Measurement and Verification: Making AMI Data Smart for Demand Response

Will Gifford, DNV GL, Arlington, VA

Abstract

The continued development of demand response (DR) in European electricity markets requires the adoption of robust measurement and verification (M&V) practices for valuing its performance as a resource. While the benefits of DR as a capacity and energy resource are becoming widely recognized and accepted, these are often seen as being equaled or outweighed by the risks associated with the associated performance uncertainty when called upon. Robust M&V practices for DR reduce this uncertainty and thereby strengthen the market and program participants' confidence that DR performance can be fairly valued both prospectively and retrospectively. This is especially critical as highly granular interval consumption data through Advanced Metering Infrastructure (AMI) becomes more widely available. M&V protocols for DR have been in place for several years in the U.S. to establish standards for program impact evaluations and settlement calculations. State commissions, wholesale electricity markets, the federal government, and other organizations in the U.S. have published M&V protocols for a variety of applications. This paper presents an overview of M&V methodologies for a number of these applications and translates them into a European context. We draw on experiences and lessons from the U.S., especially those related to M&V methods in various applications adapting with the increased availability of customer and end-use interval consumption data. We demonstrate that the role of M&V is not diminished in the context of smart meter saturation of a system, but rather may represent a shift in methodology while maintaining a target of reducing the uncertainty of DR performance.

Introduction

This paper summarizes the recommended practices for M&V of DR resources and provides guidance based on these best practices for M&V to support the current and future roles for DR in Europe. The primary source for the best practices is *Measurement and Verification for Demand Response*, prepared by ML Goldberg and GK Agnew of DNV GL for the Measurement and Verification Working Group of the National Action Plan on Demand Response (a U.S. federal working group). This reference document, published in 2013, provides guidance on various methods of M&V of DR for settlement and impact estimation, which are described as follows (Goldberg & Agnew 2013, vi):

Settlement. Determination of the demand reductions achieved by individual program or market participants, and of the corresponding financial payments or penalties owed to or from each participant.

Impact Estimation. Determination of program-level demand reduction that has been achieved or is projected to be achieved, used for ongoing program valuation and planning.

There are other valuable M&V protocol and guidance documents for DR sponsored by U.S. state agencies which may be useful alongside *Measurement and Verification for Demand Response*, including

- *Load Impact Estimation for Demand Response: Protocols and Regulatory Guidance* (2008) produced for the California Public Utilities Commission Energy Division, which provides detailed descriptions and methods for impact estimation, and

- *Protocol Development for Demand Response Calculation – Findings and Recommendations* (2003), produced for the California Energy Commission, which provides an overview of baseline approaches and benchmarks the accuracy of many baseline methods which have been used in settlement and impact estimation.

While these and other DR evaluation guidance documents are excellent resources, the primary reference for this paper is *Measurement and Verification for Demand Response* for reasons which include the following:

- Its publication in 2013 accounts for recent issues and developments in M&V for settlement and impact estimation.
- It gives approximately equal treatment to M&V for wholesale settlement purposes and impact estimation for program evaluation.
- The perspective is national rather than for a specific state or region.

The primary source for references on the European market for DR is *Mapping Demand Response in Europe Today; Tracking Compliance with Article 15.8 of the Energy Efficiency Directive* (2014), produced by Smart Energy Demand Coalition (SEDC). This report is both comprehensive and current with respect to the development of DR in Europe.

Key Differences between the U.S. and Europe with Respect to Demand Response

There are important differences between the U.S. and Europe related to DR, including:

1. **Program administration.** In the U.S., demand response programs have been implemented for several decades. In the early days of DR, a utility owning generation resources, transmission lines, and distribution wires could offer an incentive to customers to reduce demand on request (communication could consist of a phone call), or through direct load control (DLC). The utility and the participants were the beneficiaries of DR programs. This load management foundation helped to establish a foothold for DR in the U.S., where the value of DR was a real resource for utilities with capacity issues. DR is not currently as well-established in Europe, and the landscape of players is inherently more complex than the U.S. had in the early days of DR. These players include: customers, retailers, Transmission System Operator (TSO), Distribution Network Operators (DNO), Balance Responsible Parties (BRP), and DR Aggregators. This system with several groups of stakeholders is similar certain regions of the U.S. today where DR is thriving. But considering the early stage of DR in Europe, it adds complexity and slows the adoption process because currently not all parties are effectively coordinated towards a common end goal.

