The PARITY project has received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 864319

Project Acronym:	**PARITY**
Project Full Title:	**Pro-sumer AwaRe, Transactive Markets for Valorization of Distributed flexibilITY enabled by Smart Energy Contracts**
Grant Agreement:	**846319**
Project Duration:	**42 months (01/10/2019 – 31/03/2023)**

DELIVERABLE D3.3
PARITY Framework Performance Measurement & Verification Methodology

Work Package:	**WP3 – Use Cases, Requirements and System Architecture Definition**
Task:	**T3.3 – PMV methodology definition**
Document Status:	**Final v2.0**
File Name:	**PARITY_D3.3_PARITY Framework Performance Measurement & Verification Methodology_R1_V1.0_CIRCE**
Due Date:	**30.09.2020**
Submission Date:	**05.10.2020 (v1), 25.06.2021 (v2)**
Lead Beneficiary:	**CIRCE**

Dissemination Level	
Public	X
Confidential, only for members of the Consortium (including the Commission Services)	

The PARITY project has received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 864319

Authors List

Leading Author				
First Name		Last Name	Beneficiary	Contact e-mail
Victor		Ballestin	CIRCE	vballestin@fcirce.es
Co-Author(s)				
#	First Name	Last Name	Beneficiary	Contact e-mail
1	Andrea	Conserva	CIRCE	circe@fcirce.es
2	Jorge	Rueda Quintanilla	CUERVA	jruedaq@grupocuerva.com
3	Alberto	Sánchez Pérez	CUERVA	asanchez@grupocuerva.com
4	Pablo	López	CUERVA	plopez@grupocuerva.com
5	Samuel	Wingstedt	CHECKWATT	samuel.wingstedt@checkwatt.se
6	Dan-Eric	Archer	CHECKWATT	daneric.archer@checkwatt.se
7	Cruz	Borges	DEUSTO	cruz.borges@deusto.es
8	Diego	Casado	DEUSTO	dcasado@deusto.es
9	Koldo	Zabaleta	DEUSTO	koldo.zabaleta@deusto.es
10	Ana	Villar	URBENER	avillar@urbener.com
11	Olga	Marco	URBENER	o.marco@urbener.com
12	Christos	Malavazos	HYPERTECH	c.malavazos@hypertech.gr
14	Stelios	Zikos	CERTH	szikos@iti.gr
15	Fotis	Manesis	BFS	fmanesis@konkat.gr
16	Iulia	Minda	EON	iulia.minda@eon.se
17	Simon	Stukelj	AEM	simon.stukelj@eon.se
18	Niccolò	Poretti	AEM	nporetti@aemsa.ch

Reviewers List

Reviewers			
First Name	Last Name	Beneficiary	Contact e-mail
Andrea Emilio	Rizzoli	SUPSI	andrea.rizzoli@idsia.ch
Davide	Strepparava	SUPSI	davide.stepparava@supsi.ch
Marco	Derboni	SUPSI	marco.derboni@idsia.ch
Charalampos	Stauropoulos	BFS	cstauropoulos@bfs-ae.gr

Version History

Version	Author	Date	Status
0.1	Andrea Conserva, CIRCE	June 26th, 2020	Initial draft (TOC)
0.2	Andrea Conserva, CIRCE	July 10th, 2020	State of the art
0.3	Victor Ballestin, CIRCE	September 9th, 2020	Final draft for internal review
1.0	Victor Ballestin, CIRCE	September 15th, 2020	Version including comments from partners, ready for submission
2.0	Victor Ballestin, CIRCE	June 25th, 2021	Updated version taking into account 1st review comments

Legal Disclaimer

The PARITY project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864319. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

Copyright

Executive Summary

This document is an outcome of task 3.3, designed to define the Performance Measurement and Verification Methodology (PMV) to be followed by PARITY to measure and verify the response of distributed energy resources to flexibility solicited by aggregators or DSO. The report provides information from literature research to establish the state of the art in these procedures. In particular, relevant well-established protocols such as the International Performance Measurement and Verification Protocol (IPMVP) and Federal Energy Management Programme (FEMP) protocols, the most used at international level for Measurement and Verification (M&V) projects have been analysed, and also previous EU projects on the subject such as eeMeasure, Moeebius, OrbEEt, HOLISDER and FLEXCoop.

Moreover, an analysis of baseline estimation methodologies in previous projects has also been performed to identify the main issues detected in previous projects in M&V in DR (provided in Annex). The key aspect to overcome these barriers is the definition of this baseline. Most issues encountered when developing a baseline are related to the selection of representative days as the basis for estimation avoiding non representative consumption and the definition of adjustments' types and windows. The PARITY models developed for each of the DR systems sort out this issue by providing a continuously automatic normalization and calibration of the baseline that uses data from the smallest number of recent days that yield high accuracy. Furthermore, since automated flexibility is considered, the models will register the timing of the signals in order to avoid ramp or transition periods in the system's demand in preparation for the event and discard them from the reference period. The models will take into account human actions and levels of occupancy.

The main result of this work is the PMV methodology, also taking in consideration the most common recommendations found for M&V in DR events. The methodology is composed by 3 phases and 9 total steps, from the definition of the events and systems affected to the PMV report.

PARITY develops further on the current state of the start of M&V methodologies applied in DR by assessing flexibility potential and impact for a wider range of systems and building characteristics. Besides, the methodology is applicable to different grid conditions are their consequent dispatch signals, adapting the systems involved, comfort conditions and remuneration to each scenario.

Furthermore, an analysis of the application of this methodology in the different use cases and grid conditions has been made, detailing potential DR systems associated, variables that affect demand and monitoring and actuators needed. In addition, a cost-benefit analysis is detailed for each of these scenarios.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Term	Description
VPP	Virtual Power Plant
M&V	Measurement and Verification
DR	Demand Response
EEM	Energy Efficiency Measure
IPMVP	International Performance Measurement and Verification Protocol
CBA	Cost-Benefit Analysis
HDD	Heating Degree Days
CDD	Cooling Degree Days
ESCO	Energy Service COmpany
PMV	Performance Measurement and Verification
EEM	Energy Efficiency Measure

1. INTRODUCTION

1.1 Scope and objectives of the deliverable

An inaccurate measurement of the participation of users on demand response events can result in one of two things: either the participation in the programme is overestimated, reducing the benefits of the programme, or underestimated, therefore lowering participation and the potential benefits of the system. It is crucial to have a transparent, simple and reliable M&V methodology in order to have a balanced and effective system. The PARITY PMV has been designed with these principles in mind.

