme peak demands.

2.3 Reliability of Demand Response as a Resource

In reporting results, FSC attempted to provide information that enable readers to assess the reliability of demand response resources. The term 'reliability' is often used loosely in discussing system operations and long-term planning. For clarity, throughout the report, FSC distinguished between operational reliability and capacity realization. The distinction is highly analogous to generator operations since a key goal is to enable comparisons between generation and DR resources.

For operation reliability, the key metric is the load reduction as a percentage of the scheduled demand reduction resources, after accounting for the unavailability that was preannounced to the system operator. For long-term planning, the key metric is demand reductions as a percentage of the nameplate capacity. Throughout this report, the summer demand reduction capability reported by aggregators for each contributor (Summer contracted MW) is treated as the nameplate capacity.

Operational reliability affects the ability to balance the grid on a short-term basis. It is affected by the accuracy of demand and supply forecasts and by the amount of operating reserves in place that enable the system to recover from shocks such as transmission and generator forced outages. The primary evaluation output of operational reliability is whether the scheduled resource is delivered when called upon. In the context of generators, notification of scheduled outages or prolonged forced outages allows the system operator to schedule additional resources for operations. In addition, the system operators will have operating reserves available in case of unforeseen generator or transmission outages. In context of contractual DR resources, operational reliability factors out planned nonperformance reported to the system operator as well as partial de-rating of resources reported by the DR-3 participants. These are analogous to preannounced generator outages in that participants are notifying the system operator not to expect or rely on them as a resource for a specific period. The operational reliability is the share of the scheduled resources that is indeed delivered. The operational reliability is based on the accuracy of the resources' predictive models, how well the resource is incorporated into operations and whether the expected resources are indeed delivered.

In contrast, the capacity realization rate is designed to assess the extent in which the resources can be factored into mid-term and long-term planning. For generation, the standard practice in planning is to derate nameplate capacity for both scheduled and unscheduled outages. More sophisticated approaches factor in the timing and likelihood of scheduled and forced outages. For contractual DR resources, both planned and unplanned nonperformance are factored into the capacity realization rate and are used to derate the contracted load reduction resource.

In using *ex ante* load reduction estimates for planning purposes, it is also critical to link DR resources to relative need for capacity or the time-differentiated capacity value. While the *ex ante* reflect the net load reduction capability, they do not directly account for the timing and magnitude of load reductions. They also do not account for limitations on the maximum event duration, hours of availability, amount of advance notice, number of consecutive event-days the resource is available and other factors. From a capacity perspective, resources available in the summer mid-afternoon hours, which are the hours of system peak, generally have higher value than resources in off peaking hours. Resources that can sustain reductions (or supply) for longer periods also have greater value because the risk of installed capacity shortages during system peaking conditions sometimes include more than four hours. For planning, it is necessary to estimate and incorporate the extent to which DR resources serve as capacity without increasing the likelihood of unserved electricity – a value more commonly known as effective load carrying capacity. This typically requires subsequent analysis and adjustments and is done in the long-term planning analysis.

2.4 Report Structure

The remainder of this report is divided into five main sections. Section 3 describes the regression model development, summarizes the evaluation approach employed for each program and describes the checks FSC conducted to ensure results are accurate and valid. More detailed discussions about the evaluation methodology for each program are presented as appendices. Sections 4 and 5 respectively present the DR-2 and DR-3 results. Each of these sections provides background information about the program, participant mix and events called in 2012 (if applicable). This is followed by a discussion of the 2012 *ex post* load reduction results, including how load impacts varied by industry, geography and size. Sections 4 and 5 finish with a presentation of the *ex ante* load reduction estimates for long-term planning, which incorporate adjustments for scheduled nonperformance and deviations from scheduled demand reductions.

3 Evaluation Methodology

To calculate load reductions for demand response programs, the participants' load patterns in the absence of program participation must be estimated. For most DR programs, this is accomplished by using pre-enrolment data, observing behavior during event and non-event-days, use of an external control group or a mixture of the aforementioned. The most rigorous method for impact evaluations is a well executed experiment with random assignment to either a control or treatment condition. While randomized experiments are rarely feasible for actual programs, there are multiple available methods for assessing reductions that approach the rigor of experiments. The best available method is a function of the program characteristics, available data, the ability to incorporate research design elements and statistical methods.

In all instances, the evaluation used unperturbed load data prior to any program enrolment and post-enrolment data. None of the evaluations used an outside control group. Instead, the customers were used as their own control group. DR-2 and DR-3 are each dominated by very large commercial and industrial customers. In Ontario, there are a limited number of similar facilities, impairing the ability to develop an external control group.

Regression methods were employed to analyze DR-2 and DR-3 customer load patterns with or without event participation for load shifting. The analysis consists of applying regression models separately to each set of customer load data at the hourly level. Because the coefficients are customer specific, they can better explain the variation in individual customer production and/or occupancy patterns, weather sensitivity, price responsiveness, enrolment dates, planned nonperformance and event-day dispatch patterns. Four models were tested for each customer. Different models had different temperature variables but all models took into account the previous weeks' load. Wholesale customers had models with market price variables and GAM eligible customers had models that took into account GAM event-days. After conducting out-of-sample testing, the model that best predicted actual load on proxy event-days (days similar to event-days but are not event-days) was used in the analysis.

While the evaluation relies on regressions to estimate demand reduction, settlement for DR-3 and DR-2 relies on day-matching baselines. Baselines are used for settlement because they can be calculated quickly and are easier to understand without technical training in statistical and econometric methods. Regression techniques have several distinct advantages over the day-matching baseline methods typically used for DR program settlement. First, they can help identify the key drivers and predictors of load patterns and load reductions. Second, regression results provide more robust estimates of load reductions and are not as sensitive to biases in the reference load. Third, they can be used to predict load reductions for operations or for long-term planning by factoring in expected weather, system and market conditions, if available. In other words, they can better predict load reductions on a day-of or day-ahead basis in order to assist operations as well as predict load reduction resources under system peaking conditions in order to assist in long-term planning. For both regression and baseline methods, it is critical to conduct a number of validity tests to determine the ability of the evaluation approach to produce accurate results.

The remainder of this section summarizes the methods used for the DR-2 and DR-3 evaluations and discusses the tests and checks undertaken to ensure reduction estimates are accurate. Appendices A and B provide more detailed discussions of the methodology for each evaluation and present the results of several tests conducted to ensure the models produced accurate results.

3.1 DR-2 Analysis Approach Overview

DR-2 is a unique program and presents several challenges from an evaluation perspective. Once a customer is enrolled in DR-2, it is not possible to observe their behavior absent the requirement to shift loads since, based on program rules, customers shift loads on a daily basis. Given the limited number of facilities of similar size and industry, it is also not possible to develop a valid external control group. As a result, the evaluation relies heavily on customer electricity use patterns prior to their enrolment on DR-2. The introduction of DR-2 daily load shifting creates a shift in the hourly electricity usage patterns and volatility that should occur in tandem with enrolment in the program. Technically, this analysis approach is referred to as an interrupted time series and is the primary evaluation method for DR-2.

The key to an interrupted time series analysis is knowing the exact point at which an intervention occurred. In the case of DR-2, the timing of when customers were expected to change load patterns is well known. So is the expected change in demand. In addition, participants are required to eliminate volatility in the hourly use patterns during the peak period. As a result, with the interrupted time series analysis, it is possible to observe: 1) if the customer load shapes change in tandem with their start date and 2) if the volatility in the electricity use during load shift hours is also reduced in tandem with their start date.

Part of the challenge for DR-2 is the multiple interventions to participant load that occurred since 2007. In 2007 and 2008, they participated in DR-1, which altered their behavior during event-days. From March to October 2009, they were engaging in load shift during the transition period. For November and December 2009, there was a different intervention in place – namely, the DR-2 full program rules were in effect and the settlement baseline calculations were updated to incorporate the DR-1 evaluation data. A strong point of this evaluation is that the timing of each intervention (i.e., shift in DR participation) is well documented. The downside is that disentangling the effect for each of the programs requires us to rely on historical participants that are distant from the evaluation period.

Since the start of the DR-2 program, participants have consumed less electric power than they did prior to enrolling in DR-2. The difference can be observed in two ways. First, the number of days when the facilities shut down has increased. However, participants inform the IESO of planned nonperformance days when they are not available. Second, during the days they are in operation, their overall consumption is generally lower.

The introduction of DR-2 coincided with the economic downturn in 2009. While the Canadian economy started to recover in 2010, the frequency of the facility shut downs indicate that DR-2 participants, all of which are pulp and paper plants, have not. In fact, one of the three plants enrolled in DR-2 permanently shut down in April 2011. The timing of the downturn raises the question of whether or not changes to electricity consumption are entirely due to DR-2 or can be attributed to other factors.

The regressions provide the hourly reductions of the program but the load shifting to off peak periods does not always match the reduction during the peak period. This reflects the fact that some customers cannot fully shift their load reductions to off peak periods due to the physical capacity of the plant. Most of these customers exhibit step-like changes in demand driven by the underlying processes in operation. Given the rules that shifting has to occur to periods outside of the 7 AM to 7 PM period, these customers cannot replicate their baseline (pre-DR-2) power consumption and at the same time comply with program rules.

To assess the degree of load shifting, FSC focused solely on changes in the load shapes and developed a process where the electricity consumption with and without DR is identical by construction. To do so, FSC had to normalize the pre-enrolment, transition and DR-2 period consumption to be equivalent. This process is simple and, in essence, subtracts out differences in overall consumption and allows FSC to focus solely on changes in the load shape. It also assumes that none of the reductions in overall consumption are due to DR-2.

In the evaluation, FSC used a balanced approach that acknowledges the lack of a conclusive answer about the impact of DR-2 on electricity consumption. As a lower bound, FSC estimated load shifting assuming that changes in electricity consumption levels cannot be attributed to DR-2. As an upper bound, it was assumed that DR-2 led to the lower consumption levels during days when plants were in operation.

3.2 DR-3 Analysis Approach Overview

The DR-3 load reductions were estimated through regression methods using data from event and non-event-days and the available pre-enrolment data. Individual customer regressions were developed using data from 2010-2012 and aggregated to the settlement account and entire program level.

Based on the program dispatch pattern, the DR-3 program naturally produces an alternating or repeated treatment design. This means the analysis dataset includes a small number of treatment (event) days and a large number of control (non-event) days. Individual customer regressions use information from the control days to predict what would have happened on treatment days had the treatment not been in effect. Although events only occur for a few hours on each event-day, the entire event-day is evaluated to estimate both load reductions during event hours and load shifting to non-event hours.

The high concentration of load and participation in DR-3 was a significant factor in the selection of individual participant regressions. Changes in the number and mix of participants, as well as differences in the available data, made it difficult to analyse aggregated data. Moreover, individual participant regressions can better address variation in the size, production schedules and processes of participants.

A primary focus of the analysis was the actual load reductions relative to the contracted and scheduled load reductions – the realization rate. The contracted load reductions did not vary much for most participants. Comparing the event load reductions to scheduled load reductions allowed the regressions to assess the extent to which participants over or underperformed.

4 DR-2 Program *Ex Post* and *Ex Ante* Evaluation

DR-2 is a contractual load shifting program in which participants specify the load shift window and amount of load shifting. The program does not provide compensation for changes in the overall level of electricity consumption that may be associated with economic fluctuations or other factors. Load reduction during peak hours must be accompanied by shifting to the off peak of 7 PM to 7 AM in order to receive payment. While the program rules avoid having to determine if changes in electricity consumption are attributable to participation in the program, attribution is at the core of program evaluation.

This report contains the third full year analysis of the DR-2 program, which was officially implemented in November 2009. The participants visibly changed their load shape and eliminated volatility over the peak hours in conjunction with the implementation of the program.

The 2010 and 2011 analysis showed that customers do not return to normal, pre-DR-2 electricity use patterns during nonperformance days. In other words, DR-2 nonperformance is not equivalent to zero electricity output by a generator. It raises the question: Can the demand reductions, including changes to electricity consumption, be attributed to the program? Or is it better to assume that changes in electricity consumption levels are due to other factors besides DR-2?

It is important to note that customers cannot fully shift the loads given the production capability during off peak hours. If the participants follow the rules of the program, they are unable to fully make up the load that they shift to off peak hours.

The *ex ante* impact estimates use a balanced approach. As a lower bound, FSC used estimates of load shifting that assume any changes in overall electricity consumption levels cannot be attributed to DR-2. As an upper bound, FSC assumed the demand reductions and the lower consumption levels from DR-2 observed between 2009 and 2012 will be sustained in better economic times.

The rest of this section is organized as follows: the remainder of this section provides a brief summary of the key results; Section 4.1 provides a brief history of the DR-2 program and outlines the current contracts; Section 4.2 discusses the load reduction results; Section 4.3 discussed the differences between load reductions and the more conservative load shifting estimates; Section 4.4 presents the *ex ante* results for the 2012 evaluation; Section 4.5 offers conclusions and recommendations. Appendix A provides insight into the details and validation of the empirical model.

4.1 Program History and Customer Characteristics

DR-2 is a contractual load shifting program in which participants specify the load shift window and amount of load shifting. An example of load shift would be to reduce production below normal levels during the on peak period and to undertake that production during the off peak hours. The peak period for the program is extensive, lasting from 7 AM to 7 PM and providing participants a wide degree of flexibility for load shift hours. On the other hand, participants that contract DR for a sub-set of hours such as 12-4 PM cannot shift electricity to other hours in the peak period. Participants of the program can contract to reduce a predetermined amount of load for a minimum period of 4 consecutive hours up to a maximum of 12 consecutive hours. There are three options for participation: summer months, winter and summer months or all year.

DR-2 resources are not dispatchable since the program is not event-based. The program initially started in 2009 and was in a transition phase from March through October 2009. The transition allowed OPA and participants to determine whether the initial program rules were effective or required modification. It also allowed participants to test their load shifting capabilities and adjust the amount of load shifted and the peak participation hours to reflect their capabilities. However, reductions in payments due to partial or full non-compliance were not applied throughout the transition phase. The program rules were modified in October 2009 and the program officially launched in November 2009.

DR-2 participants notify OPA and the IESO of nonperformance days. They enable the IESO to better operate the system and schedule alternate resources to meet the higher demand for those days. Preannounced nonperformance leads to reductions in the participant payments. Unscheduled nonperformance leads to even larger payment reductions.

Participants receive compensation for both availability and utilization. The availability payments are designed to reflect the capacity value from the resource. The utilization payments reflect changes in the cost of electricity associated with shifting usage from higher-priced periods to lower-priced periods. Payments differ for summer and winter months.

For settlement purposes, compliance with the contracted load reductions is determined through a monthly baseline method. Failure to comply with the contract requirements of the DR-2 program can result in set-offs against potential revenue. The payment calculations are complex but can be reduced to a few basic components:

- Pre-DR-2 hourly consumption data is used to determine the customer load shape for each month;
- The customer must change its load shape – load reduction during peak periods must be accompanied by shifting to off peak periods;
- The load shift estimates are adjusted for changes in overall monthly consumption;
- Participants must provide consistent reductions during the selected peak hours during each day;
- Participants must reduce volatility over selected peak hours – spikes in demand above specific thresholds lead to payment setbacks; and
- Electricity use during peak hours that are not selected for shifting cannot exceed baseline levels – i.e., all shifting must occur to off peak hours.

As mentioned earlier, the program is relatively new and was in a transition phase for most of 2009. Importantly, all the direct participants transitioned from DR-1 to DR-2. From 2007 through February 2009, these customers experienced a number of events wherein customers perturbed their naturally occurring load patterns by reducing load. In 2007, 1,800 DR-1 event hours were called and 1,200 event hours were called in 2008.

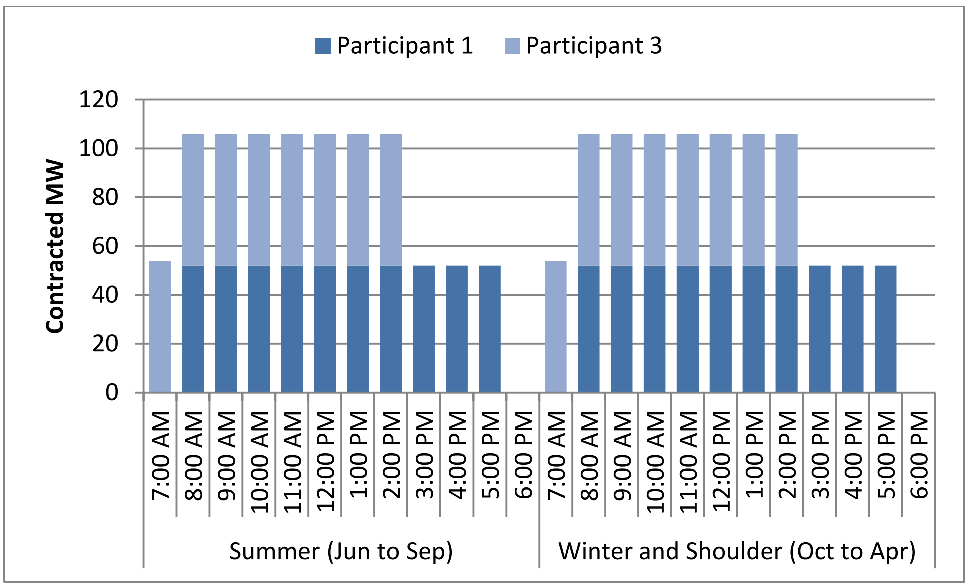
Initially, DR-2 had three contributors, but in 2011, one of them with 13 MW of contracted shifting closed their facility. This leaves just two participants with 52 and 54 MW of contracted shifting.

Table 4-1 summarizes the contracted load shift amount and window for 2012 by season. Importantly, the load shift contracted hours vary by participants and do not perfectly overlap. For example, the contracted load reductions from 7-8 AM, 8 AM to 3 PM and 3-6 PM are 54 MW, 106 MW and 52 MW, respectively. Figure 4-1 visually depicts the aggregate contracted load reductions by hour.