2. **Rates.** While residential time-of-use rates[1] are still in a nascent state in the U.S., they are well-established in Europe. This variable rate structure already offers customers some level of incentive to reduce electricity consumption during peak periods. While residential time differentiated rates programs in the U.S. have been growing in recent years, DLC and other incentive based programs continue to be a means of circumventing the relatively static rate structure to shave peak load when needed.

3. **Regulation.** The European Commission Energy Efficiency Directive, Article 15.8 requires member state regulators and Transmission System Operators (TSOs) to allow consumer access to markets for DR, to enable participation by aggregators and to enable and encourage

1 Time-of-use rates give the cost of consuming a unit of energy in a given time period, where the cost for a given unit (e.g. kWh) varies throughout the day, as opposed to time invariant rates which stay the same throughout the day.

program development. However, by 2014 only Belgium, Great Britain, Finland, France, Ireland, and Switzerland had reached a level where DR was a viable commercial offering (SEDC, 4). In the U.S., on the other hand, there are national and state level regulatory frameworks mandating or facilitating DR to serve as an alternative to fossil fuel generation resources for capacity, energy, and ancillary services when cost effective. Additionally, wholesale markets recognize the value of DR resources aggregated from residential or small commercial customers, offering market programs and protocols incentivizing participation.

The role of M&V in DR can vary depending on the specific contexts, but universally its purpose is to mitigate the risk of imbalance in the costs and benefits of the resources among the stakeholders which are impacted. The principles outlined below for M&V of DR resources will apply to programs in both the U.S. and Europe.

Overview of DR M&V Methodologies

Measurement and verification methodologies for DR resources are used for financial settlement and load impact estimation. Virtually all program administrators benefit from impact estimates of the program resource, and they may be required by markets to verify the accuracy of a nominated resource prior to financial settlement for its market value. For example, a DLC program consisting of an aggregation of residential customer participants may be required to submit impact estimates for the program at summer peak with a prescribed level of statistical precision to be eligible to bid the resource in a capacity market auction. The M&V requirements for settlement of this DLC resource, however, may simply consist of verification data demonstrating that the program dispatch signal was deployed to the participant population during capacity events.

Impact estimation can be conducted after DR events or on close to a real-time basis. For program resources which are monitored in near real-time, telemetry systems utilize measurement equipment tied to impact estimation algorithms which feed into communications channels back to system operators. Ancillary services programs have telemetry requirements so that the system operator can monitor the dispatchable load resource in near real-time. Telemetry requirements provide detail about the frequency of readings from the end-use resource providing DR which are then communicated remotely through a communication channel such as a cellular network.

Impact Measurement for Settlement

Table 1 summarizes of the M&V needs for the main types of DR programs in the U.S., from the perspective of both participating customer to aggregator, and market participating aggregator to the wholesale market. Note that in the U.S. the aggregator is often the electric utility retailer. A third party aggregator may also act as a turnkey service provider for the utility. A utility may then decide to offer the program resource managed by the third party aggregator in markets.

Capacity and ancillary services are the most common uses for DR programs. Demand resources used as capacity services are obligated to be available for dispatch by the system operator over a defined period of time (FERC 2012). For ancillary services, which provide fast response support to provide reliable transmission of electricity to customer loads (FERC 2012) the resource is also paid to be on standby for dispatch by the system operator, but there are more stringent requirements on the time required for the resource to ramp up to the expected level of load reduction. Resource telemetry is also generally required of ancillary services.

Table 1. M&V Needs for DR Programs Integrated into Wholesale Markets in the U.S.