1.2 Structure of the deliverable

Before presenting the PARITY PMV methodology, the different grid conditions considered in PARITY are presented in Section 2, since these conditions will determine the procedures and requirements of the DR events. Considering the insights previously gathered, Section 3 presents the 3 phases and 9 total steps that compose the PARITY PMV. In Section 4, the PMV specificities for each of the grid conditions, use cases and DR systems comprising PARITY are analysed. This section ponders the different monitoring and actuation needs, variables that affect calculations and end-user reporting adapted to each of the scenarios considered in the PARITY project. The main conclusions are summarised in Section 5. A self-contained analysis of the state of the art in M&V methodologies, considering both energy efficiency and DR, is provided in Annex 1. Both existing protocols and previous EU projects on the matter have been studied. Moreover, since baseline definition is the key aspect in any M&V methodology, the existing practices for baseline construction are presented in Annex 2, by focusing on the different approaches to historical data analysis, reference periods and adjustments used to improve accuracy, highlighting the most successful practices encountered.

1.3 Relation to other tasks and deliverables

The methodology is applied in each of the grid conditions and use cases defined in D3.1.

The reporting structure of the DR event M&V results will serve as an input for the impact assessment in Task 8.6.

2. PRE-ANALYSIS OF GRID CONDITIONS IN DEMO SITES

The assessment of demand reduction of end users' standard loads during a DR event requires the definition of a specific PMV methodology. This can apply to end-users that will participate to DR events by allocating the energy consumption at the most cost-effective times in an automated way, both in wholesale and balancing markets. In PARITY, the reduction of demand will be dependent on the participation in each of the defined status of the grid:

- Normal grid conditions (Green state) → under green state this configuration should allow P2P transactions to be executed within previously defined limits/constraints, where multiple actors can actively participate. In this scenario, the DR event will be conditioned by the available systems and their minimum comfort conditions in order to limit the impact on the end user.
  These comfort conditions will need to be adapted to the building's characteristics and, since end-users cannot always explicitly specify their comfort boundaries, this will be determined by service level agreements with a by-pass option.
- Critical grid conditions (Orange state) → when the priority is given to the DSO as a buyer of flexibility. P2P energy transactions are no longer allowed.
- Emergency grid conditions (Red state) → where the DSO assumes control. In this case, the minimum comfort conditions will not be considered.

3. DESIGN OF THE PARITY PMV

The definition of the PARITY PMV is required to provide a fair and accurate remuneration method for the assessment of consumers' response to DR events to future PARITY final users and aggregators.

The methodology proposed takes into consideration the methodologies and protocols applied in the last years, as shown in Annex 1, both in European projects and in the American energy markets. IPMVP and FEMP have the highest influence in the field of M&V protocols used to assess the impact of EEM. In the recent years, EU funded projects, such as OrbEEt and Moeebius, have combined these two methodologies to establish and implement a new hybrid method. In America, the North American Energy Standards Board (NAESB) has provided recommendations for M&V applied to DR with the following two main objectives: (a) to identify the most appropriate M&V methodologies for each type of DR event, in order to determine demand reduction quantities, and (b) to provide a common terminology for the definition of measurement methods and DR events.

In this case, thanks to the algorithms and forecasting models developed in the project, the users' actions and behaviour are modelled and can be predicted in a very accurate way, due to the continuous update of the model based on real-time data. To define these models, the baselines estimation methodologies for M&V procedures in previous projects have been analysed and presented in Annex 2. In addition, the calibration takes into account typical information from sensors (e.g. temperature, luminance, humidity, etc.) and also considers the users' feedback on actions undertaken by PARITY control system on the dwelling's systems participating to DR and automated control events providing a more solid basis for estimations. The implementation of this procedure avoids the need for control and evaluation groups of customers needed in other methodologies. The PARITY PMV has been structured in three main phases, as in FEMP and Moeebius methodologies. These are the Ex ante analysis, Implementation and Ex post assessment phases. Each of these phases is composed of three steps that are described next.

3.1 Ex ante analysis

a) Definition of Demand Response events and main factors for remuneration.

The aggregator/DSO has to define at which types of DR event the customer will potentially participate (normal, critical and emergency grid conditions), including also information about their frequency or foreseen schedule during a year or along the duration of the contract between costumer and aggregator. At the same time, also the remuneration information for each condition (i.e. if it will be done monthly, yearly and the unit price) and the time of event notification (e.g. 2 hours before the event, day before the event, etc.) has to be agreed. For the latter, despite PARITY solution provides automated response to DR events (without requiring users interaction), sending a notification to the users before the beginning of the event to inform them about the start of the DR event is not needed, but it is recommended in order to address potential issues related to transparency and user friendliness in PARITY models.

b) Definition of Demand Response toolkits and minimum comfort factors & conditions.

In this step, the electrical systems that will be used for participation in DR events should be defined based on the type of DR events and the pilot sites. The electrical systems that will be involved have to be audited in order to collect their most relevant information (e.g. type, nominal power, efficiency, etc.).

In case of participation in DR events in normal and/or critical grid conditions, for each type of use that will be affected by DR events, an agreement should be made between the aggregator and prosumers on minimum comfort conditions that must always be maintained, in order to avoid any possible discomfort. These comfort conditions will need to be adapted to the buildings characteristics. Since end-users cannot always explicitly specify their comfort boundaries (often driven by intrinsic behavioural factors) this will be realized through more intuitive service level agreements, also allowing the users to by-pass system automated control actions. The minimum acceptable comfort conditions that will be defined by the users and/or inferred by the PARITY comfort profiling engine will feed the PARITY model to optimize the consumptions as well as the demand reductions during DR events. In addition, since comfort conditions can vary along a year, PARITY models will update the initial parameters set by the users without affecting their comfort. This will be possible thanks to the users' reaction to automated actions undertaken by PARITY solution on dwelling's systems. This information will be collected by PARITY models that will automatically learn, which are the optimal comfort conditions at any time.

In case of participation in DR events in emergency grid conditions, the end-users have to specify potential schedule-related limitations and/or a maximum amount of annual DR events.

c) Definition and analysis of static and dynamic variables that affect the demand and that need to be measured.

All the variables that need to be monitored in order to apply the assessment of demand reduction according to the types of DR events and systems that provide a response, should be defined in this step. The variables will also be used for the generation and auto-calibration of PARITY forecast models and are typically related to interior and exterior climate conditions (e.g. temperature, luminance, humidity, etc.) and to user behaviour (e.g. level of occupancy, schedule of electrical equipment, etc.). Expected result after applying this step is the specification of a set of variables and of their dependencies with energy uses.