Table 4-1: Distribution of Peak Load and Contracted Load Shift by Participant as of December 2012

Period	Participant Name	Contracted MW	Contracted Hours
Summer	Participant 1	52	8 AM to 6 PM
	Participant 3	54	7 AM to 3 PM
	TOTAL	**106**	**Varies by hour**
Winter and Shoulder	Participant 1	52	8 AM to 6 PM
	Participant 3	54	7 AM to 3 PM
	TOTAL	**106**	**Varies by hour**

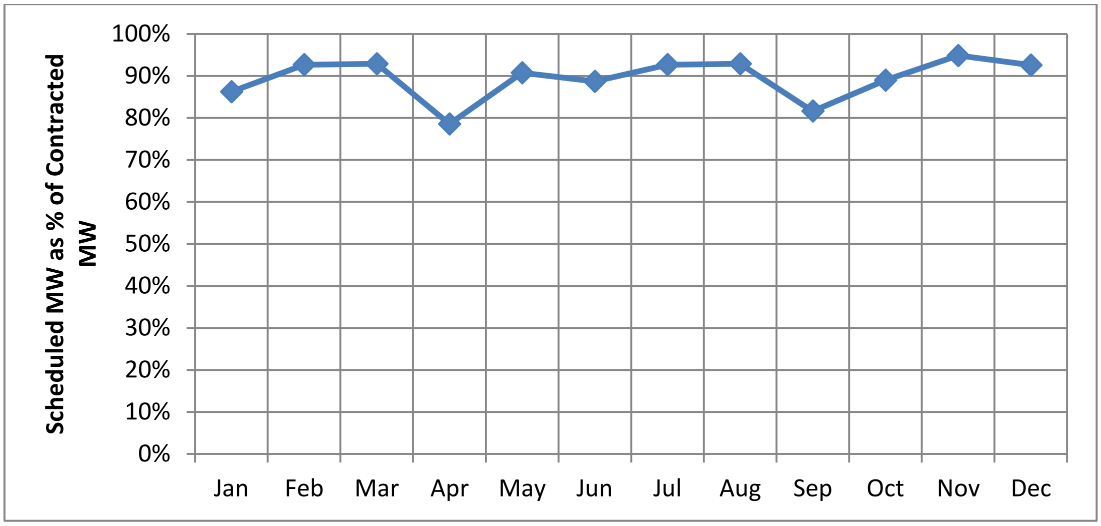
Figure 4-1: Hourly Distribution of 2012 Contracted Load Reduction (MW)



Most of the contracted summer demand reductions are in early morning hours when DR is less valuable rather than in the afternoon hours. Of the top 50 system load hours in each year between 2006 and 2012, 89% of them occurred in the summer months of June to September between 7 AM and 7 PM. However, 74% of those hours occurred between the 6-hour window of 12-6 PM. In other words, the peaking conditions that drive the need for additional capacity are more likely to occur in the afternoon hours. Higher incentives for the afternoon summer hours could induce participants to provide larger demand reductions for those hours and substantially increase the value of the program.

Figure 4-2 shows the scheduled performance as the percent of the contracted MW for each of the months in the contract periods. The contributors scheduled fewer nonperformance days in 2012 than in 2011.

Figure 4-2: DR-2 Average Scheduled Performance by Month in 2012

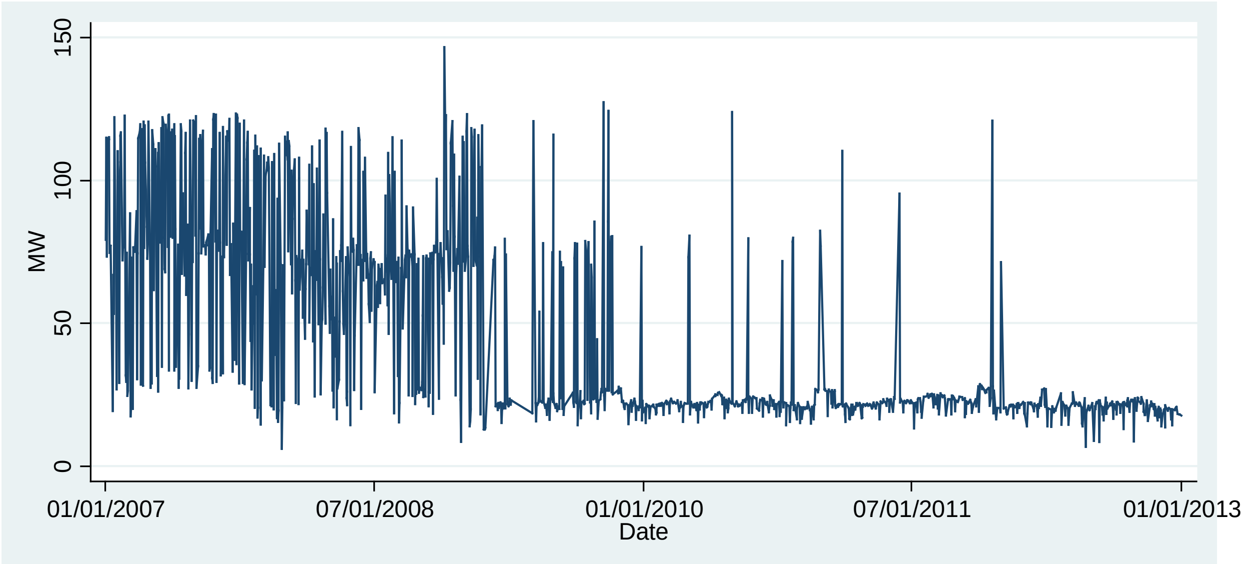


Historically, nonperformance by DR-2 has meant less demand on the electric grid – a positive development for a program designed to reduce demand during peak hours. In other words, DR-2 nonperformance is not analogous to generator nonperformance. When a generator is unavailable, it does not introduce any electricity supply into the system. When a DR-2 participant is unavailable, they typically have even lower demand than during days when they actively engage in load shifting. Nonperformance days correlate highly to facility shut downs. During DR-2 transition and enrolment periods, participants shut down the main processes at their facilities in almost all scheduled nonperformance days. Almost no shutdowns occurred outside of scheduled nonperformance.

The two remaining participants are unable to match their 2006 to 2008 electricity consumption levels even if they operated each weekday using their current DR-2 electricity use patterns. Put differently, even without any nonperformance days, they are unable to replicate pre-DR-2 consumption levels while following the DR-2 schedule. Contributors are unable to fully shift their loads to off peak hours due to their production capacity and this leads to reduced consumption.

There is strong evidence that the DR-2 Program is reducing load volatility during the contract hours. Figure 4-3 illustrates the load for Participant 3 for 11 AM to 12 PM from January 2007 to December 2012. The figure shows that load spikes during this hour are relatively common before the beginning of the DR-2 program, with peak usage often above 100 MW. After the commencement of DR-2, these loads rarely spike above 30 MW during this hour.

Figure 4-3: DR-2 Participant 3 Load for 11 AM to 12 PM from January 2007 to December 2012

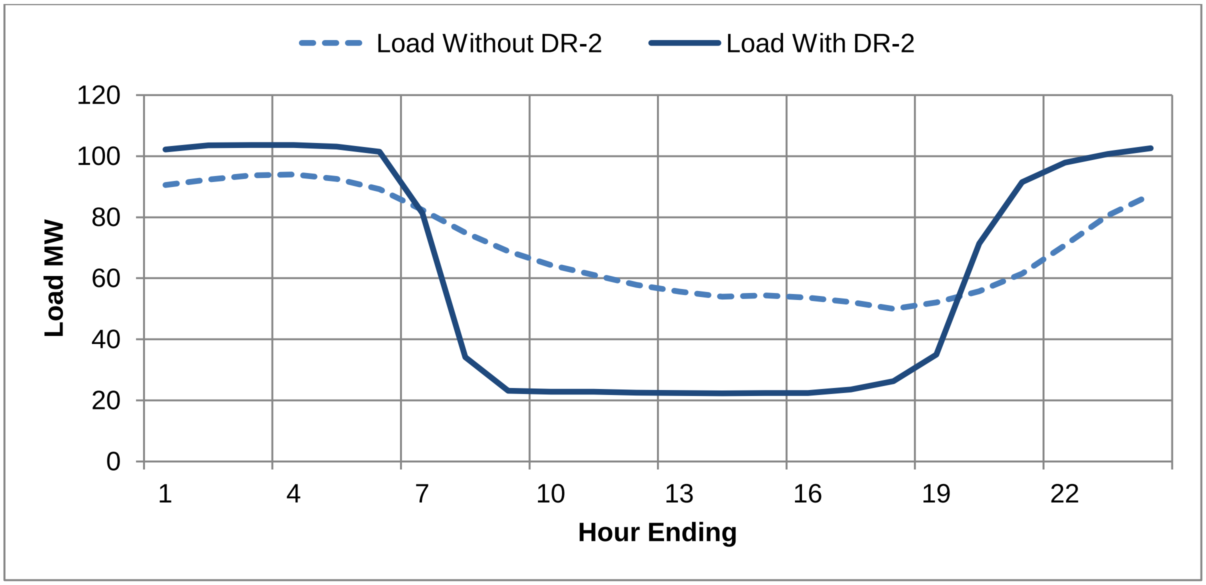


4.2 2012 *Ex Post* Load Reduction Results

In 2012, participants delivered demand reductions both for hours in which they were and were not contracted. In other words, they reduced demand for periods in which they were not paid. This was mainly in order to comply with the rule that load could only be shifted to hours outside of the 7 AM to 7 PM window. Basically, the participants could not replicate the load patterns in the baseline months because their load patterns are step-like and driven by a discrete number of processes. The baselines were calculated by averaging days with a mix of active processes. Because of their step-like load patterns, some participants cannot replicate the load patterns in the baseline and have to choose between exceeding it or providing reductions during hours they have not contracted.

Figure 4-4 illustrates this pattern with the participant contracted to shift load from 8 AM to 6 PM. In practice, they altered their electricity across the entire peak period from 7 AM to 7 PM, providing two extra hours of demand reductions. This customer is clearly reducing demand, but also does not fully shift all of the demand reduction to off peak hours. If they did, the increases in electricity consumption (MWh) during off peak hours would exactly match the decreases during contracted hours. Likewise, the participant contracted to provide reductions between the hours of 7 AM to 3 PM does not return to their pre DR-2 load until 6PM.

Figure 4-4: Comparing Participant 1 Load Before and After DR-2



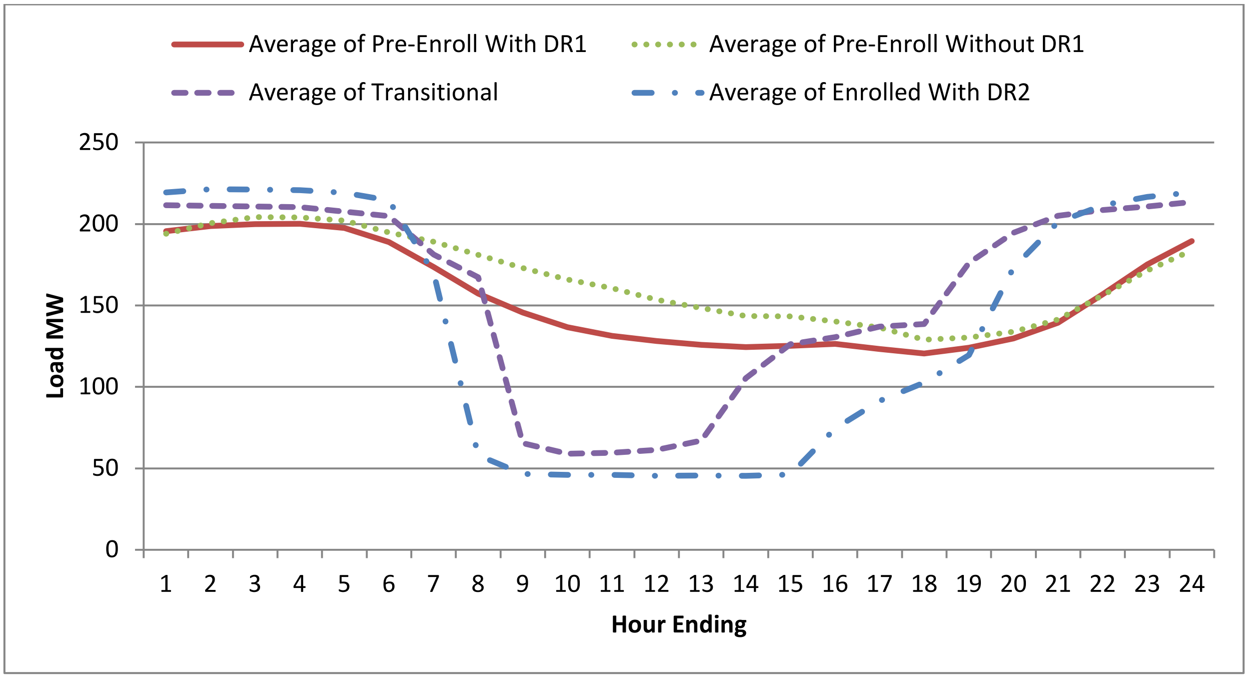
4.3 Load Shifting vs. Load Impacts

A key challenge for DR-2 is to distinguish between changes in consumption levels unrelated to the program and changes in hourly usage patterns. To avoid compensating participants from changes in overall output, the DR-2 rules make it clear that reductions in usage do not constitute load shifting, they must be accompanied by increases in use during off peak hours. In contrast, the regression models can control for changes in economic conditions and estimate the change in electricity use patterns, including reductions attributable to the program. The primary risk of doing so is confounding the program reductions with economic conditions if they are highly correlated.

Overall, electricity consumption levels of DR-2 participants changed relative to their pre-enrolment level consumptions, which are used to calculate settlement baselines. Both the average daily usage, as well as the number of days in which operations were shutdown, vary significantly between pre and post DR-2 enrolment. For almost all days when operations shut down, participants scheduled it as a nonperformance day with the IESO. Once those days are factored out, average daily consumption level for the pre and post DR-2 period are similar.

Figure 4-5 shows the average weekday hourly load for DR-2 participants excluding nonperformance days and days when the facility was shut down. The graph excludes scheduled nonperformance and days when the facility shut down during the pre-enrolment period. It compares the time span prior to any participation in DR programs, the period when they participated in DR-1 (with DR-1 reductions added back in), the DR-2 transition period and the DR-2 contractual period. It highlights several key issues. First, although the graph shows unfiltered results, it is clear that customers reduced load during peak hours and engaged in load shifting during the transition and DR-2 contractual periods. This was also observed at the individual participant level. Second, customers have a limited amount of load to shift and some of their response is in fact reductions in consumption. However, DR-2 only provides compensation for shifting. Given the industrial processes in place, the maximum aggregate load for DR-2 participants is approximately 250 MW.

Figure 4-5: Average Hourly Load for Aggregate DR-2 Participants by Pre-enrolment and Participation Period (Excludes Nonperformance Days)



Importantly, the meter data presented above does not control for differences in electricity market prices, economy (aside from shutdowns) and other factors that vary across different time periods. The evaluation impact estimates do account for those factors, to the extent possible. To estimate load reductions, historical pre-enrolment load patterns from 2006-2008, which include the DR-1 reductions added back in, were employed to predict the counterfactual – that is, the expected load levels for DR-2 participation months. Based on the regression results, the load shifting to off peak periods does not always match load reduction during the peak period.

As a conservative cross check to the regression based load reduction estimates, FSC estimated the amount of shifting without the influence of changes in consumption. To do this, the load shapes were normalized by dividing the hourly load for each participant by the average hourly load for all qualifying days in the month. This process essentially subtracts out differences in overall consumption – the average value for all shapes is equal to one – and allowed FSC to focus solely on changes in the load shape. The load shapes can be directly compared in their normalized form or scaled for the actual consumption in the month. In this instance, the reference or counterfactual load shape was scaled to reflect the actual consumption observed during each DR-2 month. By construction, the load decreases over the peak period window exactly match the increases in load outside of the peak window. Figure 4-6 compares the normalized loads with and without DR and reflect the load shifting under the strict rule that load reductions, even if attributable to the program, do not count as load shifting.

Figure 4-7 shows the hourly load shifting for the transition and contract period. The regression predicted counterfactual and load with DR have been rescaled so demand shifted from peak hours exactly matches the demand shifted to off peak hours.

Figure 4-6: Hourly Load Shifting For DR-2 Participants (Rescaled Counter-factual) For Contract Period, Excluding Planned Nonperformance Days

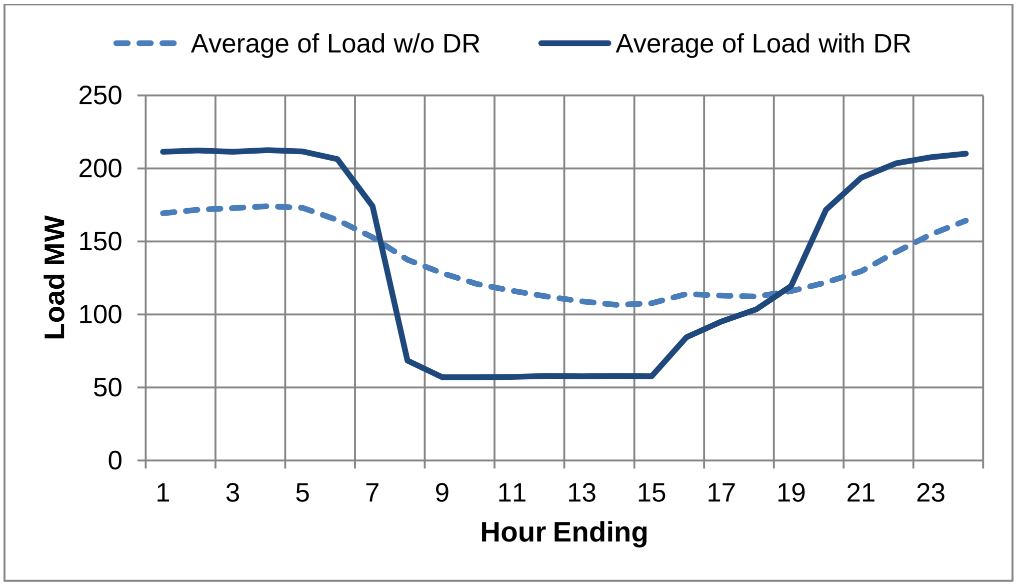


Table 4-2 compares the total contracted and day-ahead scheduled resources to the actual load reductions and shift attribution for the contract period. The load reductions across the 7 AM to 7 PM period were, on average, 91.9% and 81.2% of the load reduction contracted for the summer and winter months respectively. In contrast, the shift attribution was lower – 56.1% and 55.4% – for the summer and winter months respectively.