<table>
<tr><th>DR Program Type</th><th>Program Description</th><th>Typical Market Program Type</th><th>M&V Needs - Aggregator to Participant</th><th>M&V Needs - Market to Aggregator</th></tr>
<tr><td>Demand Bidding and Buy-Back, Peak Time Rebates</td><td>Participant is paid for reduced demand during an event compared to a baseline</td><td>Capacity</td><td>Individual baselines for program participants for measuring load reductions. Hourly premise interval load data is usually required.</td><td rowspan="3">Baselines for measuring aggregate load reduction during events, and models for prospective impacts of the programs under different conditions.

Hourly interval data is usually required either for full population or from a statistically designed sample of participants. Aggregated premise level data is preferred by market, but end-use data may be used for some programs.</td></tr>
<tr><td>Direct Load Control</td><td>Participant is paid an annual incentive for allowing certain end-use equipment to be shut off or cycled during events</td><td>Capacity</td><td>Verification that the participant did not override the curtailment signal. Hourly premise or end-use interval load data is usually required.</td></tr>
<tr><td>Interruptible Load</td><td>Participant is compensated for reducing load to a designated baseline during events or pays a penalty for some level of deviation from baseline</td><td>Capacity or Energy</td><td>Measured deviation from the designated baseline for each participant. Hourly premise interval load data is usually required.</td></tr>
<tr><td>Spinning Reserves</td><td>Program to provide a resource for energy supply and imbalance within a few minutes of dispatch</td><td>Ancillary Service</td><td rowspan="2">Telemetry for the dispatchable resource in near real-time, and metering for settlement following events. Data interval requirements are sub-hourly.</td><td rowspan="2">Telemetry for the dispatchable resource in near real-time, and metering for settlement following events. This may apply for the full population of program participants or for a statistical sample designed for</td></tr>
<tr><td>Non-Spinning Reserves</td><td>Program to provide a resource for energy supply and imbalance within 10 to 30 minutes of dispatch</td><td>Ancillary Service</td></tr>
</table>

Regulation	Program which allows a system operator to increase or decrease load for the resource on a real-time basis during a designated period	Ancillary Service		acceptable levels of precision. Data interval requirements are sub-hourly.
Time-Based Rate	Program where customers on the rate/tariff are charged different prices during times of the day to reflect the cost of generation and delivery	Not typically offered in market programs	N/A	N/A

Accurate baselines are critical for credible DR resource valuation through settlement, and they also must balance accuracy, flexibility, simplicity/comprehensibility, and reproducibility. (Goldberg & Agnew 2013, 32) This can be challenging considering the wide variety of load profile characteristics of participating customers.

If settlement baseline methods are to be assigned based on customer type, the assignment is most effective if it is based on observable load characteristics and customer rate class2, rather than on a reported business category or customer segment. Key qualities that can be determined from the customer's load data include weather sensitivity, seasonality unrelated to weather, and variability unrelated to season or weather (Goldberg & Agnew 2013, 35).

Weather Sensitive Loads. For weather sensitive loads, baseline methods based on rolling averages should use days in the baseline that are of the same season and have similar weather, and use an additive adjustment. A day-of-event additive adjustment scales a baseline up or down by a fixed amount to match the actual load in a designated adjustment window, prior to the event. For example, suppose an event is called at 15:00 on a given day, and an adjustment window of 12:00 – 14:00 was designated. If the baseline load was 1 MW lower than the actual load for the participant, the additive adjustment would increase the baseline load profile by 1 MW throughout the day. This mitigates the risk of the resource from this participant from being undervalued in settlement calculation; with rationale that prior to the event notification the baseline and actual load profiles should be the same. In the opposite example case where the event day load is above the baseline, the baseline should be adjusted up to meet the actual load profile to mitigate the risk of the resource from this participant being overvalued. Aggregator programs that only use a positive day-of-event adjustment at the participant level and have program impacts summed across the participants can severely overstate the impact of the overall program.

Seasonality Unrelated to Weather. This property could be exhibited for the seasons of the year in the literal sense, where loads follow similar patterns within each season. On the other hand, there could be other repeating patterns in the data, such as for days of the week. For example, a load profile could be very different on Fridays from other weekdays. It can be difficult to identify complex seasonality in customer load data used for settlement impacts.