3.2 Implementation

d) Analysis of existing monitoring system and specification of metering points' and sensors' characteristics.

In this step, an evaluation of the monitoring system (if any) already installed in the building will be performed. The evaluation has foreseen the collection of information such as communication infrastructure and protocols, mode of transmission, installed devices and measured parameters. In case there are smart appliances installed, their characteristics have to be audited as well. After the collection of this information, the variables identified in the previous step as those that need to be monitored as well as the electrical systems that will participate to DR events will allow to specify the new monitoring system's characteristics (e.g. performance, accuracy, communication protocol, etc.). Different monitoring requirements can be needed depending on the participation in each different grid condition. For instance, in case of participation in emergency grid conditions it is expected that few variables are needed to be monitored and, consequently, less complex monitoring devices are required to be installed. Moreover, in this phase, the most appropriate location for each sensor should be defined.

e) Analysis of the technical and economic reliability of individual loads measurements.

In this step, the economic and technical reliability of the PARITY monitoring and control system installation should be assessed. This analysis has to be performed considering the audit realised in the previous steps as well as the definition of the monitoring system specifications (e.g. location of the sensors, communication protocol, etc.). Considering that in PARITY it is expected that the measurements of loads is individual (following in this sense a similar approach to Option B of IPMVP protocol), this step will provide relevant information to verify that PMV methodology can be implemented successfully. Thus, pre-identifying and addressing potential barriers that can arise during the PARITY solution implementation is the main objective of this step. With the aim of considering the application of the PMV in the three grid conditions, different economic reliability analysis should be performed according to the arrangements settled in the Step 1.a (i.e. different CBA scenarios have to be defined for each grid condition and DR event).

f) Conduct post-installation verification activities for algorithm calibration.

After the monitoring and control equipment has been installed, control of the system's operation status is required to check that all components operate as planned and to rectify any detected problem. Following this activity, a period for the calibration of the PARITY models for the HVAC, DHW and Artificial Lighting flexibility estimation, is needed before starting the participation in DR events.

Delving deeper into the PARITY models, concerning the HVAC systems flexibility, calibration of three core-models is prerequisite: (1) the thermal comfort model, which revolves around the automated learning of occupants comfort boundaries; (2) the space thermal model, that assimilates how the space (and broadly building) thermally behaves under certain conditions; and (3) the HVAC device model, that predicts the amount of electric energy consumed under specific HVAC control actions and space thermal conditions (e.g. space air temperature) of the building.

For the former, a black-box, classification model, that produces comfort boundaries based on the user actions on electric HVAC devices or its presence in a space under certain ambient conditions, is introduced. To estimate the comfort boundaries, the Gaussian Naïve Bayes method is applied, a supervised Machine Learning (ML) classification algorithm that follows the Bayes' theorem assuming conditional independence between the features. The notion is that the thermal comfort can be inferred and forecasted on the basis of prior knowledge of building occupancy, ambient conditions and control commands on heating units, e.g. temperature setpoints, ON/OFF, etc. The posterior probability is calculated based on the prior probabilities with the assumption of conditional independence of the features, which refers to the predictors.

For the space thermal modelling, the adoption of a grey-box model has been anticipated (although black-box models will also be investigated). The grey-box model structures are derived from resistance–capacitance (RC) networks analogue to electric circuits to describe the dynamics of the systems. Thereby the distributed thermal mass of the space is lumped to a discrete number of capacitances, depending on the model type. Research studies have concluded to a plethora of grey box model types of different orders. In the most simplified, 1^{st} order model, the entire thermal mass of the space is lumped to a single capacity. As such, no distinction can be made between the fast dynamics of the indoor air and the slow dynamics of the wall mass. The introduced space thermal model is based on a 2^{nd} order model that considers this difference by including a second capacity. As every grey-box model, the space thermal model consists of a set of continuous stochastic differential equations expressed in a state-space form together with an output equation:

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

where, $x(t)$ is the state vector of the dynamic system, $u(t)$ is a vector containing the measured inputs of the system and $y(t)$ the measured output of the system.

Following the grey-box model structure described above, a calibration – or identification in terms of state-space modelling – process is applied to estimate the values of the adjustable parameters that form the matrices A, B, C and D. Note that the identification process is implemented in a way that allows repeated evaluation and adjustment of these parameters at regular time intervals (i.e. every week). Besides the grey-box structure definition, historical/measured data of the inputs (u) and output (y) signals of the model is prerequisite. Pre-processing of that data precedes the identification procedure, where data-detrend is performed. Data-detrend facilitates the linear grey-box model estimation and addresses their failure to capture arbitrary differences between the input and output signal levels.

The HVAC modelling approach selection is determined by the HVAC type. Currently, two different approaches have been investigated to model (1) air-to-air, constant volume heat pumps; and (2) air conditioning split units with an inverter. For the former, density-based non-parametric clustering (e.g. DBSCAN) and other multilevel (non-parametric) classification methods are investigated, while for the latter, Gaussian Processes, a generic supervised learning method designed to solve a regression problem, is introduced. These algorithmic approaches have been selected due to their non-parametric nature. In contrast to parametric models that assume a fixed model structure and a finite set of parameters (e.g. polynomial functions or neural networks), thus bounding the expressiveness of the model, non-parametric models assume an infinite-dimensional vector of parameters, where the amount of information that can be captured from the model grows as the amount of data increases. In addition, non-parametric models allow us to use the same models in all the experimental setups, regardless of the specifics of the HVAC system, whereas in the case of parametric models, a laborious manual model selection process has to be performed. Note here that more advanced classification and/or regression modelling of components might have to be supported on an as-needed basis.

The artificial lighting flexibility estimation requires proper modelling (and calibration) of the visual comfort and artificial lighting devices. Similar to the thermal comfort, the visual comfort modelling relies on the Naïve Bayes method, where two main classes (comfort & discomfort) and a single feature (the Illuminance in Lux) are used. The artificial lighting devices are modelled as quadratic functions of the status (on, off) and the dimming level, that estimate the artificial lighting consumption values for different values of the device status and the dimming level. A polynomial regression problem is solved to identify (calibrate) the unknown parameters of the aforementioned quadratic function.

With respect to the electric water heaters flexibility, calibrated models for the domestic hot water demands forecasting and the thermal behaviour of tanked electric water is foreseen. For the former, regression methods for medium-term time series forecasting are investigated, while for the latter, a 1st order state space model is being developed, where a system identification process, identical to the one performed for space thermal model, is required.