Table 4-2: 2012 Hourly Comparison of Load Reduction to Shift Attribution

Season	Hour Start	Hour End	Contracted MW	Scheduled MW	Load Reduction MW	% Contracted MW	Load Shift MW	% of Contracted MW
Summer	7:00 AM	8:00 AM	54	44.5	111.8	207.0%	74.7	138.3%
	8:00 AM	9:00 AM	106	95.2	114.6	108.2%	79.6	75.1%
	9:00 AM	10:00 AM	106	95.2	103.2	97.3%	70.5	66.5%
	10:00 AM	11:00 AM	106	95.2	94.1	88.8%	63.6	60.0%
	11:00 AM	12:00 PM	106	95.2	83.2	78.5%	55.2	52.0%
	12:00 PM	1:00 PM	106	95.2	72.0	68.0%	46.7	44.1%
	1:00 PM	2:00 PM	106	95.2	67.8	63.9%	43.1	40.6%
	2:00 PM	3:00 PM	106	95.2	66.4	62.7%	42.3	39.9%
	3:00 PM	4:00 PM	52	50.8	51.1	98.3%	26.8	51.5%
	4:00 PM	5:00 PM	52	50.8	43.4	83.5%	19.4	37.3%
	5:00 PM	6:00 PM	52	50.8	43.2	83.1%	18.0	34.5%
	6:00 PM	7:00 PM	0	0.0	38.2	.	10.8	.
	Contracted hours			**79.3**	**72.0**	**74.1**	**94.5%**	**45.9**
	7 AM to 7 PM			**86.5**	**78.5**	**77.4**	**94.5%**	**49.1**
Winter	7:00 AM	8:00 AM	54	38.5	84.8	156.9%	63.9	118.4%
	8:00 AM	9:00 AM	106	90.5	77.8	73.4%	58.7	55.3%
	9:00 AM	10:00 AM	106	90.5	72.6	68.5%	54.1	51.0%
	10:00 AM	11:00 AM	106	90.5	67.7	63.9%	49.7	46.9%
	11:00 AM	12:00 PM	106	90.5	66.0	62.3%	48.4	45.7%
	12:00 PM	1:00 PM	106	90.5	65.9	62.2%	48.4	45.6%
	1:00 PM	2:00 PM	106	90.5	64.5	60.8%	47.2	44.6%
	2:00 PM	3:00 PM	106	90.5	66.4	62.7%	49.0	46.3%
	3:00 PM	4:00 PM	52	52.0	51.3	98.6%	33.8	65.0%
	4:00 PM	5:00 PM	52	52.0	37.8	72.6%	21.2	40.7%
	5:00 PM	6:00 PM	52	52.0	24.0	46.2%	8.5	16.3%
	6:00 PM	7:00 PM	0	0.0	14.4	.	-1.6	.
	Contracted hours			**79.3**	**69.0**	**57.8**	**75.3%**	**40.1**
	7 AM to 7 PM			**86.5**	**75.3**	**61.7**	**75.3%**	**43.9**
Shoulder	7:00 AM	8:00 AM	54	34.5	68.7	127.2%	52.4	97.0%
	8:00 AM	9:00 AM	106	81.4	79.3	74.9%	66.8	63.0%
	9:00 AM	10:00 AM	106	81.4	71.2	67.2%	58.9	55.5%
	10:00 AM	11:00 AM	106	81.4	66.2	62.4%	54.1	51.1%
	11:00 AM	12:00 PM	106	81.4	62.9	59.3%	50.7	47.9%
	12:00 PM	1:00 PM	106	81.4	60.5	57.1%	48.3	45.6%
	1:00 PM	2:00 PM	106	81.4	59.4	56.0%	47.1	44.4%
	2:00 PM	3:00 PM	106	81.4	61.1	57.7%	48.7	45.9%
	3:00 PM	4:00 PM	52	46.9	42.6	82.0%	28.8	55.3%
	4:00 PM	5:00 PM	52	46.9	32.5	62.4%	18.3	35.3%
	5:00 PM	6:00 PM	52	46.9	24.8	47.7%	10.2	19.7%
	6:00 PM	7:00 PM	0	0.0	9.6	.	-8.1	.
	Contracted hours			**79.3**	**62.1**	**53.2**	**68.5%**	**39.7**
	7 AM to 7 PM			**86.5**	**67.7**	**57.2**	**68.5%**	**44.0**

The scheduled demand reduction and its delivery vary across the hours of the day. Both are generally higher in the morning hours and lower in the afternoon hours. For example, from 7-8 AM, participants provided 180% of the scheduled resources. In fact, both participants reduced demand during this hour even though they were not contracted to do so. Conversely, during the 2-6 PM summer critical hours, when system peaks are more likely to occur, participants only delivered 77.3% of the scheduled load reduction. Figure 4-7 illustrates the impacts and shifts relative to the scheduled reductions by hour for the 2012 summer months. While flexibility is one of the key features of the program, it can produce more value by increasing the incentives for the more critical hours (and decreasing them for less critical hours) and/or narrowing the eligible hours.

Figure 4-7: 2012 Hourly Comparison of Load Reduction to Shift Attribution From June 2012 to September 2012

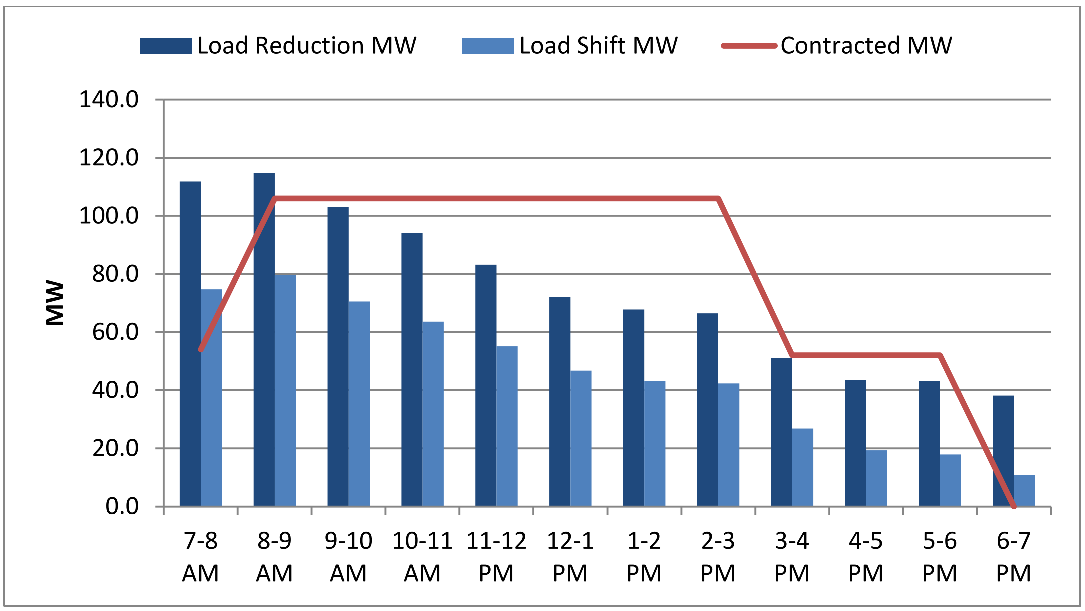


Table 4-3 and Table 4-4 summarize the *ex post* load reductions and the shift attribution, by hour, for the monthly system peak days. They highlight several issues. As mentioned earlier, DR-2 program load reductions are higher than the shift attribution. The load reductions reflect the reductions attributable to the program based on the statistical model. The shift attribution provides a more conservative estimate without the potential of confounding program reductions with economic conditions.

Table 4-3: 2012 Hourly DR-2 *Ex Post* Load Reductions by Month For Monthly System Peak Days

Day Type	Month	Hour Ending										
		8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM
Monthly System Peak Day	Jan	109	97	92	86	84	85	84	87	69	48	29
	Feb	96	84	77	71	69	68	67	69	48	31	15
	Mar	105	95	87	83	80	78	75	77	59	49	37
	Apr	90	98	90	86	84	82	79	81	58	52	44
	May	89	99	89	84	78	72	69	70	43	37	35
	Jun	103	110	100	93	85	78	72	72	55	51	51
	Jul	98	114	103	96	85	76	70	71	58	54	50
	Aug	118	111	103	96	87	77	71	70	52	43	43
	Sep	115	108	98	91	84	76	72	72	58	45	42
	Oct	64	76	65	60	55	50	48	50	32	19	7
	Nov	79	84	77	72	67	66	65	68	47	27	13
	Dec	81	84	78	72	70	70	68	71	54	37	18
	2012	**95**	**97**	**88**	**83**	**77**	**73**	**70**	**72**	**53**	**41**	**32**

Table 4-4: 2011 Hourly DR-2 *Ex Post* Shift Attribution by Month For Monthly System Peak Days

Day Type	Month	Hour Ending										
		8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM
Monthly System Peak	Jan	74	65	62	56	55	56	55	58	41	24	8
	Feb	79	68	62	57	55	55	54	56	35	18	2
	Mar	78	70	63	60	57	55	53	55	35	25	13
	Apr	60	72	65	62	59	59	56	58	34	27	20
	May	67	80	71	67	60	55	52	54	24	18	16
	Jun	57	66	59	54	48	43	39	39	23	19	19
	Jul	59	79	71	65	56	50	46	47	33	30	27
	Aug	86	81	73	67	60	52	46	45	26	18	16
	Sep	85	79	71	65	59	52	48	49	33	19	15
	Oct	59	75	63	58	52	46	44	45	26	12	-1
	Nov	64	73	67	61	57	56	55	58	35	16	1
	Dec	60	65	60	54	52	53	51	54	37	21	3
	2011	69	73	65	60	56	53	50	51	32	20	11

4.4 Load Reductions for Planning – *Ex Ante* Results

Ex ante load reductions are most appropriate for demand response (DR) programs since they reflect the insurance or option value of demand reductions. They are the demand reductions that can be expected under extreme conditions and are used for system planning and can be used to assess program cost effectiveness (typically weather or electricity price related). DR programs are designed to help address electricity system capacity issues when the system is under the most stress.

The *ex ante* load reductions are based on the amount of contractual load response and the degree to which participants have historically delivered those load reductions. The load reductions are used as an upper bound for *ex ante* impact and shift attribution is used as a lower bound, conservative estimate. The shift estimates assume that any changes in consumption are not due to the program and effectively subtract out changes in electricity consumption. The DR-2 participants are not weather sensitive and have a contractual amount of load shifting for a specific period. As a result, the *ex ante* load reductions factor in the amount of contractual demand reduction and the degree to which these customers have historically delivered scheduled demand reductions. In practice, derating the demand reduction for scheduled nonperformance is a conservative approach given that these customers shut down facilities on those days and decrease demand on the electric system.

The *ex ante* forecasts incorporate all of the data during the contract period: from November 2009 to December 2012 for the active participants as of January 2013. Participant 2 is dropped from these estimates because they are no longer active. Table 4-5 illustrates the load reduction estimates between 2006 and 2012. It also factors in an additional downward adjustment compared to historical 2012 load reductions to account for the lack of information about performance under better economic conditions. On average, during summer months OPA can expect 87% and 71% of the contracted load reduction to be scheduled and delivered. During winter months, OPA can expect 66% and 51% of the contracted load reduction to be scheduled and delivered. The corresponding values for shoulder months are 87% and 57%.

Table 4-5: 2012 Hourly DR-2 *Ex Ante* Load Reductions by Season

Season	Hour Start	Hour End	Contracted MW	Scheduled MW	*Ex Ante* Load Reduction Estimate	90% Confidence Interval	
						Lower Bound	Upper Bound
Summer Months	7:00 AM	8:00 AM	54	45.0	84.8	65.2	104.7
	8:00 AM	9:00 AM	106	93.9	87.4	68.9	106.4
	9:00 AM	10:00 AM	106	93.9	79.0	61.6	96.8
	10:00 AM	11:00 AM	106	93.9	72.9	56.3	89.8
	11:00 AM	12:00 PM	106	93.9	65.3	49.8	81.2
	12:00 PM	1:00 PM	106	93.9	58.6	44.2	73.4
	1:00 PM	2:00 PM	106	93.9	53.9	40.0	68.1
	2:00 PM	3:00 PM	106	93.9	54.2	40.3	68.4
	3:00 PM	4:00 PM	52	43.7	38.8	24.6	53.2
	4:00 PM	5:00 PM	52	43.7	31.8	17.5	46.3
	5:00 PM	6:00 PM	52	43.7	30.1	15.3	45.0
	6:00 PM	7:00 PM	0	0.0	23.3	7.8	39.0
	AVERAGE		**79.3**	**69.4**	**56.7**	**41.0**	**72.7**
Winter Months	7:00 AM	8:00 AM	54	46.7	74.6	62.3	87.5
	8:00 AM	9:00 AM	106	93.9	68.4	57.2	80.2
	9:00 AM	10:00 AM	106	93.9	63.0	52.4	74.2
	10:00 AM	11:00 AM	106	93.9	58.6	48.4	69.4
	11:00 AM	12:00 PM	106	93.9	57.0	47.0	67.5
	12:00 PM	1:00 PM	106	93.9	56.9	47.1	67.3
	1:00 PM	2:00 PM	106	93.9	55.8	46.1	65.9
	2:00 PM	3:00 PM	106	93.9	57.7	48.0	68.0
	3:00 PM	4:00 PM	52	43.7	39.4	29.8	49.7
	4:00 PM	5:00 PM	52	43.7	25.3	16.0	35.2
	5:00 PM	6:00 PM	52	43.7	10.7	1.9	20.2
	6:00 PM	7:00 PM	0	0.0	-1.0	-10.2	8.7
	AVERAGE		**79.3**	**69.6**	**47.2**	**37.2**	**57.8**
Shoulder Months	7:00 AM	8:00 AM	54	43.0	65.4	54.8	76.5
	8:00 AM	9:00 AM	106	93.9	73.1	63.3	83.5
	9:00 AM	10:00 AM	106	93.9	65.0	55.8	74.8
	10:00 AM	11:00 AM	106	93.9	60.8	51.9	70.3
	11:00 AM	12:00 PM	106	93.9	56.4	47.9	65.5
	12:00 PM	1:00 PM	106	93.9	53.5	45.2	62.4
	1:00 PM	2:00 PM	106	93.9	51.3	43.2	59.9
	2:00 PM	3:00 PM	106	93.9	53.0	44.7	61.7
	3:00 PM	4:00 PM	52	43.7	32.2	23.9	41.0
	4:00 PM	5:00 PM	52	43.7	21.9	13.4	31.0
	5:00 PM	6:00 PM	52	43.7	13.9	5.0	23.4
	6:00 PM	7:00 PM	0	0.0	-0.2	-9.6	9.8
	AVERAGE		**79.3**	**69.3**	**45.5**	**36.6**	**55.0**

4.5 Conclusions and Recommendations

The DR-2 program led to distinct changes in customer loads and reduced the volatility of use during peak hours. The effect of DR-2 on customers' load is evident without complicated analysis and can be observed by simply comparing load patterns and volatility before and after implementation of DR-2. However, reductions during the peak window (7 AM to 7 PM) were not fully matched by shifting to off peak hours (7 PM to 7 AM). Overall, electricity consumption for DR-2 participants is approximately 30% lower than their consumption prior to enrolling in DR-2. Given their production capacity, the participants are unable to fully shift reduction during peak periods to off peak periods and fully comply with DR-2 program rules. In addition, a significant portion of reductions occurred outside of customers' contracted windows. Simply put, customers did not match the scheduled reductions well. This is in part due to the process driven, step-like load patterns of participants.

Ex ante impact estimates use a balanced approach to account for the reduced energy consumption. As a lower bound, FSC used estimates of load shifting that assumed any changes in overall electricity consumption levels cannot be attributed to DR-2. As an upper bound, FSC assumed the demand reductions and the lower consumption levels from DR-2 observed between 2009 and 2012 will be sustained.

FSC does not recommend continuing the DR-2 contracts with existing participants beyond the current contract period. There are two main reasons for this recommendation. The estimate of the load shifting and energy saving relies on constructing baselines using data from 2009 and earlier. The baselines, and by connection the shifting and savings estimate, becomes less reliable as time elapses. The second reason is because discontinuing the program allows OPA to determine whether or not the change in load shapes persist without continued payments. Since the program led to fundamental changes in production processes and schedule of participants, they may continue consuming less power during on peak hours and shifting it to off peak hours.

OPA should consider whether to continue a load shifting program in the future but with a revised design and different target contributors. Load shifting can provide significant benefits, particularly given the current state of Ontario's electricity system, which currently has surplus generation during off peak hours on shoulder months. A revised load shifting program should allow aggregation across customers and, ideally, focus on Class B customers, which are not exposed to wholesale market prices. A revised load shifting program also should strongly consider narrowing the load shifting period to hours with the most values or providing higher incentives for the most critical hours.

5 DR-3 Program *Ex Post* and *Ex Ante* Evaluation

5.1 Program Background

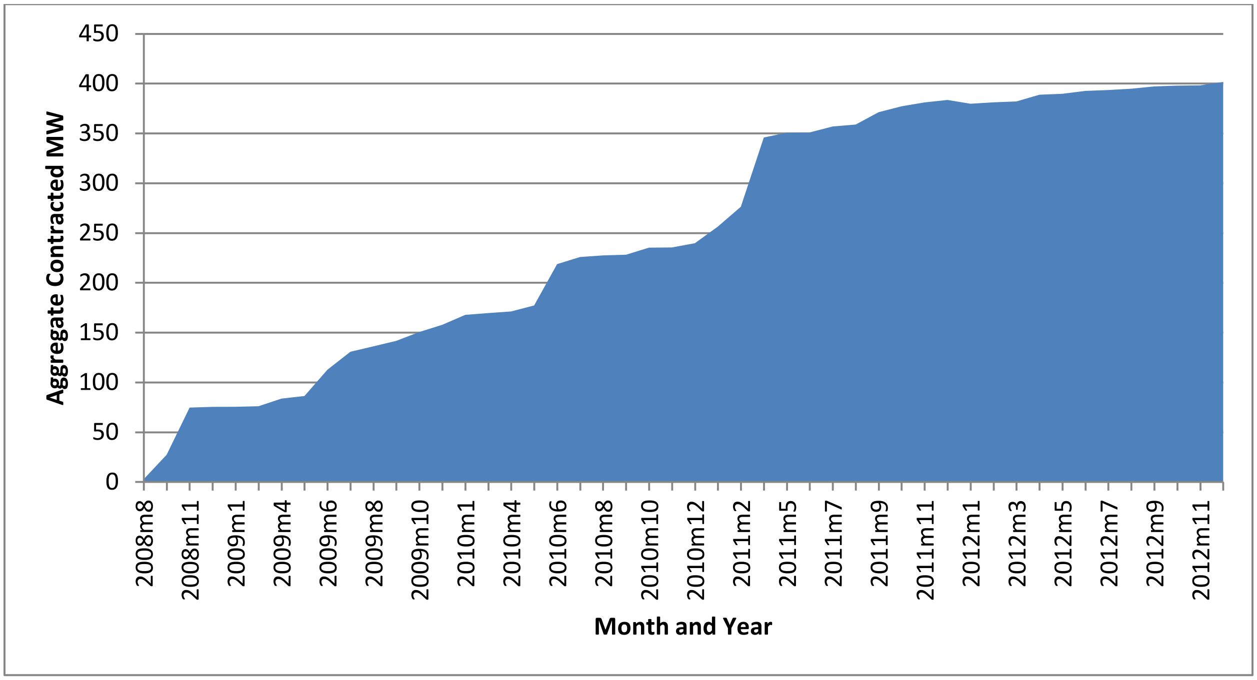
DR-3 allows participants and aggregators to enter into contractual agreements for load reductions with OPA. Participants can choose to enroll directly with OPA, provided they meet minimum load reduction criteria, or can participate through an aggregator. Under DR-3, an aggregator or direct participant must commit to a specific load reduction amount for either the 100- or 200-hour option per year. In 2010, OPA began to offer an additional option for 100-hour customers, which restricts activations to the 12-6 PM time period (hereafter referred to as the "12-6 Group"). OPA has some discretion regarding the timing of events and currently is determining event-days based on IESO day-ahead supply cushion estimates. In exchange for load reductions, the DR-3 program makes both availability (capacity) and energy payments. Unlike OPA's demand buy-back program (DR-1), OPA can reduce payments if participants fail to provide the contracted load reduction or are unavailable to provide load reductions (the equivalent of a scheduled outage). For settlement purposes, compliance with the contracted load reductions is determined through day-matching baseline methods.