Highly Variable Loads. For loads that may be volatile in the designated pre-event adjustment window, but exhibit a strong linear relationship with temperature and/or humidity, an

2 A rate class is a set of customers who pay the same cost per unit of electricity consumed (e.g. per kWh) and other costs per unit of peak demand (i.e. per kW).

additive adjustment based on the corresponding temperatures from the baseline and actual event days may be used. The baseline temperature refers to the average of the temperatures on the days that feed into the baseline. For example, suppose a baseline is defined to be the average over three days, with average temperatures in a specified time window 35 ºC, 32 ºC, and 34 ºC, and on the event day the actual temperature in that time window was 36 ºC. The baseline load would be adjusted by the ratio of the actual temperature to the three-day average (36/33.67 = 1.07), so each load interval in the baseline would be increased by seven percent.

Loads that do not have clear relationships with weather or exhibit seasonality are challenging for programs and markets. Often these loads are significant, but the measured impacts do not seem to reflect actual reductions from a baseline, because the notion of a baseline for these loads does not hold. For these participants, who tend to be medium to large commercial and industrial customers, we recommend there be a standard of predictability that should be met. Engaging these kinds of customers directly may potentially offer insight that can help create a solution for all parties (Goldberg & Agnew 2013, 37).

Baselines. DR baselines are designed to represent what the event day load would have been, had the DR event not occurred. They can be produced to represent interval load for specific end user participants or for the aggregated load of the whole program. Baselines make use of seasonal patterns within the interval load series over time and external influences on the load.

The load for most customers will have similar hourly profiles over certain day types. For example, residential load profiles tend to follow similar shapes over weekdays where the peak occurs in one of a few hours of the day and the trough also in one of a few hours. The load shape may be a bit different on weekends and holidays than weekdays, and there may be other seasonal influences over the course of a year. The outside weather conditions have a great influence on the load as well. For commercial and industrial customers, there may be pronounced seasonality associated with the days of the week, and other factors which are tied to operations and production schedules. It is critical for all of the drivers of load to be accounted for in a baseline prescription for settlement.

It is beyond the scope of this paper to present all, or even many specific baseline methods for settlement used in U.S. markets currently. Appendix B of Goldberg & Agnew 2013 provides a detailed presentation of this.

Impact Estimation for Program Evaluation

Most DR programs are subject to program evaluation in the U.S., which serves the purpose of comparing program performance to its stated goals and providing actionable information for administrators to make adjustments to maintain or improve performance, or in some cases to end the program. Impact evaluations of DR programs typically use M&V methodologies that may be more complex than those used for settlement calculations, for the following reasons:

- Settlement calculations are generally more time sensitive than program evaluations.
- Savings for settlement calculations need to be prescriptive so that they are transparent to the relevant parties, whereas program evaluations can use methodologies that require statistical or engineering judgment based on experience.

Impact estimation consists of a comparison of load from program participants during DR events to baseline counterfactual load, representing the expected load for the program with all conditions the same except without an event being called. *Ex post* impact estimation looks at what happened in past events and *ex ante* estimation is concerned with impacts for upcoming events, usually considering varying conditions such as weather, day types, and participant characteristics. For many years, regression models capturing these relationships have been effective in the U.S. for program impact evaluation. In more recent years, randomized control trial (RCT) evaluation frameworks have been used for certain types of programs, especially those whose primary purpose is to provide capacity during system emergencies. For these types of programs, control groups of participants matched on certain characteristics can provide extremely accurate baselines. These programs do, however, require some portion of the program population to remain on the "sidelines" during events, or to utilize non-program participants for the baseline control group. An alternative is to match non-participants to program participants when smart meter data is widely available.

Role of Smart Meter Data

At the end of 2013, the penetration rate of smart meters was 22% in Europe with a target of 80% penetration by 2020.3 Smart meter data do not magically make DR participants out of non-participants. But when the meters are backed by suitable information technology so that the meters not only function for billing but also for analysis, it can open up new opportunities.