After the short introduction to the PARITY modelling approaches above, it is becoming apparent that the models' calibration implies the need for (1) state-space models system identification; and (2) supervised machine learning, regression and classification, algorithms evaluation and testing.

Within PARITY, the common denominator and starting point for all models' performance evaluation and testing is to perform k-fold cross-validation, a widely used method for testing the performance of a prediction model to be identified. This technique separates the data into k sets: the training sets and the validation sets. Training sets are used to train the models, while the trained models are applied to the data of the validation sets to investigate their capability on predicting the output using as inputs data that have not been considered in the training phase. Within PARITY, 10-fold cross-validation is applied on data that are not partitioned randomly, but sequentially into 10 sets. The cross-validation procedure is particularly useful when models are identified based on real measured data. The main advantage of cross-validation procedure is its ability to mitigate the presence of potential outliers and to handle overfitting issues that may occur. It is worth mentioning here that a whole year period for the cross-validation would be desired, since it would streamline the training of dedicated models: a whole year dataset allows to quantify the seasonality impact on the models' accuracy.

After the parameters' identification of each model based on the cross-validation datasets, each model is validated by means of relevant indicators. For the regression models, two widely used metrics are used to quantify the identified model's performance: (1) the root mean square error ($RMSE$) and the level of fit (FIT). A well-identified model corresponds to low values of $RMSE$ for the validation set and high values of FIT for the training set. $RMSE$ is defined as follows:

$$RMSE = \left(\frac{1}{n} \sum_{k=1}^{n} \left(\widehat{Y_k} - Y_k \right)^2 \right)^{1/2}$$

where, n is the number of samples of the validation set and $\widehat{Y_k} - Y_k$ represents the difference between the output predicted by the data-driven model and the respective measured value. FIT is the normalized root mean square error ($NRMSE$) measure of how well the response of the model fits the training data, expressed as the percentage:

$$FIT = 100(1 - NRMSE)$$

where, $NRMSE = \left(\frac{1}{m} \sum_{i=1}^{m} \left(\frac{\hat{Y}_k - Y_k}{Y_k} \right)^2 \right)^{1/2}$ and m is the number of samples of the training set.

For the classification models' performance evaluation, the $F1 - score$ metric is used, calculated as follows:

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

where, $TP,\ FN$ and FP are the true positives, false negatives and false positives, respectively, as they are estimated and provided by the confusion matrix, used to describe the performance of a classification problem on the validation-sets, for which the true values are known. The $F1 - score$ values range from 0 to 1, with 0 and 1 being the worst and best value, respectively.

Having achieved acceptable values[1] of the aforementioned metrics for the cross-validation datasets, the initial versions of the identified models are becoming available, however they are continuously self-calibrated further. In fact, the models are self-calibrated with measured data that monitor not only energy consumptions or interior conditions, but also users' behaviour. Depending on how much will vary the users' behaviours (and in general the variables that define the model) data from the last few days to few weeks, are used to increase the models' short-term prediction capability.

3.3 Ex Post analysis

g) Testing of the system in a DR event to validate model accuracy and reliability.

Once each one of the previously defined PARITY models is calibrated (Step 2.c), the prediction accuracy achieved by the model auto-calibration needs to be verified and validated in one or more DR test events (test phase) and under different grid conditions. At least one test should be carried out for each type of electrical system participating to the DR programme (i.e. those defined in step 1.b). During the test, the demand will be shifted by a predetermined value that then will be compared with the amount of demand flexibility estimated by the PARITY model. The main target at this step of the methodology is the overall performance evaluation of the calibrated model in terms of prediction accuracy but also to ensure that the developed model is reliable; that means the model should have the so-called "generalization capability" and can maintain the same levels of prediction accuracy under different ambient conditions and occupancy patterns, as well as for different device model brands (for example in the case of HVAC modelling, the same predictive model should capture with similar accuracy level the power consumption of a DAIKIN and a SIEMENS brand HVAC device). While existing guidelines and sources like IPMVP and ASHRAE Guideline 14, provide and recommend some quantitative requirements for accuracy between the model and the data used for its calibration, however, they do not provide a general means of evaluating model performance during the test phase. Considering the existing guidelines, the nature, the structure as well as the dependency of the calibrated models to real-time data acquired from the installed infrastructure, in PARITY PMV methodology the metrics that are proposed for the proper accuracy evaluation of all the developed models are: i) the coefficient of determination R^2, ii) the coefficient of variance of root mean squared error (CVRMSE), iii) mean bias error (MBE) and iv) the range normalized root mean squared error (RNRMSE). The aforementioned performance indicators is described and analysed below. In the following mathematical formulas, n corresponds to the number of samples of the data set used in the test phase, $\widehat{Y_k}$ corresponds to the predicted output of the PARITY model while Y_k stands for the respective measured (actual) values:

- Coefficient of determination (R^2) is a metric that indicates how well the prediction model fits the data during the test phase. Mathematically can be computed as follows:

[1] The performance metrics' acceptable values are highly affected by the cross-validation data completeness and quality; thus, indicative values will be defined in near future, after collecting and analysing the IoT data streams.

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(Y_k - \widehat{Y_k})^2}{\sum_{k=1}^{n}(Y_k - \mu)^2}$$

When the value R^2 is close to 1 then the model is perfect and there is a strong correlation between the model output and the data, whereas values closer to 0 indicate that there is no correlation between the model predictions and the measurements. Hence, a value of R^2 closer to 1 is desirable.

- Coefficient of variance of root mean squared error (CVRMSE) is derived by normalizing the root mean square error (RMSE) of the model predictions and it is a unit-less metric and it is preferred over RMSE because of it eliminates the dependency with the scale of data. The formula of this indicator is given by:

$$CVRMSE = \frac{\sqrt{\frac{\sum_{k=1}^{n}\left(Y_k - \widehat{Y_k}\right)^2}{n}}}{\mu} \times 100\%$$

In general values lower than 30% are acceptable but that depends of the specific model which is under evaluation and the noise of the actual values.

- Mean Bias Error (MBE) indicates how well the model predictions match the actual values. The mathematical representation which describes MBE metric is the following:

$$MBE = \frac{\sum_{k=1}^{n}(Y_k - \widehat{Y_k})}{\sum_{k=1}^{n} Y_k} \times 100\%$$

Values greater than 0 indicate that the model underpredicts the actual values while negative values suggest that the model overpredicts the actual values.