DR-3 participants notify OPA and the IESO of any short-term fluctuations in load reduction capability due to facility maintenance or down time. These days are classified as nonperformance days and are analogous to scheduled generator outages. They enable the IESO to better operate the system and schedule alternate resources for those days. Nonperformance days lead to reductions in the participant payments. The payment reductions are higher if an event is called during a participant nonperformance day. Unscheduled nonperformance – failure to meet contractual obligations during events – leads to even larger payment reductions.

5.2 Program Participation

DR-3 added 89 new contributors in 2012, accounting for 27 MW of summer contracted load reductions. By the end of 2012, there were 526 total contributors in the DR-3 program with an aggregate summer contracted load reduction of 408 MW. Since program launch, the DR-3 participant mix and load reduction capabilities have evolved substantially. The largest change in load reduction capability occurred with the enrolment of three large direct participants in September and October of 2008. In 2010, a nearly 40 MW increase in load reduction capability occurred in June with the addition of 44 contributors by aggregators. This increase included a single large new participant with a contracted load reduction of 25 MW. In 2011, a large portion of the new customers provided less than 1 MW of load reduction. However, three customers contributed over 10 MW each with the two largest adding 28 and 26 MW, respectively. In 2012, similar to 2011, most new contributors provided less than 1 MW of load reduction. The largest new contributor has a summer contracted MW of 6.5 MW.

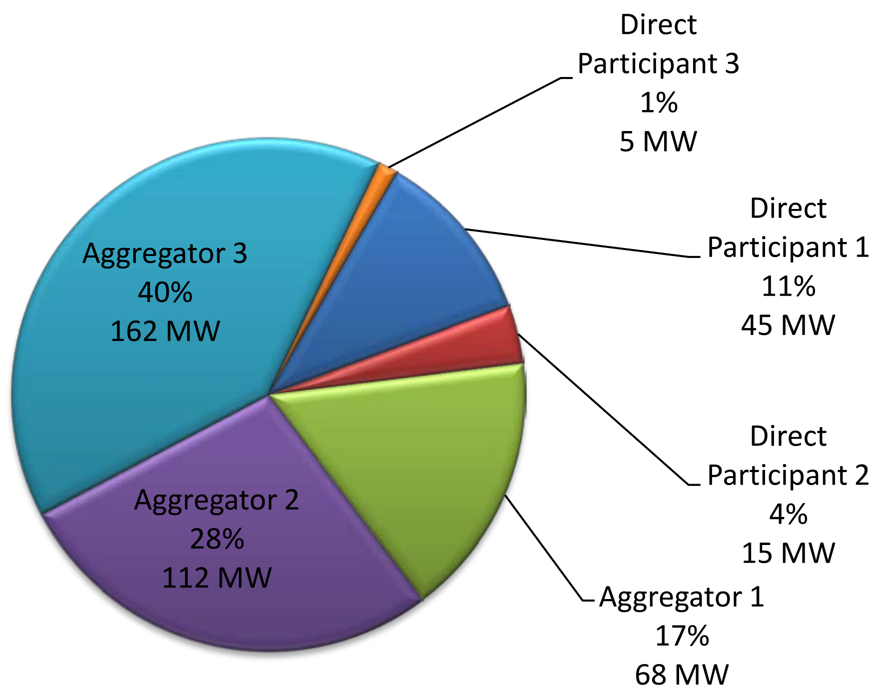
Figure 5-1: Aggregate Contracted Load Reduction by Month and Year



Despite substantial growth in the number of contributors in 2012, load reductions remain concentrated. As of December 31, 2012, the top 20 contributors accounted for 61% of the summer contracted load reduction whereas 89% of the contributors accounted for less than 1 MW of summer contracted load reduction.

Figure 5-2 shows the distribution of summer contracted load reduction among aggregators and direct participants as of December 2012. The percent of historical load that each aggregator and direct participant reduces varies substantially. For example, Direct Participant 1, is contracted to reduce 45 MW or 100% of its historical load. In contrast, Aggregator 3, is contracted to deliver 162 MW of load reduction though it averages 446 MW of historical load – a 36% reduction. The percent impacts have two main implications. As discussed earlier, error in settlement baselines are magnified with smaller percent load reductions. In addition, it is more difficult to accurately distinguish smaller percent load reductions from normal variation electric loads (background noise).

Figure 5-2: Distribution of Contracted Load Reduction by Aggregator/Direct Participant As of December 2012



5.3 2012 Events and Event Conditions

Table 5-1 shows the events that were called in 2012. There were 5 events called in 2012 compared to 11 in 2011, 10 in 2010, 6 in 2009 and 14 in 2008. The first event for 2012 was called on June 20 and the last one on September 6. As noted above, participants must choose between the 100-hour (normal or 12-6) and 200-hour options, which represent the maximum number of event hours per year that can be called. Over the estimation period, the 200-hour commitment group was called for all 5 events, the 100-hour group was also called for all 5 events and the 12-6 group was called for 3 events.

Table 5-1: 2012 Event-day Information

Event Date	Event Start [1]	Event End [1]	100-hour Group	200-hour Group	100-hour, 12-6 Group	Total Summer Contracted MW	Scheduled MW
6/20/2012	2:00 PM	6:00 PM	Activated	Activated	Activated	378	373
6/21/2012	2:00 PM	6:00 PM	Activated	Activated	Activated	364	357
7/17/2012	2:00 PM	7:00 PM	Activated	Activated	Activated	382	375
9/5/2012	3:00 PM	7:00 PM	Activated	Activated	–	366	355
9/6/2012	3:00 PM	7:00 PM	Activated	Activated	–	366	358

[1] Program start and end times varied by DR-3 option. The values reflect the earliest and latest DR-3 activation.

5.4 2012 *Ex Post* Load Reduction Results

For the 5 events called in 2012, 84% of contracted load and 86% of scheduled load reductions were delivered. Contributors scheduled 98% of the contracted reductions for event days. At its peak on June 21, the DR-3 Program delivered 357 MW of load reduction, representing 98% of contracted MW or 100% of scheduled MW. Because of the changing mix of participants, the amount of contracted and delivered load reduction varied depending on the event-day. Table 5-2 shows the average hourly load reductions for each of the event-days called in 2012.[6]

Table 5-2: DR-3 Program 2012 Load Reductions by Event-day

Event Date	Group	Summer Contracted MW	Scheduled MW	Estimated Load Reduction (MW)	% of Summer Contracted MW	% of Scheduled MW
6/20/2012	100/200/12-6	378	373	356	94%	96%
6/21/2012	100/200/12-6	364	357	357	98%	100%
7/17/2012	100/200/12-6	382	375	330	86%	88%
9/5/2012	100/200	366	355	262	71%	74%
9/6/2012	100/200	366	358	262	71%	73%
Average Event-day		371	364	313	84%	86%

Some of the variation in performance is explained by the fact that aggregators manage loads to meet contractual obligations according to settlement baselines. Inaccuracies in the settlement baselines affect the extent to which the scheduled load reduction is delivered.

Table 5-3 summarizes the 2012 average *ex post* load reductions for both aggregators and direct participants. The table first shows the three direct participants, then the three aggregators and the average for both groups at the bottom. Direct participants ranged from 61% to 98% of summer contracted MW and of scheduled MW. 2012 had less planned nonperformance than 2011 so there is little difference between percent scheduled and percent contracted. Aggregators ranged from 64% to 100% in terms of contracted MW delivered. On average, aggregator participants supply 85% of contracted MW compared to direct participants, who supplied 81% of contracted MW.

6 All subsequent DR-3 analysis will use the subset of customers who had complete data for the months of June 2012 through September 2012. FSC used 312 contributors out of a total of 513, who accounted for 93% of the total summer contracted MW present by June 2012.

Table 5-3: Summary of 2012 Load Reductions by Type of Participant

Participants	Type	Summer Contracted MW	Scheduled MW	Impact (MW)	% of Summer Contracted MW	% of Scheduled MW
1	Direct Participant	45	45	38.2	85%	85%
2	Direct Participant	15	15	14.6	98%	98%
3	Direct Participant	5	5	3.1	61%	61%
1	Aggregator	65.5	65.5	65.5	100%	100%
2	Aggregator	101.7	94.9	65.2	64%	69%
3	Aggregator	138.9	138.2	126.7	91%	92%
Aggregator Participant		**306.1**	**298.6**	**257.5**	**85%**	**87%**
Direct Participant		**65.0**	**65.0**	**55.8**	**81%**	**81%**

Table 5-4 shows the 2012 *ex post* load reductions for the average event by IESO region. DR-3 participants are spread throughout OPA's territory. The South Central and Toronto regions have the most settlement accounts and are among the top three biggest in terms of summer contracted MW. The Northeast region has the highest average summer contracted MW per customer at 24 contracted MW per customer. IESO regions also differ in the percentage of scheduled load reductions that they deliver. The Northwest region delivered only 37% of contracted MW while the Niagara region delivered 111% of the contracted MW.

Table 5-4: Summary of 2012 Load Reductions by IESO Region

IESO Region	# of Settlement Accounts	Summer Contracted MW	Estimated Load Reduction (MW)	% of Summer Contracted MW
East	7	26.7	24.8	93%
Essa	4	12.2	8.7	71%
Georgian Bay	1	2.8	1.4	50%
Long Point	4	5.0	4.8	96%
Niagara	6	14.2	15.8	111%
Northeast	5	118.6	124.6	105%
Northwest	1	0.1	0.0	37%
Ottawa	4	27.7	19.1	69%
South Central (DR)	12	79.5	73.2	92%
Toronto	16	113.7	97.3	86%
West	11	30.8	23.3	76%

Finally, impacts are broken down by industry. Table 5-5 shows the *ex post* results broken down by industry. The industry with the greatest reference MW is the "Industrial Tools/Metal Work/Electronics

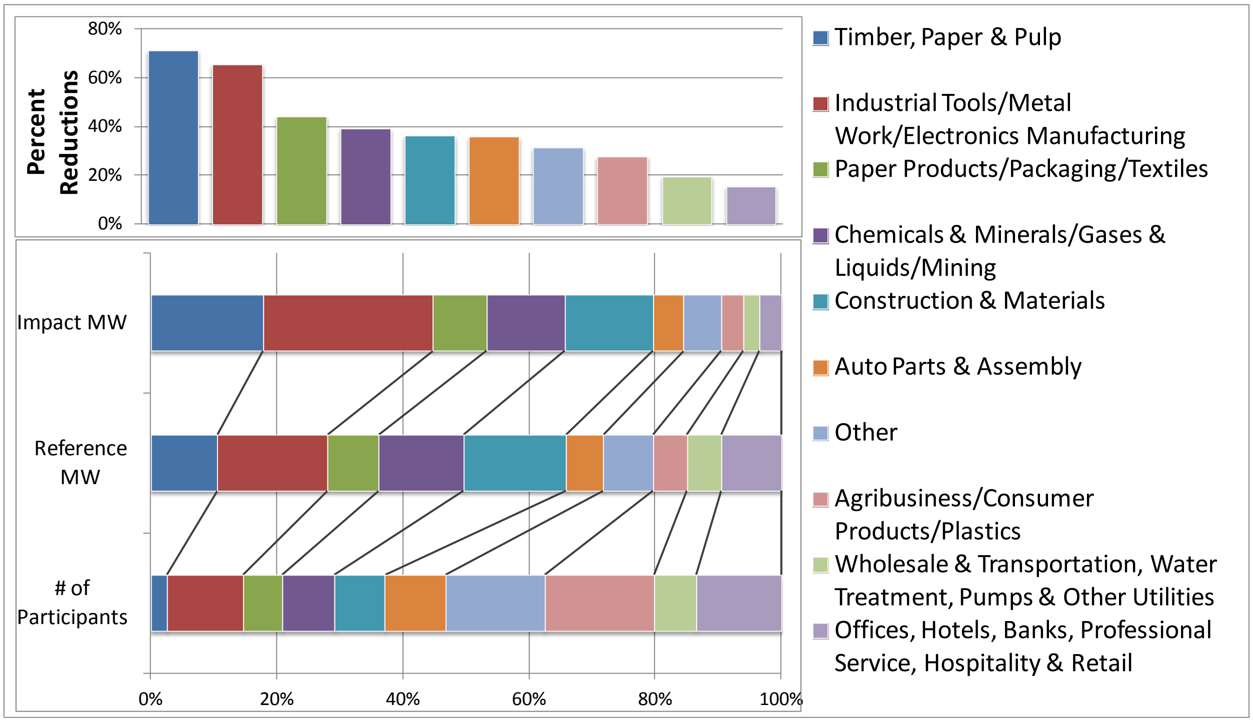
Manufacturing" group. Participants who fall under that category tend to give about 65% of reference MW. The next largest category is "Construction & Materials" and participants in this group tend to deliver 36% of reference MW. However, "Timber, Paper & Pulp" has the highest percent reductions at 71%. Percent delivered ranges from 15% to 71% for these industries.

Table 5-5: Summary of 2012 Load Reductions by Industry

Industry	# of Contributors	Reference (MW)	Impact (MW)	% of Reference MW
Agribusiness/Consumer Products/Plastics	54	44.6	12.4	28%
Auto Parts & Assembly	30	47.2	16.9	36%
Chemicals & Minerals/Gases & Liquids/Mining	26	111.1	43.4	39%
Construction & Materials	25	133.1	48.4	36%
Industrial Tools/Metal Work/Electronics Manufacturing	38	143.0	93.6	65%
Offices, Hotels, Banks, Professional Service, Hospitality & Retail	42	78.0	12.0	15%
Other	49	65.4	20.6	31%
Paper Products/Packaging/Textiles	19	67.2	29.5	44%
Timber, Paper & Pulp	8	87.3	62.2	71%
Wholesale & Transportation, Water Treatment, Pumps & Other Utilities	21	44.5	8.7	20%

Figure 5-3 visually depicts the program's concentration of customers, loads and percent reductions across industries. For example, customers in the "Timber, Paper & Pulp" category accounted for 3% of customers but were responsible for 11% of the total reference MW and 18% of the total impact. In other words, these customers were not only larger than average, but provided deeper demand reduction than average. The reverse is true for "Offices, Hotels, Banks, Professional Service, Hospitality & Retail." They were 13% of the program customers, but accounted for 9% of the total reference load and 3% of the reductions. In other words, these customers were smaller than average, and also reduced a smaller share of their loads than the average customer.

Figure 5-3: Concentration of Customers, Loads and Demand Reductions by Industry



5.5 Multi-year Performance of DR-3

Analyzing DR-3 performance over multiple years allows us to explore the reliability of aggregator programs and variability in response patterns. It is also useful for defining the load reduction capability when all resources are dispatched jointly, which are referred to as *ex ante* impacts throughout this report.

There are three main aspects to aggregator performance: the ability to build DR resources, the ability to ensure the resources remain available and the ability to deliver expected DR resources. The ability to build resources is analogous to commitments to build generators. Not all generators scheduled to be built are built on time or even built at all. Likewise, DR aggregators can meet, exceed or fall short of commitments to build new DR resources according to a prespecified schedule. The ability to build resources according to schedule is not typically factored into performance reliability of either generators or DR resources. However, it has real implications for long-term system planning.

Once DR resources are built, the key question is how reliably they perform relative to the expected reductions. In this context, reliability includes two components: preannounced, nonperformance, typically due to facility maintenance by large electricity customers; and deviations between DR resources scheduled and delivered during actual activations. These differences can be attributed to shortfalls or over delivery by aggregators, or structural flaws such as bias in the settlement baselines, which is discussed later.

From 2008 to 2012, OPA dispatched the program 44 times. In each instance, customers who signed up for the 200-hour option were dispatched, while customers on the 100-hour options were activated more sparingly and were dispatched 31 times.[7] In total, both options were dispatched jointly 31 times. Figure 5-4 depicts the weather and Ontario system load conditions when DR-3 was dispatched. Table 5-6 shows the distribution of events by year, month, day of week and start time for each option.