Residential DR is one such opportunity. Markets for DR usually require each site in the aggregated resource to be metered individually and additionally may require telemetry from each individual site in the aggregation. The cost of installing and operating this communication infrastructure for residential programs is a barrier to the program being cost effective since each individual household-level resource is quite small.

When Advanced Metering Infrastructure (AMI) is already in place, however, the barriers are reduced significantly. The key challenge becomes establishing the operational capabilities to process the metering data from the program participant population and adding any incremental technology and analytical components to the AMI infrastructure to enable telemetry at the required frequency. As penetration of smart meters grows, there becomes increased opportunity for aggregators to assemble programs consisting entirely of customers with smart meters.

The two-way communication capability of smart meters creates an opportunity which cuts across all customer sectors. Advanced metering infrastructure (AMI) systems which are not only capable of collecting data on an interval basis but also transmitting load data for processing at a frequent and fine grained level can be valuable for DR for reasons which include:

- Interval data can be processed through forecasting analytics and displayed as a "dashboard" of the dispatchable resource for system planners.
- Verification analytics can detect potential systematic errors in dispatch signals or direct load control equipment.

Smart meter communications channels can also be leveraged to provide enhanced event signal notifications and performance feedback, through display mediums such as in-home devices, smart phone applications, and internet gateways when properly connected through communications protocols. These systems can make for an improved DR engagement experience to boost participation and performance.

There are, however, new challenges that emerge concurrently with programs of large aggregations of residential customers. One is that the information technology (IT) systems of the aggregator and user of the DR resource must be paired with algorithms capable of computing load impacts from hourly data for settlement, representing perhaps hundreds of thousands of residential

3 N. Raven, *Pan European Networks: Government*, 10, http://www.paneuropeannetworks.com/GOV10/#150/z

customers who could be in a program. Telemetry dashboards can be complex to set up and may require significant IT equipment upgrades, and there are currently technical limitations to AMI communications systems that would prevent DR from being used for high frequency control applications such as regulation. Many systems may not currently be able to provide data at a sufficiently granular level to be used for spinning reserves applications which require resource ramping in less than 10 minutes. Another issue is that quality control procedures must be established for the processing of this data to help ensure the measured impacts are valid through such a large increase in scale.

Role of M&V of DR in Europe

Protocols for M&V in Europe are currently lacking or uncoordinated in many regions. In some cases TSOs will have one set of criteria, the BRP will have another, and the retailer will maintain a third (SEDC, 5). **Table 2** presents some needs for DR by types of entities that may be impacted by DR in Europe.

Table 2. DR M&V Needs of Various Stakeholders in European Markets

European Market Participant	**Market Role**	**Value of DR**	**M&V Needs**
Customers	End Users of electricity	DR offers direct compensation in return for having flexibility to shift or curtail load during designated periods	Facilitate fair compensation for performance to encourage participation in DR programs
Distribution Network Operator (DNO)	Manages the distribution grid, providing security of supply and maintaining power quality	Capacity resource in localized areas of distribution	None directly. Needs DR resources integrated into markets to be as reliable as supply based resources
Retailer	Supplies electricity to customers through contracts with suppliers	Can offer competitive advantage, as DR may be more cost effective in certain times than generation resources	None directly. Needs DR to not increase cost of providing service to customers over that of generation
Aggregator	Contracts with DR providers and sells the resource as a balancing or capacity service	Market value of the coordinated resource sold to markets	Critical for determining settlement with participating customers and for settlement of aggregate resource with markets
Generator	Supplier of electricity to the grid through contracts with retailers.	Competitor to DR resource providers	Ensure DR aggregators or direct participants are subject to the same metering and

			telemetry requirements as generators, and are subject to penalties for non-performance
Balance Responsible Party (BRP)	Resource scheduler to meet electricity demand with electricity supply	May use DR as a resource to schedule along with generation	Facilitate accurate telemetry so that scheduled resources meet expectations.
Transmission System Operator (TSO)	Manages the transmission grid, ensuring a means for adequate supply at the required power quality levels	Maintains balancing and capacity markets which DR resources can participate alongside generation resources.	Facilitate resource valuation for a fair and efficient market when both DR and generation resources are integrated

Conclusions

The following recommendations are offered concerning M&V of DR programs in Europe.