- Finally, the range normalized root mean squared error (RNRMSE) is a valid alternative, recently proposed in the bibliography for evaluation the accuracy of energy prediction models. It is a normalized form of RMSE in which the range (max(Y) – min(Y)) of actual (measured) values is used for the normalization. RNRMSE is defined as follows:

$$RNRMSE = \frac{\sqrt{\frac{\sum_{k=1}^{n}\left(Y_k - \widehat{Y_k}\right)^2}{n}}}{\max(Y) - \min(Y)} \times 100\%$$

The lower the value of RNRMSE metric is, the higher the accuracy of model prediction, while this indicator seems to provide more reliable estimates of the predictive performance in comparison to CVRMSE metric.

As a result, in order to conclude in a model that is characterized of high accuracy, the aforementioned statistical metrics should be computed and not exceed the desired limits. Thus, any deviation (representing the accuracy) between estimated demand flexibility and the actual/measured consumption is expected to be very low. Once the test phase is completed, the customer must be informed about the level of the model's accuracy and has to accept it to participate in the DR programme.

As regards the reliability assessment of the calibrated model, the methodology that is proposed to be followed in PARITY PMV is based on a sensitivity analysis concept. Sensitivity analysis corresponds to the methodology in which the uncertainty (standard deviation) of model predictions must remain stable and in low values compared to the mean of the prediction values.

The goal is to check and ensure that the calibrated model is robust. That means the model should maintain the prediction performance within the expected accuracy levels while the ambient or other external and/or interior conditions that could affect the demand flexibility present some unexpected fluctuations.

Apart from the performance evaluation of PARITY models in the test phase in terms of accuracy and reliability, at this point, another technical issue should be clarified and determined during the test phase. It is related to the reliability evaluation of the software and hardware infrastructure, but also to the specific tests that should be performed in order to ensure the validity of performance evaluation results. During the test phase, it is critically important that the following technical factors will be considered and validated:

- The correct and valid installation of all the hardware and software infrastructure,
- The status of each device or any other hardware component (e.g. sensors, gateways) that have effect on the accuracy and performance of the calibrated PARITY models,
- The proper communication and networking of all the components,
- The response time of the installed hardware devices with the main focus to be on the sensors and the actuators.

All the aforementioned factors are of paramount importance and must be considered during the test phase, as any small deviation or oversight could probably affect the performance of all the calibrated models.

h) Demand/generation flexibility assessment

PARITY models will be used for the assessment of demand flexibility. Based on recent historical data, they provide an estimation of the baseline that is continuously auto-calibrated and self-adjusted to guarantee high accuracy, as briefly presented below.

Within PARITY, the baseline and the upwards/downwards flexibility are estimated in the same manner by formulating human-centric optimization problems to be solved. For the purposes of participating in demand response, the optimization is first performed with comfort constraints, derived from the respective PARITY models (thermal, visual comfort and domestic hot water demand models), to estimate the baseline consumption, and with slightly loosened constraints so as to identify the available flexibility. For the sake of example, the baseline consumption over a day ahead time horizon is estimated by solving the following optimization problem.

$$\min_{control} \sum_{t=1}^{N} P_{HVAC}(t)$$

s.t. thermal comfort constraints at each timestep t,

HVAC constraints at each timestep t,

space thermal dynamics constraints at each timestep t.

$P_{HVAC}(t)$ refers to the HVAC power consumption at each timestep t. The thermal comfort, HVAC and space thermal dynamics constraints are specified by the homonymous self-calibrated PARITY models. At this point, should be clarified that the same procedure is to be followed for different time horizons according to the specific problem that is under investigation (i.e. for frequency regulation, intraday demand flexibility assessment etc.).

This way of baseline estimation follows the same philosophy of the approaches analysed in the Annex section. The main difference with this method is that the selection of the number of days prior to the DR event for baseline estimation is not needed, as it is performed in an automated way.

In addition, setting exclusion rules within the PMV process is not needed since the PARITY models automatically exclude outliers. The exclusion is performed not only to avoid considering values representative of extraordinary users' behaviour but also to exclude from baseline estimation, values of demand affected by the DR event. The exclusion of values from the ramp period (i.e. between notification and beginning of the DR event) can be made automatically by the PARITY solution. In fact, when it is installed, no actions are required from the users for demand shifting since both the preparation (e.g. for pre/post-heating/cooling) and the participation to the event are performed automatically. In this way, PARITY models are able to understand when measurements should not be considered for baseline construction being not representative. Thus, the assessment of demand flexibility can be made simply analysing the baseline estimated by the PARITY models without concerns about which period before the DR event should be selected for estimation (i.e. baseline windows) since it is optimised automatically by the model. Being this approach based on calibrated forecasting models, it is similar to the Option D of the IPMVP protocol, with the main difference, that in PARITY PMV, the energy loads are analysed individually and not at building/dwelling level. Furthermore, another main difference with all the M&V protocols defined so far is that the models used for the flexibility's assessment will also take in consideration under which of the three different grid conditions the participation of prosumers in DR events is happening. The consideration of this aspect it is strictly related to the remuneration process that will be performed under different criteria (previously defined in Step 1.a) according to the grid condition (e.g. the remuneration in emergency grid conditions will be different from that of normal grid conditions).

i) Definition of the PMV report

A PMV report will be issued for each prosumer after their participation in DR events or flexibility transactions. It will include the explanation of the demand reduction assessment made through the PARITY PMV. The detailed information that will be included in the report to be provided to the prosumer should be defined at this step of the methodology. Typical information included about the event is the type, schedule and duration, amount of reduced demand (kW or kWh), unitary price (€/kW or €/kWh), comfort conditions during the event (temperature, humidity, etc.), grid condition during the event (green, orange or red), remuneration information, increased amount of self-consumption rate, entity that requested the participation in the event (DSO or aggregator), etc. The report will be issued to the prosumer with a periodicity according to its preferences. For example, it could be event-based (at the end of each DR event) or periodic on a weekly/monthly/yearly basis. In general, sending remuneration information frequently on a regular basis should guarantee higher transparency.

4. PARITY PMV APPLICATION

The PARITY project comprises a wide variety of potential DR systems and grid states. This section addresses the specificities that the PMV methodology will have to address in each of the DR systems and use case scenarios defined in D3.1.

This analysis will be performed for each of the phases in the methodology. The Use Cases defined in the project are listed in Table 2.