Figure 5-4: Event System Load and Weather Conditions (2008-2012)

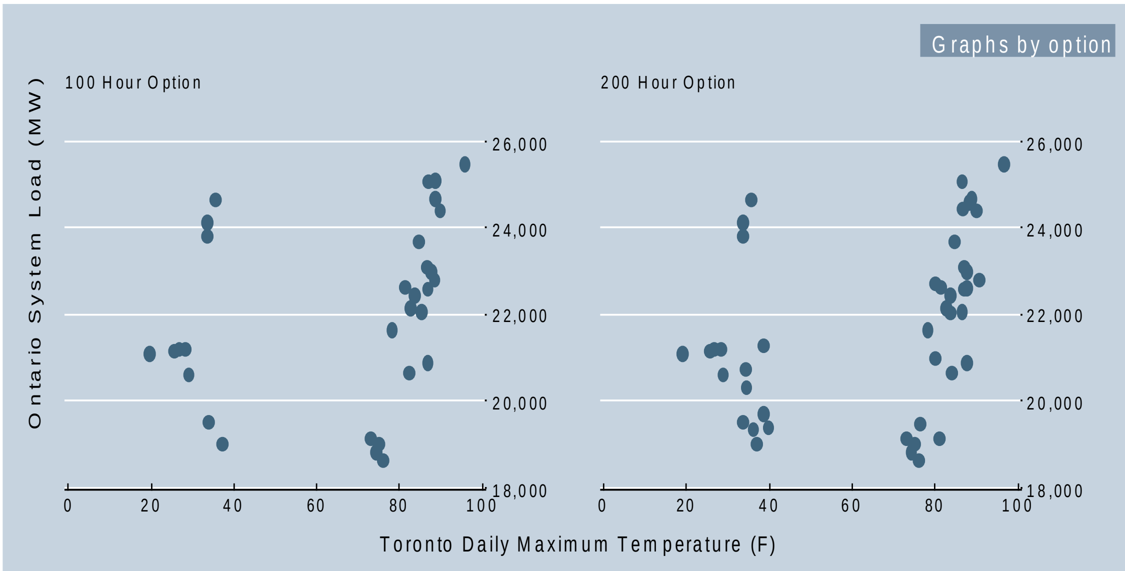


Table 5-6: Distribution of Event Conditions by Year, Month, Day of Week and Start Time

Year	100-hour Option	200-hour Option
2008	8	14
2009	4	6
2010	6	8
2011	8	11
2012	5	5
Total	**31**	**44**

Month	100-hour Option	200-hour Option
May	2	2
June	4	6
July	6	8
August	5	6
Sept	9	11
October	0	2
November	4	8
December	1	1
Total	**31**	**44**

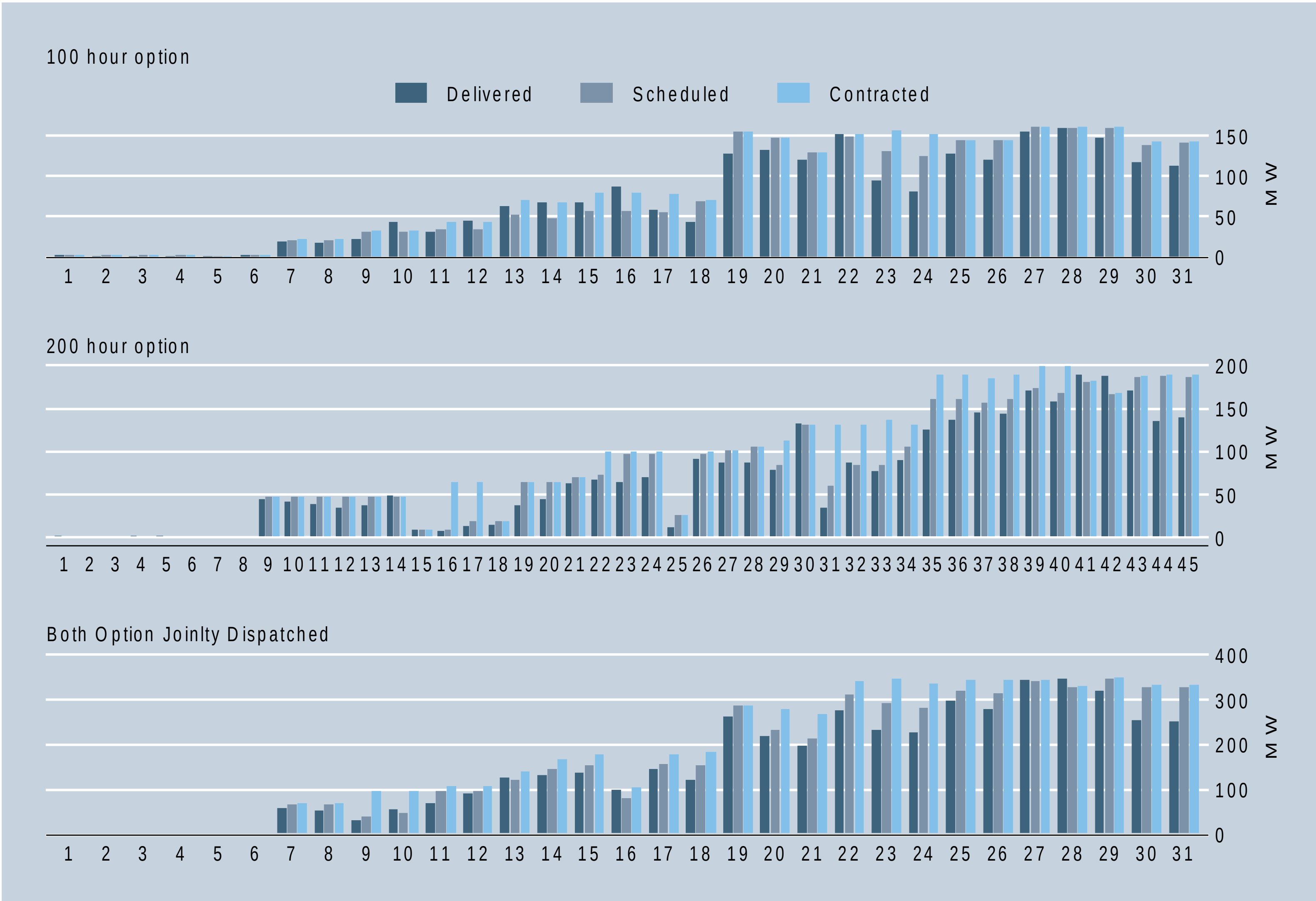
Day of Week	100-hour Option	200-hour Option
Monday	4	10
Tuesday	8	9
Wed	8	13
Thursday	8	9
Friday	3	3
Total	**31**	**44**

Event Start Time	100-hour Option	200-hour Option
12 PM	2	2
1 PM	5	7
2 PM	9	10
3 PM	8	11
4 PM	6	9
5 PM	1	5
Total	**31**	**44**

Figure 5-5 summarizes the overall performance of aggregator resources in Ontario from 2008 to 2012. It compares side-by-side the delivered, scheduled and contracted DR resources. The results are presented separately for the 100- and 200-hour options and for instances when both options were jointly dispatched.

[7] In one occasion, only customer signed up for the 100 hour, 12 pm 6 pm availability groups were dispatched. That event is excluded from the multi-year analysis because it included a small fraction of participants.

Figure 5-5: DR-3 Event Performance based on Historical Events, by Dispatch Option (2008-2012)



The delivered reductions in the graph are based on evaluation results and not on the settlement baselines. As discussed later, it highlights the importance of assessing accuracy of settlement rules in advance of signing multi-year contractual agreements. The gap between the delivered and scheduled DR resources reflects deviations between DR resources expected by the system operator and resources delivered during actual activations. A substantial portion of this difference is explained by systematic bias in the settlement baselines. The difference between delivered and contracted resources is similar to comparing how much generators can reliably deliver against nameplate capacity.

Table 5-7 summarizes aggregator performance by year for each of the options. It contains the data underlying Figure 5-5. It is useful for assessing if aggregators improved at delivering the scheduled demand reductions or improved in making the contracted resources available.

Table 5-7: DR-3 Historical Performance by Year (2008-2012)[8]

Option	Year	Events	Sites	Avg. Delivered	Avg. Scheduled	Avg. Contracted	% Delivered of Scheduled	% Scheduled of Contracted	Realization Rate
100-hour	2008	8	5.4	1.4	2.7	3.0	50.1%	92.2%	46.2%
	2009	4	22.5	30.6	33.5	44.7	91.4%	74.9%	68.5%
	2010	6	88.0	74.8	70.0	75.1	106.8%	93.3%	99.6%
	2011	8	188.0	105.4	125.6	132.5	83.9%	94.8%	79.5%
	2012	5	221.0	123.3	137.0	146.0	90.0%	93.8%	84.4%
200-hour	2008	14	2.7	1.4	1.7	1.7	82.0%	100.0%	82.0%
	2009	6	11.3	11.5	15.5	17.2	74.3%	90.3%	67.1%
	2010	9	32.8	34.3	46.3	53.3	74.2%	86.9%	64.5%
	2011	11	79.2	88.9	92.5	113.1	96.1%	81.8%	78.6%
	2012	5	52.2	123.6	131.4	133.6	94.1%	98.4%	92.6%
Both Jointly Dispatched	2008	8	7.9	2.7	4.4	4.6	61.2%	95.0%	58.1%
	2009	4	34.0	42.2	49.9	63.6	84.7%	78.4%	66.4%
	2010	6	118.3	106.8	114.3	124.9	93.4%	91.6%	85.6%
	2011	8	275.0	203.4	225.5	250.8	90.2%	89.9%	81.1%
	2012	5	273.2	246.9	268.3	279.5	92.0%	96.0%	88.3%

8 These values were calculated at the customer level and then aggregated to the average event day. These values also exclusively used July contracted MW.

In general, 2008 performance is relatively high but includes so few sites that conclusions cannot be drawn based on that year. Aggregator resources performed better in 2010-2012 than they did in earlier years; a larger share of scheduled resources were delivered and larger share of contracted resources were available for operations.

Thus far, the multi-year analysis presents data on historical event performance. However, a key factor in assessing the expected performance of DR-3 on a forward looking basis is the change in the customer mix over the past five years. Different customers have been activated a different number of times, based on their enrolment data and program option selected. If each event is treated as an independent data point, sites with a larger history of events are given disproportionate weight relative to their demand reduction commitments than newer sites.

To calculate overall performance, FSC first calculated performance for each individual site based on each sites' event history. By doing so, each site was weighed equally. In other words, this avoids overweighting the results for customers that enrolled earlier or who were dispatched more often because they enrolled in the 200-hour option. Once the individual site's historical demand reductions are calculated, they are then aggregated to estimate the overall performance factors: the share resources that is scheduled or available, and the percent of scheduled resources that are delivered. These two factors, jointly, determine the realization rate (delivered MW/ contracted MW). Table 5-8 summarizes the performance factors for each DR-3 program option and in total, when performance is calculated based on each site's event history. The table includes all sites that experienced an event between 2008-2012 and had adequate hourly or sub-hourly data for evaluation.

Table 5-8: DR-3 Performance Factors by Program Option

Program Option	Sites	Impact	Scheduled MW	Contracted MW	% Scheduled	% Delivered	Realization Rate
Option B 200-hour 12-9 PM Summer Availability	189	157.9	163.3	186.1	87.7%	96.7%	84.8%
Option A 100-hour 12-9 PM Summer Availability	204	128.6	149.7	160.1	93.6%	85.9%	80.3%
Option C 100-hour 12-6 PM Summer Availability	72	23.5	20.8	21.5	97.1%	112.6%	109.3%
Total	465	309.9	333.9	367.6	90.8%	92.8%	84.3%

Overall, DR-3 delivers 84.3% of the contracted demand reductions. On average, 90.8% of resources are scheduled and 92.8% of scheduled resources are delivered. The specifics vary for each option. The customers enrolled on the 100-hour option with 12-6 PM summer availability performed above average but are a relatively small share of the program. This information was used to develop estimates of resources available for operations, which are based on the percentage of scheduled resources actually delivered, 92.8%. It is also used to develop *ex ante* estimates for planning, which are based on the realization rate, 84.3%, which factors in both scheduled nonperformance and deviations between scheduled and delivered resources.

5.6 Load Reductions for Operations

From a system operator perspective, a key aspect of the DR-3 program is the reliability of the scheduled commitments. For a generator, once it is committed, whether it can deliver its resources is a function of whether it is in operation and power flow. In general, the generator is either operating or experiences an unforeseen forced outage. In other words, it either meets the day-ahead commitment in full or not at all. In contrast, DR-3 load reductions can deviate from the scheduled resources and the resource delivered is estimated rather than directly observed.

A primary reason why DR-3 load reductions have deviated from the scheduled MW in the past is an error in the settlement baselines – or, more specifically – the lack of adjustments to account for upward bias in the baseline method. The contracted load reduction values should not be used for scheduling resources unless the upward baseline bias is eliminated. Until upward baseline bias is corrected, the scheduled load reductions should be adjusted for baseline error. Doing so will ensure that DR-3 indeed meets its scheduled load reduction commitments.

Completely eliminating baseline bias is difficult because no single baseline provides the best fit for all contributors. One alternative is to modify the current standard baseline in order to minimize bias given the current participant mix. Such an approach improves overall accuracy in the short-term but may need recalibration as the program grows and the participant mix changes. Since 16% of the program consists of direct participants, another option is to standardize a baseline accuracy and selection process for direct participants rather than rely on the standard baseline.

In addition to adjustments for settlement baseline bias, scheduling needs to take into account uncertainty due to variation in participant performance and statistical uncertainty. These three factors were combined based on historical data.[9] Table 5-9 summarizes the amount of load reduction that can be expected for operations, along with the 90% and 80% confidence intervals. It both adjusts for settlement baseline bias and participant performance and incorporates the event-day uncertainty of reductions. It factors in the fact that, on average, DR-3 participants deliver 92.8% of scheduled resources. The confidence bands factor in the event-by-event variation in the percentage of scheduled resources that are delivered.

Table 5-9 enables OPA or the IESO to determine the amount of DR-3 resources for day-ahead commitments after factoring out scheduled nonperformance. For example, if 400 MW of the contracted resources are scheduled to perform, derating for the bias inherent in the settlement baseline and average participant performance reduces the expected reductions to approximately 350 MW.

[9] The distribution that best described the historical variation in performance – weighted by the day-ahead MW commitment – was assessed through three distributional goodness-of-fit measures. Once that was accomplished, the statistical uncertainty was incorporated via Monte Carlo simulation.

Table 5-9: Expected Load Reduction for Operations (With Confidence Intervals)

Scheduled MW	Expected Load Reduction	90% Confidence Interval		80% Confidence Interval	
		Lower	Upper	Lower	Upper
0	0.0	0.0	0.0	0.0	0.0
25	21.8	12.9	30.7	15.3	28.4
50	43.6	25.8	61.3	30.5	56.9
75	65.5	38.7	92.0	45.8	85.3
100	87.3	51.6	122.6	61.0	113.7
125	109.1	64.5	153.3	76.3	142.2
150	130.9	77.4	183.9	91.6	170.6
175	152.7	90.3	214.6	106.8	199.1
200	174.5	103.2	245.3	122.1	227.5
225	196.4	116.1	275.9	137.4	255.9
250	218.2	129.0	306.6	152.6	284.4
275	240.0	141.9	337.2	167.9	312.8
300	261.8	154.8	367.9	183.1	341.2
325	283.6	167.7	398.6	198.4	369.7
350	305.5	180.6	429.2	213.7	398.1
375	327.3	193.5	459.9	228.9	426.5
400	349.1	206.4	490.5	244.2	455.0
425	370.9	219.4	521.2	259.5	483.4
450	392.7	232.3	551.8	274.7	511.9
475	414.5	245.2	582.5	290.0	540.3
500	436.4	258.1	613.2	305.2	568.7
525	458.2	271.0	643.8	320.5	597.2
550	480.0	283.9	674.5	335.8	625.6
575	501.8	296.8	705.1	351.0	654.0
600	523.6	309.7	735.8	366.3	682.5

5.7 Load Reductions for Planning – *Ex Ante* Results

The *ex ante* load reductions are designed to reflect the load reductions that can be expected for the purpose of planning. These estimates factor in both scheduled nonperformance and the extent to which participant load reductions deviate from scheduled resources.

In the case of DR-3, many large participants are not weather sensitive but some are exposed to IESO wholesale market prices and several of them are price responsive. To a large extent, the wholesale market prices are not necessarily relevant since the load reduction is contractual. As a result, the *ex ante* load reductions factor in the amount of contracted MW and the degree to which these customers have historically delivered those load reductions. In other words, they are derated for nonperformance days and for deviations from the scheduled load reductions (which already factor in planned nonperformance).

Figure 5-6 and Table 5-10 provide the *ex ante* load reduction estimates. These summarize the load reductions to be employed for planning purposes at varying levels of total contractual load reductions in the DR-3 program. They incorporate planned nonperformance and differences between load reductions scheduled on a day-ahead basis and the amount actually delivered. Overall, 84.3% of the contracted load reduction is expected for planning. The summer capacity of DR-3 in 2012 was 408 contracted MW in the program, which equals an *ex ante* load reduction estimate of 343 MW.

Table 5-10 can also be used to estimate *ex ante* load reductions as the program continues to grow. For example, if the DR-3 program grows to 450 contracted MW, the *ex ante* load reduction estimate would be 378 MW.

As noted earlier, while the *ex ante* reflects the net load reduction capability, they do not directly account for the timing and magnitude of load reductions. They also do not account for limitations on the maximum event duration, hours of availability, amount of advance notice, number of consecutive event-days the resource is available and other factors. For planning, it is necessary to estimate and incorporate the extent to which DR resources serve as capacity without increasing the likelihood of unserved electricity – a value more commonly known as effective load carrying capacity. This typically requires subsequent analysis and adjustments and is done in the long-term planning analysis.

Figure 5-6: *Ex Ante* Load Reduction Estimate For Long-term Planning

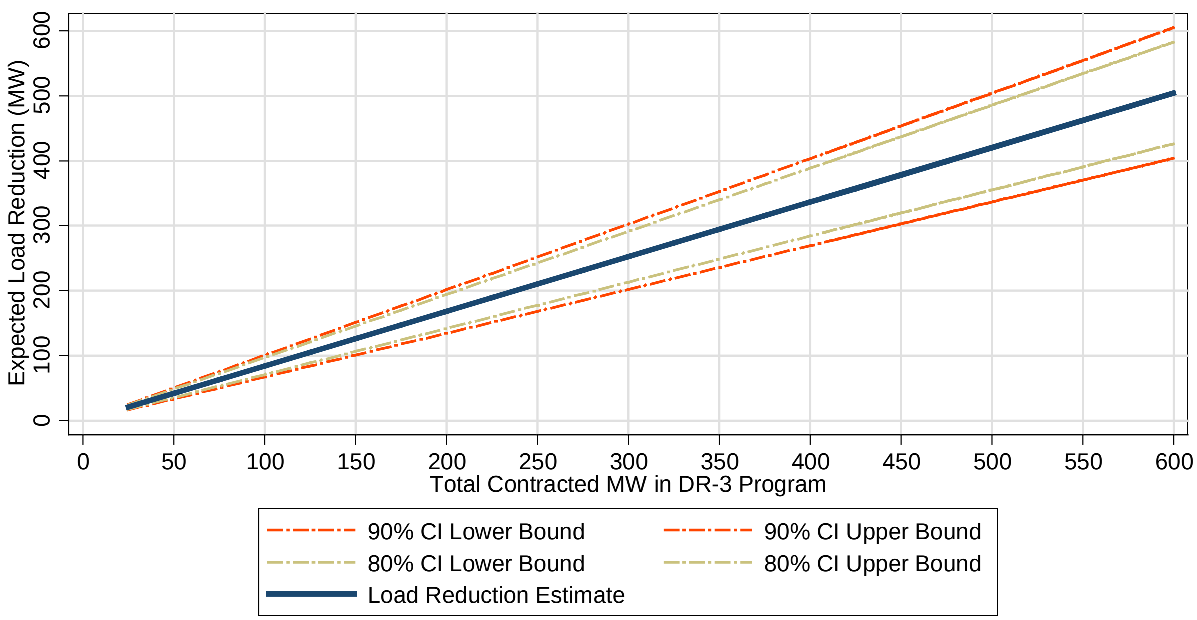


Table 5-10: *Ex Ante* Load Reduction Estimate for Long-term Planning

Total Summer Contracted MW	Load Reduction Estimate	90% Confidence Interval		80% Confidence Interval	
		Lower	Upper	Lower	Upper
25	21	17	25	18	24
50	42	34	50	36	49
75	63	50	76	53	73
100	84	67	101	71	97
125	105	84	126	89	121
150	126	101	151	107	146
175	147	118	176	124	170
200	168	135	202	142	194
225	189	151	227	160	218
250	210	168	252	178	243
275	231	185	277	195	267
300	252	202	302	213	291
325	273	219	328	231	316
350	294	236	353	249	340
375	315	252	378	266	364
400	336	269	403	284	388
408					
(2012 Summer Capacity)	343	275	411	290	396
425	357	286	428	302	413
450	378	303	453	320	437
475	399	320	479	337	461
500	420	337	504	355	485
525	441	353	529	373	510
550	462	370	554	391	534
575	483	387	579	408	558
600	504	404	605	426	582

5.8 Settlement Baseline Accuracy

DR-3 participant performance and payments are determined through settlement baselines, which need to be calculated quickly, much faster than program evaluations. In many cases, aggregators manage demand reductions to meet contractual obligations as implied by the settlement baselines. As a result, baseline inaccuracy can affect program performance. The 2009, 2010 and 2011 DR-3 evaluations found that the current baseline – which relies on the same hour load from the highest

15 of the last 20 eligible days – overestimated demand reduction by approximately 25%, on average. That is, due to baseline bias, participants could comply with a contractual obligation to reduce 100 MW by delivering 80 MW (80 X 125%=100). This section will analyze to what extent DR-3 program performance is explained by inaccuracy in the current baseline rules. These evaluations will identify several modifications to improve the accuracy of the baseline estimates that are currently being implemented.