Program Design

Programs should be developed for European markets where M&V methods are established and well-suited for the specific customer segments that will be recruited for the program. This means that data must be available to demonstrate, through M&V analytics, that programs deliver the load resource that is being claimed.

Program Evaluation

There is an understandable focus on markets and settlement for DR programs in Europe, but regulatory bodies should also take ownership of establishing goals outside of market participation for programs. Program evaluations can support cost benefit analyses from several different perspectives, including the various types entities involved in the electric power industry, customers, and societies. Regulatory bodies which have established economic and sustainability goals for their jurisdictions involving DR programs should establish budgets for tracking the performance of these programs with respect to planned targets using appropriate M&V methodologies.

Role of smart meter data

Smart meter data can be valuable, but without adequate IT horsepower and a robust analytical plan for extracting relevant information from it, there is little intrinsic benefit of the data itself. European markets should both develop a roadmap for establishing the tools and algorithms for extracting the information which will make DR resources more precise and reliable over the next several years, as well as develop shorter term plans for adding value from smart meter data over the short term.

Baselines

Appropriate baselines are critical for the success of DR in markets. We recommend that baselines be developed through load research analyses designed to detect seasonality exhibited by the load time series for the breadth of customer types that may be part of a program. This work should be done both at the outset of program's entry to participating in markets and on an ongoing basis, in parallel with impact calculations for settlement. There are well-established analytical procedures used by wholesale markets in the U.S. which assess the accuracy and variability in performance of a broad set of candidate baselines on proxy event data for program participants. See CPUC 2008 or Goldberg & Agnew 2003 for examples of this type of baseline evaluation. Similar baseline evaluations should be replicated for European markets as baselines are typically not "one-size-fits-all", and there can be regional factors which make certain baselines better choices than others. On a more immediate basis, we recommend that European markets adopt settlement baselines that have proven to be robust in the U.S. or elsewhere for the types of customers which participate in current programs. These should at minimum have adjustments for weather-sensitive loads, known non-weather seasonality, and account for highly variable loads. Examples of existing prescriptive baseline methods for settlement in U.S. markets are presented in Appendix B of Goldberg & Agnew 2013.

References

California Public Utilities Commission (CPUC). 2008. *Load Impact Estimation for Demand Response: Protocols and Regulatory Guidance.* http://www.calmac.org/events/finaldecision_attachementa.pdf. San Francisco, Calif.: California Public Utilities Commission.

Federal Energy Regulatory Commission (FERC). 2012. *FERC-731 Demand Response/Time-Based Rate Programs and Advanced Metering, Glossary*. http://www.ferc.gov/industries/electric/indus-act/demand-response/2012/instructions.pdf. Washington, DC: Federal Energy Regulatory Commission.

Goldberg, Miriam L. and Gates K. Agnew, 2013. *Measurement and Verification for Demand Response*, prepared for the Measurement and Verification Working Group of the National Action Plan on Demand Response. This report was produced for the Lawrence Berkeley National Laboratory under a contract to the U.S. DOE Office of Electric Delivery and Energy Reliability. http://emp.lbl.gov/sites/all/files/napdr-measurement-and-verification.pdf. Madison, Wisc.: DNV GL.

Goldberg, Miriam L. and Gates K. Agnew, 2003. *Protocol Development for Demand Response Calculation – Findings and Recommendations*, prepared for the California Energy Commission (CEC). http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF. Madison, Wisc.: DNV GL.

Smart Energy Demand Coalition (SEDC). 2014. Mapping Demand Response in Europe Today; Tracking Compliance with Article 15.8 of the Energy Efficiency Directive. http://sedc-coalition.eu/wp-content/uploads/2014/04/SEDC-Mapping_DR_In_Europe-2014-04111.pdf. Brussels, Belgium: Smart Energy Demand Coalition