Table 1. Definition of Use Cases

UC (initial)	Use Case title
UC1	Building-level P2H/BAB flexibility estimation & automated provision to aggregator for LFM participation
UC2	Aggregated P2H flexibility estimation and provisioning for market participation pre-qualification
UC3	EV profiling and aggregated EV flexibility estimation for market participation
UC4	Human-centric and contract-safeguarding energy and flexibility transactions in LFM, on the basis of context-aware flexibility profiles

UC5	Forecasting, scheduling and dispatch of DER flexibility for coordinated management of the LFM grid
UC6	Smart grid management using enhanced PQ services for grid instability limitation
UC7	Ancillary services provision by STATCOM to TSO for overlay network stability
UC8	Congestion management by DSO through operation of LFM to increase DER penetration
UC9	Provision of ancillary services to overlay ancillary service market operated by TSO
UC10	Participation of LFM-enabled flexibility to national wholesale energy market
UC11	Red light (emergency) grid management using automated control of distributed DER (through smart contracts)
UC12	Flexibility enhancement through synergies with neighbour distribution networks and/or LFMs.

1) **Ex ante analysis**

This step of the process determines the DR events in which the user will participate (normal, critical or emergency conditions) and the potential DR systems available for the response.

These are the grid conditions and DR systems related to each of the use cases:

Table 2. Ex-ante analysis specifications: grid conditions and systems per UC

Use Case	Grid conditions	DR systems involved
UC 1	Normal/ Critical/ Emergency	P2H: Heat pumps / HVACs
		P2H: Water heaters
		Battery Energy Storage System
		Lights
		PV
UC 2	Normal/ Critical/ Emergency	P2H: Heat pumps / HVACs
		P2H: Water heaters
		Battery Energy Storage System
		Lights
		PV
UC 3	Normal/ Critical/ Emergency	EV charger
UC 4	Normal/ Critical	P2H: Heat pumps / HVACs
		P2H: Water heaters
		Battery Energy Storage System
		Lights
		PV
		EV charger
UC 5	Normal	PV
		EV Charging
		Battery Energy Storage System
		P2H: Heat pumps / HVACs
		P2H: Water heaters
UC 6	Critical/Emergency	PV
		EV charger
		Battery Energy Storage System
		P2H: Heat pumps / HVACs

		P2H: Water heaters
UC 7	Normal	STATCOM
UC 8	Critical	PV
		EV charger
		Battery Energy Storage System
		P2H: Heat pumps / HVACs
		P2H: Water heaters
UC 9	Normal/ Critical/ Emergency	PV
		EV charger
		Battery Energy Storage System
UC 10	Normal/ Critical/ Emergency	PV
		EV charger
		Battery Energy Storage System
UC 11	Emergency	PV
		EV charger
		Battery Energy Storage System
		P2H: Heat pumps / HVACs
		P2H: Water heaters
		Lights
UC 12	Normal/ Critical	PV
		EV charger
		Battery Energy Storage System

As stated in steps b), the minimum comfort conditions of each of these systems will need to be defined to limit the demand response event in normal and critical grid conditions. Besides, the variables that affect the demand need to be accounted for as step c) reflects. Table 4 reflects these matters.

Table 3. Ex-ante analysis specifications: minimum comfort conditions and variables per system

DR system	Minimum comfort conditions	Variables that affect demand
Battery Energy Storage System	No specific comfort level for the occupant. The battery wear can only be described as a "minimum comfort condition"	Battery condition, peak demand, flexibility sales, management for increased self-consumption of PV
EV charger	EV owner's comfort boundaries to be set (e.g. minimum SoC after charging). Depending on one's usage pattern and willingness to sell flexibility, different levels of charging predictability will be acceptable.	Charger power, EV model / battery characteristics, EV state of charge (SoC), EV minimum-maximum preferred SoC, EV owners profile schedules and usage patterns, dynamic electricity pricing, geo-location (number of EVs serviced)
Lights	Luminance level set by the user or automatically selected for visual comfort based on the type of space	Weather conditions, occupancy and usage patterns, time of the day
P2H: Heat pumps / HVACs	Individual temperature comfort levels inferred from users' actions. For defining minimum comfort levels, user's preferences or User Comfort standards (ASHRAE 55, ISO 7730, EN15251) can be used.	Thermal behaviour of building, Occupancy, Weather conditions (external temperature, humidity), set-point temperature
P2H: Water heaters	Minimum preferred water temperature entered by the user, or based on comfort standards	Water heater capacity, Rated and actual power, Set-point temperature, Occupancy, Activity patterns
PV	N/A	Specifications, Irradiation, temperature

DR system	Minimum comfort conditions	Variables that affect demand
STATCOM	N/A	Local phase imbalances, active/reactive power balance forecast

2) Implementation

This phase of the process relies on the monitoring and actuation capabilities available in each system for each use case and grid condition.

In general, for each of the systems we have to account for these characteristics to determine their availability for DR:

- Response volume, availability period and duration
- Up or down response times
- Minimum baseline period between consecutive activations
- Frequency / number of activations possible time
- Potential seasonal limitations

The following table presents the monitoring needs for each of the DR systems considered.

Table 4. Implementation specifications: monitoring needs per system

DR system	Monitoring needs
Battery Energy Storage System	Charge level. Instantaneous power tracking
EV charger	Remote signal of battery SoC
Lights	Occupancy through IR-motion sensor, photosensors for natural/available lighting.
P2H: Heat pumps / HVACs	External temperature sensor and device status. Depending on the supplier, certain information on the operation of the heat pump is already visible on the device. The external sensor is standard, while the ability to vary the temperature in each individual room is optional.
P2H: Water heaters	Water Tank temperature sensor and water volume.
PV	Electrical production in real time provided by the inverter, additional values can be obtained by additional solutions offered by manufacturers (sensors/sensor cards, ...) for system monitoring. External sensors can be weather stations (humidity, wind speed/direction), temperature sensors (PV module, ambient temperature) and irradiation sensors/pyranometers for forecast in combination with online weather services
STATCOM	State of the grid through three phase VTs/CTs, capable of measuring active and reactive power, phase voltage as well as harmonics

Once the DR systems are defined the potential specificities of each UC are addressed in the Table 6.