To assess accuracy, it is necessary to know the true demand reductions that were delivered. Without knowing the true demand reduction, it is not possible to assess which baseline method produces the most accurate results. For the baseline comparison, demand reduction estimates provided for each contributor by aggregators were introduced on non-event-days with event-like conditions. That way, FSC could test how closely the estimates produced by the baseline matched the known demand reductions. In total, baseline accuracy was tested using 26 proxy event-days for class A and another set of 26 proxy event-days for class B customers. The basic steps of the analysis are as follows:

- *Select 26 event-like days:* Proxy event-days from 2010 to 2012 were selected to match the temperature and system load conditions observed in actual event-days as closely as possible.[10] However, for the GAM High-5 eligible customer group, the top 20 Ontario system load days of 2011 and 2012 were not eligible proxy event-days.
- *Simulate the load reduction:* Aggregators and direct participants provide information about the demand reductions each facility can deliver. FSC assumed each facility delivered those load reductions during the event window from 2-6 PM. To do so, FSC simply subtracted the demand reduction from the actual electricity usage of each customer during the curtailment hours of each proxy event-day.
- *Construct the candidate settlement baselines for each proxy event and estimate the demand reductions:* Multiple baselines were calculated so they could be compared side-by-side. These include the current baseline and the baselines that proved relatively accurate in prior analysis. The baselines were then used to estimate the demand reductions.
- *Assess the accuracy of each of the candidate settlement baselines:* For each of the curtailment events, FSC knew the demand patterns without curtailment and the demand reductions that were introduced. As a result, FSC could assess how accurate the baselines estimated the demand reductions.

The results focus on the accuracy of the demand reductions estimated using settlement baselines rather than on accuracy of the baselines. The sole reason baselines are calculated in the first place is to estimate demand reductions for settlement, but they are means to an end, not the end itself. In general, baseline estimates are more accurate than the demand reduction estimates they produce. A baseline that overestimates by 3% will indicate reductions of 23% when reductions are 20%. In other words, a 3% baseline error in this example can lead to a 15% error in the demand reductions (3%/20%).

Figure 5-7 and 5-8 compare the impact estimation error between the 15-of-20 baseline and the 10-of-10 with adjustment baseline for class A and class B customers respectively. Reference loads are known on proxy event-days and therefore impact error can be accurately assessed. The perfect baseline would have zero error across all proxy events because each bar represents by how much the baseline over or underestimated impacts. For example, on November 13, 2012 the 15-of-20 baseline

[10] The proxy event-days are similar to the days used to test the out-of-sample accuracy of the regression. The selection of these days is explained in Appendix B, Section 1.

overestimated impacts by 32%. This means that if actual impacts were 100 MW that day, then the 15-of-20 baseline would have estimated impacts of 132 MW. These figures show that the 15-of-20 baseline with no adjustment overestimates on the majority of proxy event-days. This is not surprising because the 15-of-20 baseline is selecting the hottest 15 day subset over the past 20. On the other hand the 10-of-10 baseline over and under predicts, resulting in less error on the average event-day. Magnitude of error is also smaller for the 10-of-10 baseline versus the 15-of-20 baseline.

Figure 5-7: Comparison of Actual Reductions vs. Baseline Reductions For Class A Customers

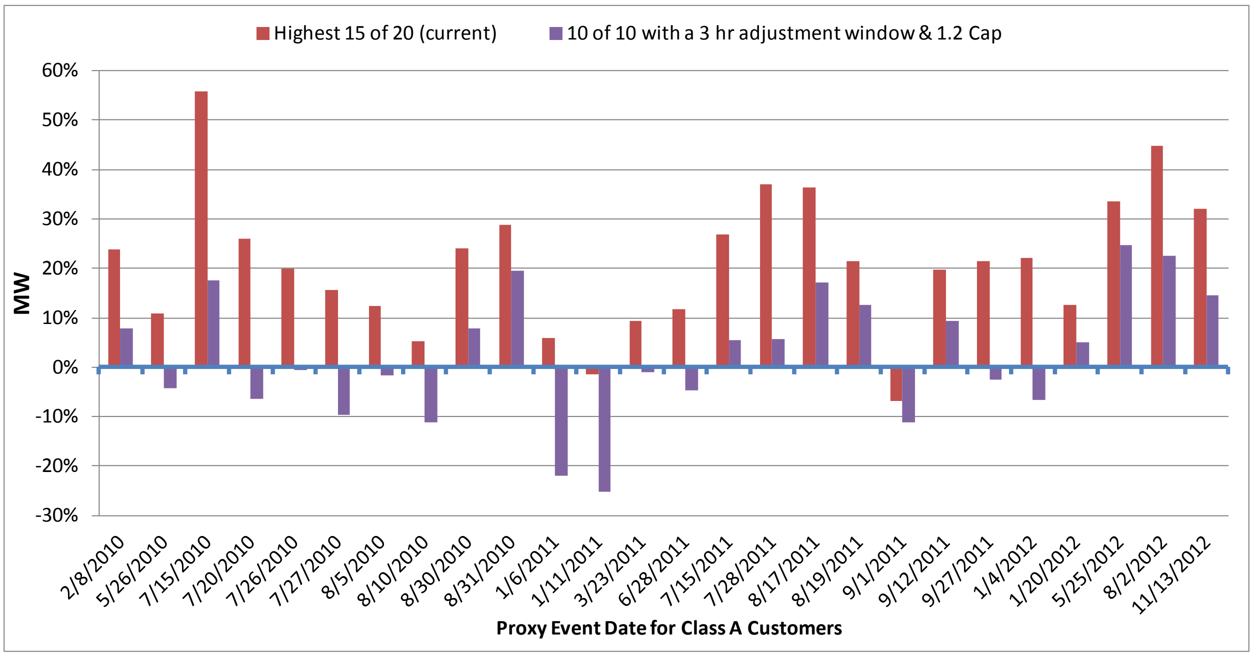


Figure 5-8: Comparison of Actual Reductions vs. Baseline Reductions For Class B Customers

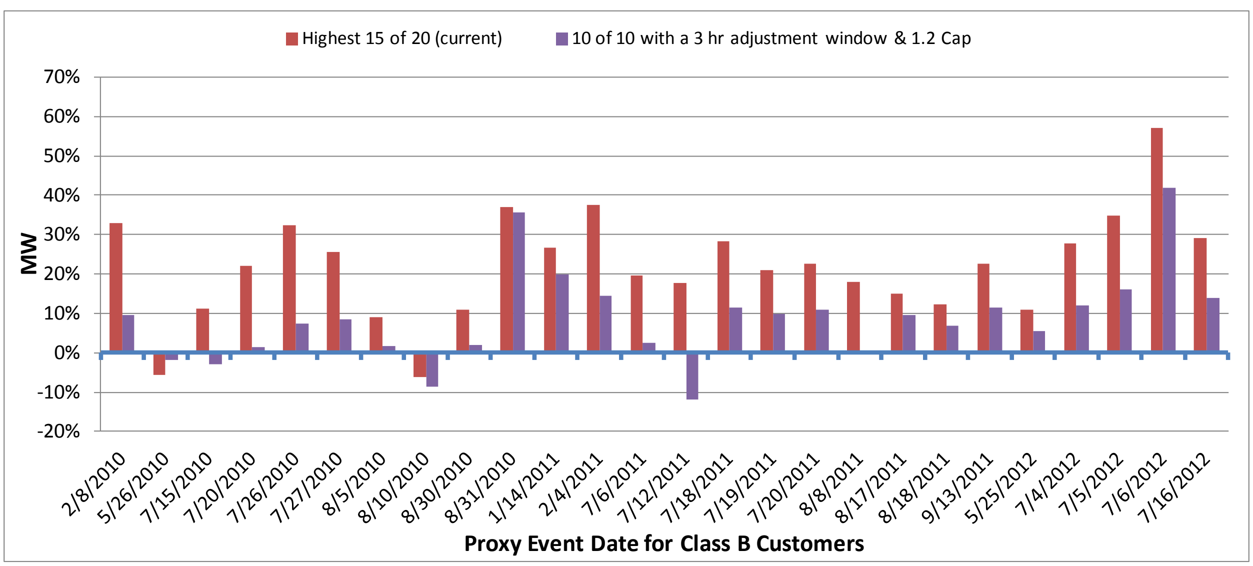
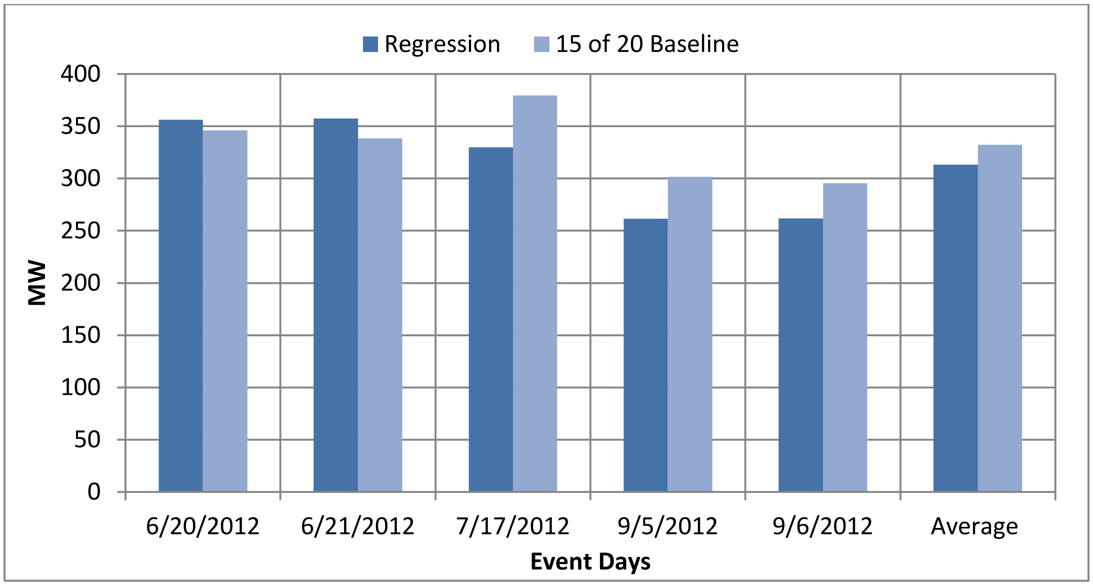


Figure 5-9 addresses how much of the underperformance detected in *ex post* evaluation can be attributed to differences in baseline calculation. The figure shows the impacts estimated by the regression baseline side-by-side with the impacts estimated from the 15-of-20 baseline for 2012 events. The 15-of-20 baseline estimated impacts were on average 19 MW greater than regression estimated impacts in 2012.

Figure 5-9: Comparison of 15-of-20 Baseline and Regression Estimated Impacts



The baseline calculation analysis provides the following conclusions:

- Bias in the current baseline (highest 15 of the last 20 eligible days) explains a large portion of the difference between the scheduled and delivered demand reductions;
- Transitioning to a more accurate baseline will better align settlements with the actual demand reductions and will reduce the amount of over and underpayments to individual DR-3 participants; and
- The attribution of DR-3 impact to Local Distribution Companies is unlikely to change since it is based on the evaluation impacts rather than on settlement baseline results.

5.9 Conclusions and Recommendations

In 2012, DR-3 performance improved although the program grew at a slower pace than in prior years. The *ex ante* load impact estimate for DR-3 is 343.6 MW,[11] representing 84% of the peak summer contracted capacity (343.6 MW out of 408 MW). The *ex ante* estimate is based on a multi-year analysis of performance and accounts for both planned nonperformance and deviations between scheduled and delivered reductions.

DR-3 resources also remains highly concentrated despite adding additional contributors. The largest 20 contributors account for 60% of the contractual demand reduction – that is, less than 5% of contributors account for the majority of the load reductions.

[11] This *ex ante* value factors in the amount of contracted load reduction, the share of contracted reduction scheduled on a day-ahead basis and the degree to which these customers have historically delivered scheduled demand reductions.

In addition, the current baseline method for DR-3 remains biased and, in aggregate, overstates demand reductions by approximately 22%. The degree of bias has not decreased as the program has grown and the participant mix has evolved. The exact effect on individual settlement accounts varies widely – some aggregators are underpaid while others are overpaid relative to their actual demand reduction. Since aggregators comply with the baseline rules, correcting the baseline should better align settlements with actual demand reductions and reduce over and underpayments. A realignment is unlikely to affect the attribution of DR-3 impacts to LDC's, since those are currently based on evaluation results, not settlement baselines.

Key recommendations resulting from the evaluation include:

- *Improve the data tracking and collection process.* The current process is highly manual and introduces significant risk for error. It also makes comprehensive analysis difficult and prolongs the settlement time frame with DR-3 participants. The data files are kept separately for each customer and month on separate spreadsheets that are not consistently labeled. As part of the evaluation, FSC was unable to obtain data for contributors that accounted for 8% of the contracted resources despite repeated efforts. Considering that nearly $35 million of settlement payments are conducted annually for DR-3, the lack of consistent and systematic data tracking can have substantial monetary implications and needs to be improved if only to reduce the potential for errors.
- *Ensure that settlement accounts reflect aggregation.* The empirical evidence shows that the accuracy of settlement estimates degrades severely when settlement accounts fail to aggregate. Ideally, no single contributor should be more than 20% of load in a settlement account. There may be exceptions for load that are proven to be highly predictable using out-of-sample testing. Currently, DR-3 participants are required to have separate settlement accounts by IESO zone. While the ability to dispatch locally is important, it is not necessary to settle by IESO zone for each event. Doing so limits the benefit of aggregation and introduces a substantial amount of payment error – many accounts are either over or undercompensated due to the inherent noise associated with disaggregated baseline results.
- *Require settlement accounts to reduce at least 10% of historical 12- 9 PM summer weekday demand.* The empirical evidence shows that settlement estimates are unreliable when the percent reductions are small. Therefore, by requiring a minimum percent reduction for settlement account, it is easier to distinguish actual demand reduction from inherent variability in loads – background noise.
- *Implement more accurate baselines.* The current baseline has repeatedly been shown to produce inaccurate estimates. In aggregate, it leads to an average 22% overestimate of demand reductions. It also leads to inherently unfair results for different settlement accounts. Some settlement accounts must over deliver to comply with settlement rules while other settlement accounts can under deliver and still comply. While no single baseline will be best for all settlement accounts, there is substantial room to improve payment accuracy across all settlement accounts.
- *Conduct tests to assess if DR-3 can be used to specify specific amounts of DR resources for specific hours under the current rules*. The current practice of calling either all customers on the 200-hour option and all customers on the 100- and 200-hour options for the same hours, does not provide the flexibility necessary for system operations.
- *Consider developing a different DR product for large customers (average monthly peak usage greater than 5 MW) that better utilizes demand reductions for operations and does not rely on settlement baselines (which are inherently less accurate for individual customers).*

Appendix A DR-2 Detailed Analysis Methodology and Validation

DR-2 is a unique program and presents several challenges from an evaluation perspective. To calculate load reductions for DR-2, the participants' load in the absence of load shifting – the counterfactual – must be estimated. For most DR programs, this is accomplished by using pre-enrolment data, observing behaviour during non-event-days (a within-customer control) or through a control group. It is not possible to observe the naturally occurring facility behaviour while customers are participating in DR-2 since the program is designed to produce load shifting on a daily basis. Given the limited number of facilities of similar size and industry, it is also not possible to develop an adequate control group.

In this instance, the pre-enrolment load contains a large number of days in which natural load patterns were disrupted due to participation in DR-1. However, the time span when customers participated in DR-1 includes many non-event periods and enables the estimation of naturally occurring load patterns through regression analysis. In addition, the naturally occurring facility behaviour can be observed for the time periods prior to participation in any DR program.

Technically, this analysis approach is referred to as an interrupted time series – a quasi-experimental method. The key to an interrupted time series analysis is knowing the exact point at which an intervention occurred. An intervention can be an event-day or a payment stimulating a change in behaviour – in this case load shifting – after a specific point. Part of the challenge for the DR-2 evaluation is the multiple interventions to participant load that occurred since 2007. In 2007 and 2008, they participated in DR-1, which altered their behaviour during event-days. From March to October 2009, they were engaging in load shift during the transition period. For November 2009 through December 2010, there was a different intervention in place – namely, the DR-2 full program rules were in effect. A strong point of this evaluation is that the timing of each intervention (i.e., shift in DR participation) is well documented and the timing of the load shift coincides with the introduction of the DR-2 rules. The downside is that disentangling the effect for each of the program effects requires us to rely on historical participants that are distant from the evaluation period.

The key weakness of an interrupted time series design is potentially confounding program effects with other factors that happen to occur during the participation period. In this case, the threat is real. The economic downturn occurred at the same time as the DR-2 program launched, first in transition mode and, subsequently, in its final form. Anecdotal evidence from participants indicate that the economic downturn affected the DR-2 participants profoundly. In addition, all of the participants faced wholesale market prices, which were substantially lower in 2009 and 2010 than in prior years.

The careful analysis of the load data show that the key difference between participant loads prior to and during DR-2 participation was the number of days facilities were shut down for the day. While enrolled in DR-2, participants provided the IESO notice of nonperformance days. These were typically, though not always, days in which the facility was shut down. For the pre-enrolment time frame, days when the facility was shut down can be identified by their load levels. In order to provide comparability across the changing economic conditions, days in which the facilities were shut down were excluded from DR-2 transitions and enrolment periods as well as from the pre-enrolment data.

Facility shut downs are factored into the long-term planning reduction estimates by derating based on the historical frequency of nonperformance days.

The regressions provide the hourly reductions of the program but the load shifting to off peak periods does not always match the reduction during the peak period. This reflects the fact that some customers cannot fully shift their load reductions to off peak periods due to the physical capacity of the plant. To obtain the amount of load shifting, FSC normalized the pre-enrolment, transition and enrolment results. This process was simple and, in essence, subtracts out differences in overall consumption and allowed FSC to focus solely on changes in the load shape. By construction, both the reference load shape and the transition period load shapes are directly comparable in terms of consumption.

A.1 Regression Model Development

The load reductions of the program were analyzed using regression analysis. The load usage patterns and load reductions were estimated using 2006 through 2012 hourly demand data for pre-enrolment, transition and enrolment periods. As noted earlier, plant shut downs were excluded from the analysis and are factored through derating for the frequency of nonperformance days. Individual customer regressions were developed for each participant. The regressions were developed with the primary goal of accurately estimating *ex post* load reductions and energy use patterns. The focus was primarily on prediction accuracy and on the robustness of the variables reflecting load reductions.