Table 5. Implementation: additional specifications for monitorization per UC

Use Case	Specifications for monitorization
UC 2 and UC 3	Requirements regarding the access to specific markets or services need to be fulfilled (pre-qualification) for each asset, such as ramping time, power/energy, temporal requirements (short-term, mid-term), balancing capabilities, etc. Communication between the aggregator and flex-buyer as well as signalling from the flex-buyer needs to be processed accordingly

Use Case	Specifications for monitorization
UC 4	Additional detailed geographical information, server equipment, software/algorithms to determine asset portfolios and assure on-time delivery within distributed LFMs. DSOs might need additional metering points
UC 5	Aggregator has to supply DSO with monitoring data from UC2-UC4; information about grid status (voltage level, voltage imbalances, grid model)
UC 6	Additional metering/monitoring equipment for reactive power and phase-voltage in real-time through the STATCOM
UC 8	As UC4, additionally detailed geographical information, real-time communication and processing
UC 9 and UC 10	Depending on the markets requirements pre-qualification criteria must be met; such as metering resolution and information storage
UC 11	Distributed/decentralized entities (servers, computers) to validate (execute, control, sign, ...) provided services agreed via smart contracts, devices (oracles) to establish these contracts; smart meters and transformer stations (HV/MV and MV/LV)
UC 12	Additional reliable and fast data interconnection between LFMs and DSO. DSO might need more processing power for network monitoring tool

Regarding actuation hardware, the system will rely on the gateway, wireless sensors network and submetering infrastructure implemented:

- Power electronics/converters
- Servers and gateways running the EMS
- VPP-platform
- BMS
- High processing power
- Fast and reliable connection
- Controllable active/reactive power output as well as power factor

In the case of DR aggregation such as UC 2 and UC 3, the third-party supplier must insert additional equipment for load verification and disconnection. In some cases, an additional device should be installed because the meter has only two contacts. Nevertheless, the network operator has priority for load disconnection.

In the case of emergency state events, an additional interface between aggregator and DSO/LEMO/LFMO might be needed, where the link needs to be fast, secure and reliable.

Table 7 states the actuators needed for each DR system in each grid condition.

Table 6. Implementation specifications: required actuators per system

DR system	Normal/critical grid conditions	Emergency grid conditions
Battery Energy Storage System	Electrical actuator to on/off charge calculated based on flexibility offer.	Electrical actuator to on/off charge calculated based on flexibility offer and TSO grid state.
EV charger	Electrical actuator to Set-on/off charge state based on flexibility offer.	Electrical actuator to Set-on/off charge state based on flexibility offer and TSO grid state.
Lights	Electrical actuator to Set-point dimmer state based on flexibility offer	Electrical actuator to Set-point dimmer state based on flexibility offer and TSO grid state.
P2H: Heat pumps / HVACs	Electrical actuator to Set-point temperature according to flexibility offer, under potential distribution	Electrical actuator to Set-point temperature calculated according to TSO energy requirement order.

DR system	Normal/critical grid conditions	Emergency grid conditions
	constraints.	Switch-off capability
P2H: Water heaters	Switch-off capability	Switch-off capability
PV	Electrical actuator to control the inverter's active and/or reactive power output.	Electrical actuator to control the inverter's active and/or reactive power output. Switch-off capability
STATCOM	The energy box that sends the signals to the STATCOM in order to give or absorb reactive power and/or The energy box that sends the signals to the STATCOM in order to balance the phases	The energy box that sends the signals to the STATCOM in order to give or absorb reactive power and/or The energy box that sends the signals to the STATCOM in order to balance the phases

Considering these implementation requirements, an initial identification of costs and benefits has been performed to each of the use cases defined in Table 8.

Table 7. Costs and benefits analysis per UC

Use Case	Cost-Benefit Analysis
UC 1	Costs: man-hours for thermal modelling/calibration; additional sensoring and actuator equipment needed most likely for Water heaters; the higher the resolution (time & measurement value), the higher the costs for equipment and processing; Benefits: detailed and dynamic modelling improves flexibility-availability By correctly regulating the heat pumps and the stratification of the buffer boilers, it is possible to have a reduction in consumption of 15%. In single-family and two-family houses these adjustments are carried out normally, with the reduction at night 22.00-06.00. The network manager, depending on the model of the pump, has 2/3 users to control: 1. pump 2. additional resistance 3. boiler resistance. Currently, the meters have only 2 contacts, it would be necessary to insert a device to supplement the cost of about 170 €. This is because the various users have different programming. If the heat pump should stop 1/2/3/4 hours, the additional resistance could be blocked even 1.5 hours; depending on the accumulation, the boiler can work even 8 hours a day.
UC 2	Costs: potentially for prequalification and metering equipment to prove the transaction Benefits: qualified for trading and delivery of services In order to give flexibility to third parties, additional luminaires must be inserted. This means having space in the electrical counting cabinet. Space that we rarely find. Therefore, in addition to the equipment of the network operator (counter and possibly the additional relay box), the external supplier must install the equipment, and possibly also the control relays, in order to have the disconnection properties. Of course, data transmission is also an important point, which has to be managed by a third-party supplier with a proper approach (e.g. Wlan, Sim, etc.). Any additional equipment that is switched on is a possible failure or malfunction.
UC 3	Costs: costs for capable infrastructure

Use Case	Cost-Benefit Analysis
	Benefits: from market transactions/network services as well as arbitraging and self-consumption Same as UC2, same space problems to insert additional equipment.
UC 4	Costs: aggregation - development of software/hardware, storage; infrastructure as UC 1-3, additional metering equipment for DSO Benefits: improved energy efficiency/less losses, possible "local" self-sufficiency/autarchy; reduced energy costs, efficient grid management If the network operator cooperates with the LFM, the control and data transmission equipment is already in place, so no additional space and data transmission costs are required.
UC 5	Costs: sensors, hardware/software to manage flexibilities Benefits: pro-active avoidance of problems in the grid In this case, since there is a need to control more than one user, there will be a need to install an additional device at a cost of about 170 €. There is enough space for the installation, the remote control table is used.
UC 6	As UC 5, there will be additional costs for capable power electronics metering equipment; fast computers to process the data. Benefits are similar to UC 5.
UC 7	Costs: similar to UC 5 and UC 6 Benefits: congestion/instability avoidance - reducing costs occurring from curtailing, losses (energy efficiency), grid reinforcement, operation (also prevention of oscillations) & maintenance; proactive grid management.
UC 8	Costs and benefits as UC 5, UC 6 and UC 7
UC 9	Costs: equipment (converters) able to meet each market’s requirements (i.e. ramping time, power/energy provision), communication infrastructure, Link DR<->Aggregator<->market; shares distributed along the chain prosumer-Aggregator-supplier/trader-BRP Benefits: revenue from trading on anc. service markets
UC 10	Costs and benefits similar but lower than UC 8 but due to less requirements.
UC 11	Costs: distributed hardware Benefits: trust, security & availability
UC 12	Costs: hardware and software for grid monitoring, data connection Benefits: synergies between LFMs and gain in flexibility; improved energy efficiency/less losses, possible "local" self-sufficiency/autarchy; reduced energy costs, efficient grid management; pro-active avoidance of problems in the grid

3) Ex Post analysis

The report on the results of the PMV should be available to the end user through a Mobile App, Desktop widget or Web Application in order to ensure the transparency of the process.