Variation in the demand for the three participants was step-like, with load levels clustered around specific values that reflect different industrial processes in operation. This creates a unique data pattern as the variation in demand is not continuous. The expected load level is best reflected by the probability that one, two or multiple processes are operating jointly. As a result, a specific regression technique – ordinal probit regression – was applied for these customers. The approach essentially predicts the likelihood that the facility is at specific demand levels. It also allowed FSC to identify the most likely facility load level and produced the expected load by factoring in the likelihood of each demand level.

At a high level, the regression models can be thought of in terms of three main components:

- *Variables that reflect the average load shape of customers (load shape variables).* These are typically accounted for by various shape indicator variables by hour of day and day of week;
- *Variables that explain deviation in hourly usage from the average load shape.* These include factors such as economic conditions, wholesale market prices, temperature and seasonality; and
- *Variables that estimate the hourly load shifting while participants were on DR-2.* These variables are designed to estimate the hourly effects of the DR-2 program throughout the transitioning and enrolment period. Separate reductions were estimated for each month and hour of weekdays.

Mathematically, the model can be expressed by the following equation:

$$
\begin{aligned}
MW_t = {} & a + \sum_{i=2}^{5} * \sum_{j=2}^{24} b_{ij} * hour_i * daytype_j + \sum_{i=2}^{12} c_i * month_i \\
& + \sum_{i=1}^{24} d_i * hour_i * DailyPriceRatio_t + e * DailyAvgPrice_t + f * TwoWeekAvgPrice_t \\
& + \sum_{i=1}^{24} g_i * hour_i * CDD_t + \sum_{i=1}^{24} h_i * hour_i * HDD_t + \sum_{i=1}^{24} k_i * hour_i * CDDsqr_t + \sum_{i=1}^{24} l_i * hour_i * HDDsqr_t \\
& + \sum_{i=1}^{24} m_i * hour_i * TSX_t + \sum_{i=1}^{24} n_i * hour_i * TSXsqr_t \\
& + \sum_{i=3}^{11} * \sum_{j=2}^{24} o_{ij} * hour_i * month_j * transtitionperiod + \sum_{i=11}^{12} * \sum_{j=2}^{24} p_{ij} * hour_i * month_j * enrolledperiod \\
& + u_t
\end{aligned}
$$

Table A-1 defines the regression variables and describes the effects they seek to identify.

Table A-1: Description of Load Reduction Regression Variables

Variable	Description
MW	Participants' estimated hourly energy usage
a	An estimated constant
b-p	Estimated parameters
$month_i$	Dummy variables for each month, designed to pick up seasonal effects
$daytype_j$	Dummy variables designed to pick up day-of-week effects. Days of the week are classified into five day types: Monday, Tuesday through Thursday, Friday and weekends and holidays
$hour_i$	Dummy variables designed to estimate the effect of variation across the hours of the day (i.e., operating schedule)
DailyPriceRatio	The ratio of the three-hour dispatch price to the daily average of the three-hour dispatch price, designed to capture the substitution effect of hourly prices within a day
TwoWeekAvgPricer	The trailing, two-week, same-hour average market price of electricity, designed to capture the effect of sustained changes in price patterns and levels on electricity consumption
DailyAvgPrice	The daily average of the three-hour dispatch price, designed to capture the effect of daily levels of market prices on electricity consumption
CDD	Cooling degree days (defined as the maximum of 0 or the average daily temperature minus 65°F), designed to reflect the impact of cooling load
CDDsqr	The square of CDD, designed to reflect nonlinearities in the relationship between temperature and electricity consumption
HDD	Heating degree days (defined as the maximum of 0 or 65°F minus the average daily temperature), designed to reflect the impact of heating load
HDDsqr	The square of HDD, designed to reflect nonlinearities in the relationship between temperature and electricity consumption
TSX	The Toronto Stock Exchange index, designed to capture the effects of the overall economy on electricity consumption

Variable	Description
TSXsqr	The Toronto Stock Exchange index squared, designed to reflect nonlinearities in the relationship between the TSX and electricity consumption
TransitionPeriod	The treatment effect of participation in DR-2's transitional period
EnroledPeriod	The treatment effect of participation in DR-2's enroled period
U	The error term

A.2 Assessment of Model Accuracy and Precision

The accuracy of the regression model was assessed by testing the robustness of the load reduction coefficients and comparing regression predicted load with metered load.

In order to ensure that the regression-based load reductions were not an artifact of the final regression model, multiple regression specifications were tested and their effect on the reduction variables was carefully analyzed. The load reduction variables did not vary much in magnitude or statistical significance with the inclusion or exclusion of explanatory variables. In other words, the load reduction estimates were robust.

Figure A-1 compares actual load, regression-predicted load with DR and regression predicted load without DR for the DR-2 transition and official contractual period in 2009. The predicted load with DR closely matches the metered load for both periods. The graphs exclude scheduled nonperformance days. The program load for the contractual period (Nov 2009 through Dec 2012) is substantially lower than in the transition period. This reflects the fact that one of the direct participants shut down their facility for the last four months of 2009, while another participant shut down for eight months during 2010 and nine months during 2011.

Figure A-1: R-2 Comparison of Regression Predicted and Actual Load for 2012

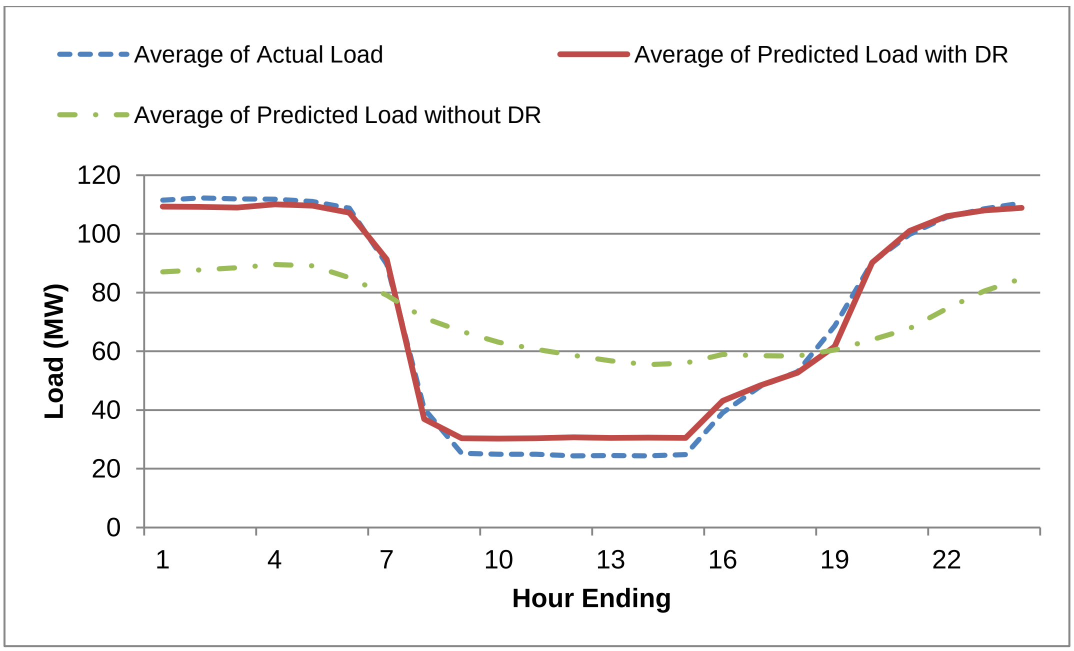


Figure A-2 compares actual load and regression predicted load for pre-enrolment weekdays. The accuracy of the regression model with the pre-enrolment data is critical since it reflects the ability of the regressions to estimate load outside of DR-2 conditions and produce the counterfactual.

Figure A-2: DR-2 Comparison of Regression Predicted and Actual Load for Pre-enrolment Weekdays

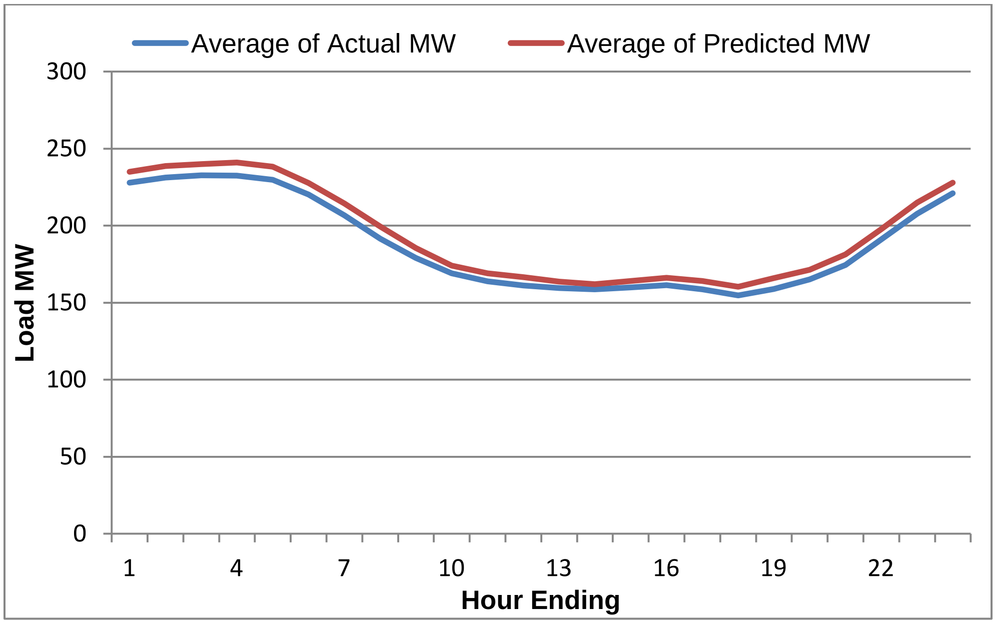
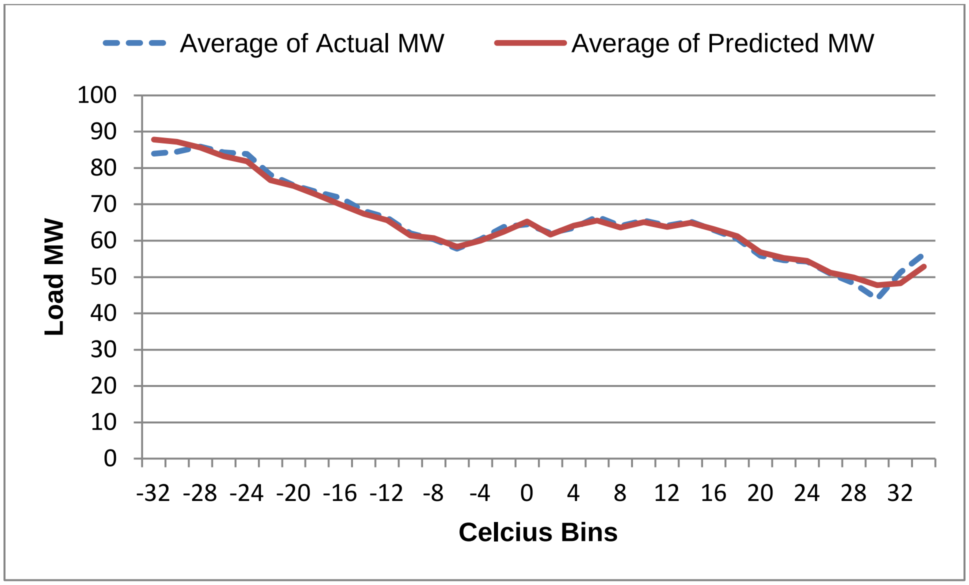


Figure A-3 shows the actual and regression-predicted load across the temperature spectrum for all days, including pre-enrolment, transition and contract period. The regression predicts participant load accurately for both hot temperatures and for cold temperatures. The ability to accurately predict participant load during summer peaking conditions provide added confidence that the regression will accurately quantify summer load reductions.

Figure A-3: DR-2 Comparison of Regression Predicted and Actual Load By Temperature



Appendix B DR-3 Detailed Analysis Methodology and Validation

The DR-3 load reductions were estimated through regression methods using data from 2010-2012 event and non-event-days and the available pre-enrolment data. Individual customer regressions were developed for each participant and aggregated to the settlement account level and the entire program level.

Regression methods were used because they have several distinct advantages over day-matching or baseline methods. First, regression results provide more robust estimates of load reductions and are not as sensitive to biases in the reference load. Second, they can help identify the key drivers and predictors of participant load and load reduction behaviour, including factors such as market prices and weather. Third, they can be used to predict load reductions for operations or for long-term planning by factoring in expected weather, system and market conditions, if available. In other words, they can better predict load reductions on a day-of or day-ahead basis in order to assist operations, as well as predict load reduction resources under system peaking conditions in order to assist long-term planning. Finally, they are not subject to gaming behaviour and can, in fact, detect gaming behaviour.

The regression models were developed with the primary goal of accurately estimating *ex post* load reductions and energy use patterns. The focus was primarily on the accuracy of predictions and on the robustness of the variables that drive the load reductions.

B.1 Regression Model Development

Individual participant regressions were used rather than aggregate or panel models for several reasons. The high concentration of load and participation in DR-3 was a significant factor in the selection of individual participant regressions. Changes in the number and mix of participants, as well as differences in the available data, made it difficult to analyze aggregated data. Moreover, individual participant regressions can better address variation in the size, production schedules and processes of participants. Because of the diversity of participant load patterns, four different models were tested for each customer. The model that most accurately predicted impacts under a false event-day experiment was chosen for each customer. A regression (of the same model) was run separately for each hour to cut down on autocorrelation between the hours.

Figure B-1 depicts the process of creating proxy event-days, testing the regression and choosing the final model for each customer. First, 26 non-event-days with high system load were identified to serve as proxy event-days: one matched to each event-day from 2010-2012. No events were called on these days but conditions were similar to event-days. Next, each of the four models was run for each customer and used to predict the unperturbed load on each proxy event-day. Actual events and proxy event-days were not used in the regression. The model that predicted the closest to the actual load on proxy event-days was chosen as the final model for each customer. Once the best model was identified for each customer, the final regression was run for each customer on a dataset that included the actual events.

Figure B-1: Comparison of Regression Predicted and Actual Load

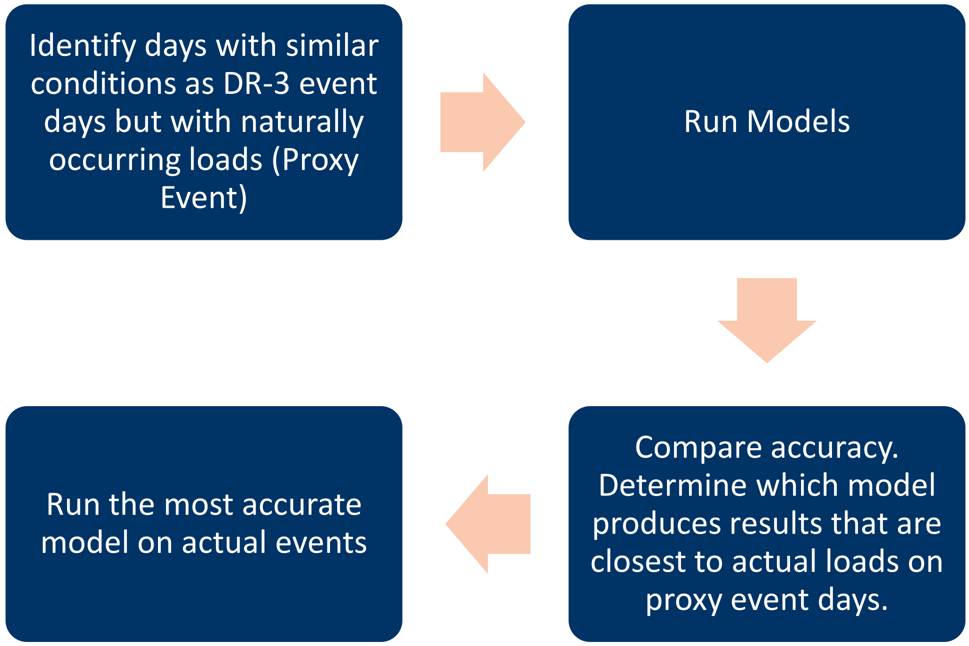
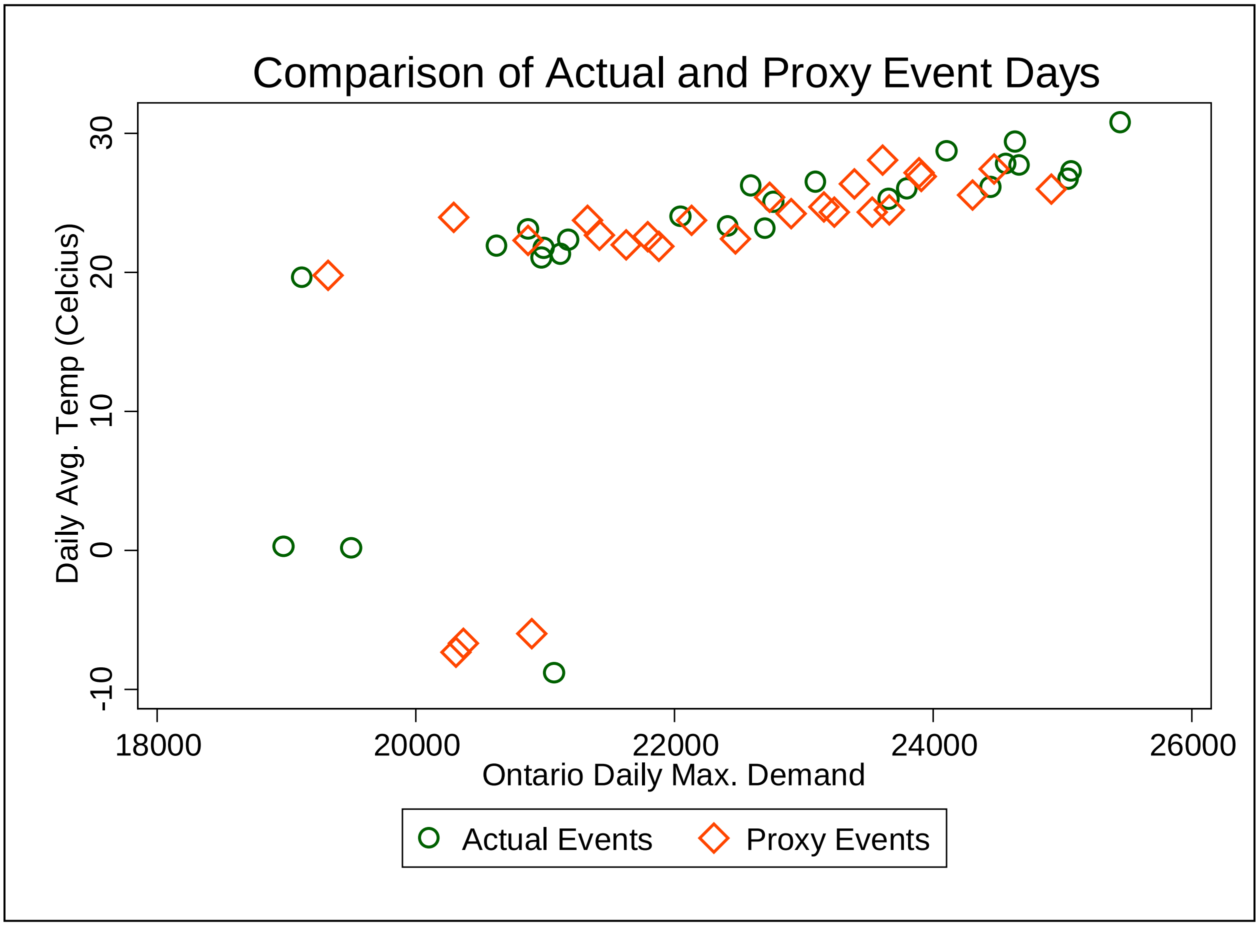


Figure B-2 compares actual and proxy event-days by plotting maximum system load by daily average temperature (°C). The green circles represent actual events while the red diamonds show proxy events. As the figure shows, proxy event-days are very similar to event-days based on these two measures.

Figure B-2: Comparison of Actual and Proxy Event-days



Of the four models tested, all included:

- Day-of-week variables that identified the average load shape of participants in the absence of curtailment events;
- Month variables to capture seasonal variation in load;
- Variables representing prior participant load to predict current load including the average same-hour load for non-event weekdays in the prior four weeks;
- The two-hour ahead price if the contributor participates in the wholesale market; and
- Event variables for each hour of each event-day.

The first model included only the variables listed above. The three other models included different measures of temperature including cooling and heating degree days, cooling degree and heating degree hours and total daily cooling and heating degree hours.

The simplest model used to estimate load reductions is shown below:

$$kW_t = a + \sum_{j=1}^{5} b_j * daytype_j + \sum_{i=1}^{12} c_i * month_i + d_i * FourWeekkW_t + \sum_{j=1}^{26} e_j * EventDay_j + f_t * \ln(DailyAvg\,Price_t) + g_t$$

The log of the average daily price was only included for wholesale customers. Also, GAM eligible customers had a variable similar to "EventDay" that indicated GAM High-5 days.

Table B-1 defines the regression variables and describes the effects they seek to identify.

Table B-1: Description of Load Reduction Regression Variables

Variable	Description
kW	Participant's estimated hourly energy usage
a	An estimated constant
b though g	Estimated parameters
$daytype_j$	Dummy variables designed to pick up day-of-week effects (Tues-Thurs are grouped together because these days are not different from each other)
$month_i$	Dummy variables for month of the year, designed to pick up seasonal effects
FourWeekkW	Four-week moving average of a participant's hourly energy usage
$EventDay_j$	Dummy variables that capture the impact of each of the 10 event-days in the estimating sample
Ln(DailyAvgPrice)	Natural log of the daily average two-hour ahead wholesale market price (only included if the participant is a wholesale market customer)
h	The error term

B.2 Assessment of Model Accuracy and Precision

The accuracy of the regression models was assessed by comparing regression predicted load with actual load on event-like days. These proxy event-days mimic DR-3 event-days in temperature and system load. Considering that DR-3 events are usually called on high system load days during the summer, it is important that the model predicts accurately for those days. Rather than running the model on all of the available load data, a group of 26 proxy event-days with high system load is withheld from the estimation. Although these 26 days are not included in the estimating sample, the model is used to predict load on those days.

Figure B-3 and B-4 shows the results, on average, for class A and class B customers respectively. Using the best of the four tested models for each customer, the analysis was on average able to predict the false event-day reference loads within 6% for class A customers and 3% for class B customers. As seen in the figures, the model accurately predicts load even if those days are not included in the estimating sample.

This validation process most closely aligns with what is expected of the model in the analysis. The *ex post* analysis estimates load reductions by predicting what load would have been if an event was not called. Basically, out-of-sample predictions are generated for days in which actual, unperturbed load data is unavailable. Therefore, out-of-sample validation using randomly selected non-event-days with high system load is a logical test to determine which model is most accurate.

B-3: Actual and Predicted Load on Proxy Event-days For Class A Customers

Date	Actual (MW)	Model (MW)				
		1	2	3	4	Final
8-Feb-10	381	400	405	403	400	404
26-May-10	420	370	369	370	373	371
15-Jul-10	360	386	389	387	385	389
20-Jul-10	368	401	397	398	400	401
26-Jul-10	360	353	348	349	351	351
27-Jul-10	384	368	368	369	369	368
5-Aug-10	431	395	400	399	395	397
10-Aug-10	424	402	401	403	401	401
30-Aug-10	383	371	387	391	376	382
31-Aug-10	385	392	408	412	396	402
6-Jan-11	371	401	402	402	401	402
11-Jan-11	383	394	399	399	394	398
23-Mar-11	532	481	478	482	481	477
28-Jun-11	530	509	508	507	508	506
15-Jul-11	556	505	498	501	505	503
28-Jul-11	499	541	541	549	541	544
17-Aug-11	488	559	554	554	557	557
19-Aug-11	490	530	522	525	528	527
1-Sep-11	615	574	566	572	569	572
12-Sep-11	544	539	533	535	537	535
27-Sep-11	576	558	560	562	563	565
4-Jan-12	443	449	448	450	450	449
20-Jan-12	453	427	436	440	426	435
25-May-12	522	546	539	545	542	541
2-Aug-12	507	529	527	529	527	524
13-Nov-12	507	497	495	497	497	495
Avg. Proxy Event	458	457	457	459	457	457
Bias	MPE	-0.3%	-0.3%	0.2%	-0.3%	-0.1%
Goodness-of-fit	MAPE	5.8%	5.9%	6.0%	5.8%	5.8%
	Normalized RMSE	6.9%	7.0%	6.9%	6.8%	6.9%

B-4: Actual and Predicted Load on Proxy Event-days For Class B Customers

Date	Actual (MW)	Model (MW)				
		1	2	3	4	Final
8-Feb-10	135	137	139	138	137	138
26-May-10	175	173	168	171	172	172
15-Jul-10	195	188	187	188	188	188
20-Jul-10	187	188	188	186	187	187
26-Jul-10	180	181	181	180	181	181
27-Jul-10	186	188	188	187	188	188
5-Aug-10	193	195	193	196	195	196
10-Aug-10	197	194	192	195	194	195
30-Aug-10	187	191	190	193	192	193
31-Aug-10	180	198	197	200	198	200
14-Jan-11	174	177	176	177	178	176
4-Feb-11	180	179	179	179	179	178
6-Jul-11	213	207	207	208	207	208
12-Jul-11	207	208	208	209	209	209
18-Jul-11	209	200	201	204	202	203
19-Jul-11	215	210	209	211	210	210
20-Jul-11	218	208	209	212	210	211
8-Aug-11	212	214	213	214	214	214
17-Aug-11	224	227	226	226	226	227
18-Aug-11	225	227	226	227	227	228
13-Sep-11	251	240	237	240	240	240
25-May-12	276	264	255	263	264	263
4-Jul-12	277	267	265	271	267	269
5-Jul-12	274	266	265	268	267	267
6-Jul-12	265	258	258	265	259	263
16-Jul-12	273	255	253	259	256	257
Avg. Proxy Event	212	209	208	210	209	210
Bias	MPE	-1.3%	-1.8%	-0.8%	-1.2%	-0.9%
Goodness-of-fit	MAPE	2.6%	2.9%	2.4%	2.5%	2.5%
	Normalized RMSE	3.6%	4.1%	3.3%	3.5%	3.4%

It is important to note here the difference between baseline and impact error because impact error tends to be larger than baseline error. Baseline error is what is displayed above and is the percent difference between the predicted reference load and the actual load that occurred that day. Baseline error factors into impact error but it isn't the whole story. Impact error is the percent difference between predicted and actual impacts. Baseline error factors into how different the two impact values are, but it is the size of the impact that determines how that error is scaled. For example, if there are two customers with actual impacts of 10 MW and 20 MW each, a baseline underestimation of 10 MW would result in 100% and 50% impact errors. Simply put, the larger the impacts the smaller the impact error is given a fixed baseline error.

Appendix C Detailed Baseline Accuracy Results

Table C-2 shows the demand reduction estimates produced by the different baseline rules and compares them to the known reductions simulated in each proxy event-day for class A customers. Table C-3 is the equivalent table for class B customers. It also summarizes the observed bias and accuracy of each of the baselines tested. The mean percent error (MPE) indicates whether a baseline has a tendency towards over or underestimating the known demand reductions (bias). A value of +10% indicates the baseline tends to overestimate the actual impacts by 10%. The mean absolute percentage error (MAPE) measures how closely the estimates match the actual values (precision or accuracy), regardless of the direction of the errors. It can be interpreted as the typical magnitude of the errors for individual events. For example, a baseline with a high MAPE value, but low mean percent error, would be considered inaccurate but unbiased. Table C-1 illustrates the differences in bias and precision with hypothetical examples. Ideally a baseline is both unbiased and precise.

Table C-1: An Illustrative Example of Bias and Accuracy

Proxy Event	Actual Reduction	Unbiased but Imprecise	Unbiased and Precise	Biased and Imprecise	Biased but Precise
1	100	150	95	50	95
2	100	50	105	100	95
3	100	150	95	50	95
4	100	50	105	100	95
Average	**100**	**100**	**100**	**75**	**95**
Bias (MPE)		**0%**	**0%**	**-25%**	**5%**
Goodness-of-fit (MAPE)		**50%**	**5%**	**25%**	**5%**

The current baseline shows the highest amount of bias (MPE) and the least accuracy among the baselines tested for both customer groups. On average, the settlement baseline would estimate 100 MW when in fact 78 MW are being delivered (a 22% upward bias) for GAM eligible customers and 77 MW for non GAM High-5 Customers. It is also the least accurate and has the largest error for individual proxy event-days. Baselines that rely on the past 10 eligible days (excluding weekends, holidays and event-days) are more accurate (MAPE) and exhibit less bias (MPE). Modifying the baseline will better align the demand reduction estimates used for settlement to actual reductions and it will also help reduce under and over payments to participants.

Table C-2: Comparison of Baseline Accuracy, Class A Customers

Proxy Event-day	Simulated Reduction	Highest 15-of-20 (Current)	10-of-10 Without Adjustment	10-of-10 With a 3-hour Adjustment Window & 1.2 Cap	10-of-10 With a 6-hour Adjustment Window & 1.2 Cap
2/8/2010	213.5	264.6	223.3	230.1	227.2
5/26/2010	209.1	231.9	195.7	200.2	204.3
7/15/2010	243.9	380.1	310.8	286.5	293.6
7/20/2010	243.9	307.1	241.0	228.5	227.9
7/26/2010	215.5	258.6	192.4	214.4	208.3
7/27/2010	215.5	249.0	183.4	194.8	196.5
8/5/2010	216.0	242.8	195.1	212.4	207.1
8/10/2010	216.0	227.2	178.8	191.8	190.3
8/30/2010	216.0	267.7	222.3	233.1	217.8
8/31/2010	216.0	278.2	233.2	258.3	250.1
1/6/2011	207.9	220.1	157.1	162.5	162.2
1/11/2011	207.9	204.7	154.4	155.5	155.2
3/23/2011	279.3	305.4	268.4	276.7	280.4
6/28/2011	268.4	300.0	257.1	255.8	262.3
7/15/2011	274.0	347.5	265.3	289.1	289.0
7/28/2011	274.0	375.1	279.9	289.3	293.2
8/17/2011	274.0	373.5	340.6	321.2	321.5
8/19/2011	274.0	333.1	301.9	308.5	302.9
9/1/2011	274.0	255.0	218.7	243.4	229.9
9/12/2011	274.0	327.8	290.2	299.9	296.9
9/27/2011	274.0	332.7	255.9	267.0	272.9
1/4/2012	212.8	259.9	202.0	198.6	202.9
1/20/2012	186.4	210.0	180.7	195.7	197.2
5/25/2012	269.3	359.6	327.2	336.0	325.9
8/2/2012	253.0	366.0	321.5	309.7	312.1
11/13/2012	234.3	309.5	260.9	268.2	272.1
Average	240.1	291.8	240.7	247.2	246.2
Bias (Mean Percent Error)		**22%**	**0%**	**3%**	**3%**
Goodness-of-fit (Mean Absolute Percentage Error)		**21%**	**0%**	**2%**	**2%**

Table C-3: Comparison of Baseline Accuracy, Class B Customers

Proxy Event-day	Simulated Reduction	Highest 15-of-20 (Current)	10-of-10 Without Adjustment	10-of-10 with a 3-hour Adjustment Window & 1.2 Cap	10-of-10 with a 6-hour Adjustment Window & 1.2 Cap
2/8/2010	49.6	66.0	60.2	54.4	55.2
5/26/2010	53.6	50.6	45.9	52.6	52.7
7/15/2010	63.6	70.6	53.6	61.7	62.7
7/20/2010	63.6	77.6	59.9	64.4	65.0
7/26/2010	63.6	84.2	71.1	68.3	65.7
7/27/2010	63.6	79.9	66.2	69.0	66.1
8/5/2010	64.7	70.5	58.6	65.8	64.0
8/10/2010	64.4	60.4	47.7	59.0	57.0
8/30/2010	65.1	72.2	66.6	66.4	64.9
8/31/2010	65.1	89.2	82.9	88.3	88.1
1/14/2011	63.3	80.2	67.7	75.9	76.5
2/4/2011	68.3	94.1	78.7	78.3	78.6
7/6/2011	72.1	86.3	74.4	73.8	75.4
7/12/2011	72.1	85.0	65.5	63.6	63.0
7/18/2011	72.1	92.6	75.1	80.5	79.5
7/19/2011	72.1	87.3	71.6	79.1	78.8
7/20/2011	72.1	88.5	71.1	80.0	79.7
8/8/2011	72.6	85.5	73.1	72.6	74.6
8/17/2011	72.6	83.4	74.8	79.5	78.2
8/18/2011	72.6	81.4	73.8	77.5	76.8
9/13/2011	76.9	94.3	77.2	85.7	86.6
5/25/2012	86.1	95.5	84.1	90.9	92.7
7/4/2012	87.0	111.2	87.6	97.6	96.5
7/5/2012	87.0	117.3	93.7	100.9	102.3
7/6/2012	87.0	136.6	112.6	123.4	123.9
7/16/2012	85.0	109.8	83.9	96.9	95.0
Average	70.6	86.6	72.2	77.2	76.9
Bias (Mean Percent Error)		**23%**	**2%**	**9%**	**9%**
Goodness-of-Fit (Mean Absolute Percentage Error)		**22%**	**2%**	**9%**	**8%**

Appendix D Glossary

Term	Definition
Contracted Load Reduction Capability (Contracted MW)	Load reduction amounts customers contract to provide. In the absence of nonperformance, this is the amount by which they will aim to reduce load.
DR	Demand Response programs provide load reductions or load shifting for specified hours in exchange for incentives. They include both dispatchable and nondispatchable programs.
DR-1	OPA's voluntary demand bidding program.
DR-2	OPA's permanent load shifting program.
DR-3	OPA's event-based load reduction program.
Event Window	Consecutive hours when a program provides load reduction. In DR-1, event-days and hours are determined by comparing strike prices to the pre-dispatch price. In DR-3 it is determined ahead of time by the IESO based on the day-ahead supply cushion. For DR-2, the event window is based on load shifting contracts with participants.
***Ex ante* Load Reductions**	Load reduction capability under a standard set of conditions. For contractual resources, the *ex ante* impacts factor in historical scheduled nonperformance and historical deviations between day-ahead committed resources and actual load reductions delivered by participants.
***Ex Post* Load Impacts**	Actual load impacts attributable to a demand response resource for a given year, but do not necessarily reflect the load reduction *capability* of the DR initiative. Effectively, historical *ex post* results are tied to specific conditions that occurred for that given event, including weather conditions, market prices and the number of participants dispatched through the DR program.
Initiative	A Conservation & Demand Management (CDM) offering focusing on a particular opportunity or customer end-use.
Gross Verified Energy/Demand Savings	Verified annual energy/demand savings that results directly from program-related actions taken by participants in a CDM initiative, regardless of why they participated.
Gross Verified Lifetime Energy Savings	Summation of the verified gross energy savings that persist over the effective useful life of all measures associated with the implementation of a CDM initiative, regardless of why the measures were implemented.
Levelized Delivery Cost	A metric that expresses delivery costs per unit of energy saved on an annualized basis, expressed in $/MWh, taking into account the lifetime energy savings associated with the implementation of a CDM initiative.
Load Reductions	Changes in load attributable to participation in a given demand response program.
Load Reductions For Operations	Load reduction estimates used for program operations. The estimates factor in historical differences between scheduled and delivered resources.

Term	Definition
Load Reductions For Planning	This term is synonymous with *ex ante* load reductions. They provide load reduction capability under a standard set of conditions and are recommended for planning and cost effectiveness analysis. The estimates factor in historical scheduled nonperformance and differences between scheduled and delivered load reductions.
Load Shifting	Substitution of electric load from one period to another.
Net to Gross Ratio	The ratio of net savings to gross savings, which takes into account free-ridership and spillover.
Net Verified Annual Energy/Demand Savings	Verified annual energy savings directly attributable to a CDM initiative, taking into consideration the net to gross ratio.
Net Verified Lifetime Energy Savings (MW)	Summation of the net verified energy savings that persist over the effective useful life of all measures directly attributable to a CDM initiative, taking into consideration the net to gross ratio.
Program Administrator Cost (PAC)	Cost test that measures benefits and costs from the perspective of a program administrator.
Program Realization Rate	A comparison of observed or measured (evaluated) information to original reported savings which is used to adjust the gross savings estimates.
Scheduled Nonperformance	Planned periods when customers will not participate in demand response programs. It is synonymous with planned nonperformance. The system operator is informed in advance of scheduled nonperformance. Scheduled facility shutdowns for maintenance are an example.
Settlement Baseline	The method for determining payment for participants. Settlement baselines produce load reduction estimates that may differ from load reductions estimated in impact evaluations. They are less complex and are designed to quick-settle payments. Importantly, if the settlement baseline is biased, participants may still be in compliance for settlement but actually deliver less load reduction than contracted.
Scheduled Resources	Reflects the amount of load reduction expected for the event and factors in planned nonperformance by program participants.
Shift Window	Consecutive hours during which customers participating in load shifting programs will reduce load, shifting it to hours outside the shift window.
Total Resource Cost (TRC)	Cost test that measures benefits and costs from a societal perspective.
Unscheduled Nonperformance	Deviations between day-ahead contracted resources and actual load reductions delivered. The gap between the committed resource and actual load reductions can be due to multiple factors including biased settlement baselines and underperformance by participants.