The information provided for each event should contain the following aspects:

- Type of event (e.g. aFRR, RR, etc.)
- Systems activated and conditions set:

  - EV charger: previous charge status and time switched-off
  - Lights: dimming status before vs. after
  - P2H: Heat pumps / HVACs: set-point temperature before vs. after
  - P2H: Water heaters: DHW volume and temperature before vs. after
  - PV: active and reactive power output before vs. after
- Schedule of the event and duration
- Visual representation comparing between expected demand, as calculated through the baseline, and actual demand during the event
- Amount of reduced demand (kW or kWh) and/or increased amount of self-consumption rate in case of PV systems
- Unitary price (€/kW or €/kWh) and other remuneration information
- Comfort conditions during the event:
  - EV charger: charge status vs. minimum agreed in that schedule
  - Lights: illuminance data vs. minimum based on occupation
  - P2H: Heat pumps / HVACs: set-point temperature vs. minimum/maximum agreed
  - P2H: Water heaters: DHW volume and water temperature vs. minimum agreed
- Grid condition during the event (green, orange or red)
- Entity that requested the participation in the event: DSO or aggregator.

5. CONCLUSIONS

The main output of Task 3.3 is the defined steps of the PARITY PMV methodology. The procedure is comprised of three phases: ex-ante analysis, implementation and ex-post analysis.

The ex-ante analysis serves as the basis for the whole procedure, gathering available data and setting the event at hand. This phase consists of three steps: definition of DR events and criteria for remuneration, definition of DR systems and minimum comfort conditions and finally the identification of static and dynamic variables that affect the demand and that need to be measured.

The implementation phase first analyses the existing monitoring system and specification of metering points' and sensors' characteristics. Then an analysis of the technical and economic reliability of individual loads measurements is made and finally post-installation verification activities for algorithm calibration.

Ex-post, there are three final steps. First, the model is tested in a DR event to validate accuracy and reliability. Then demand/generation flexibility is assessed by comparing model and actual demand and in the end the PMV report is defined.

This methodology stems from the analysis of the current state of the art in M&V applied to energy efficiency and DR. The PARITY PMV addresses the barriers identified in this analysis:

- Baseline calculation use occupancy and human actions as dependent variables and for this reason are measurable and improve accuracy.
- The selection of historical data for baselining calibration and adjustment is made automatically, with the automated selection of the number of prior days for reference.
- The PARITY PMV does not need to perform adjustments to the baseline (and select an adjustment window) between notification and activation of the DR event. This prevents potential manipulation by users to artificially increase the perceived demand flexibility contributed.
- The PARITY PMV methodology is a blend of Options D and B of the IPMVP. The high accuracy calibration models serve as a simulation similar to Option D, but instead of building-level assessment, PARITY analyses demand load by load like Option B.

Besides, the PARITY PMV can assess flexibility potential and impact for a wider range of systems and building characteristics than precious methodologies. Different grid conditions are their consequent dispatch signals are considered and the systems involved, comfort conditions and remuneration adapt to each case.

Each of the systems, grid conditions and use cases considered in PARITY has been run through the PMV methodology, analysing the specificities and needs in each situation regarding required monitoring and actuators and variables affecting consumption.

The methodology ends with the reporting structure of the DR event results that serves two purposes: first, a transparent interface with end user that fosters trust and shows the reliability of the system and second, information that feeds impact assessment in Task 8.6.

6. REFERENCES

ERCOT Demand Side Working Group (Freeman, Sullivan & Co.). «Empirical Data on Settlement of Weather Sensitive Loads.» 2012.

International Organization for Standardization (ISO). «ISO 17741:2016 General technical rules for measurement, calculation and verification of energy savings of projects.» 2016.

n.d. https://evo-world.org/en/about-en/history-mainmenu-en.

s.f. https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp.

AEIC Load Research Committee. *Demand Response Measurement & Verification.* 2009. https://www.smartgrid.gov/files/demand_response.pdf.

American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). «Guideline 14-2002 – Measurement of energy and demand savings.» 2002.

Association of Edison Illuminating Companies (AEIC). «Demand Response Measurements & Verification.» 2009.

Australasian Energy Performance Contracting Association. «A best practice guide to measurement and verification of energy savings. A companion document to 'A best practice guide to energy performance contracts'.» 2004.

California Energy Commission . «Protocol Development for Demand Response Calculation. Draft and Findings. .» 2002.

Efficiency Valuation Organization (EVO). «International Performance Measurement & Verification Protocol. Concepts and Options for Determining Energy and Water Savings, Volume I.» 2010.

—. «International performance measurement & verification protocol. Concepts and practices for determining energy savings in new construction. Volume III Part I.» 2006.

—. «International performance measurement & verification protocol. Concepts and practices for determining energy savings in renewable energy technologies applications. Volume III Part II.» 2003.

—. «International performance measurement & verification protocol. Concepts and practices for improved indoor environmental quality. Volume II.» 2002.

—. «Uncertainty Assessment for IPMVP - Release of a New IPMVP Application Guide.» 2018.

Efficiency Valuation Organization (EVO),. «International performance measurement & verification protocol. Concepts and practices for improved indoor environmental quality. Volume II.» 2002.

Efficiency Valuation Organization. *International Performance Measurement and Verification Protocol. Concepts and Options for Determining.* January de 2012. http://www.eeperformance.org/uploads/8/6/5/0/8650231/ipmvp_volume_i__2012.pdf.

Enernoc. «The Demand Response Baseline.» 2009.

European Comission. s.f. http://iet.jrc.ec.europa.eu/energyefficiency/european-energy-service-companies/monitoring-and-verification.

European Commission. *Moeebius project.* s.f. http://www.moeebius.eu/.

—. *The ICT PSP methodology for energy saving measurement - a common deliverable from projects of ICT for sustainable growth in the residential sector.* April de 2012. http://eemeasure.smartspaces.eu/static/files/eemeasure_residential_methodology.pdf.

European Committee for Standardization (CEN). «EN 16212:2012 'Energy efficiency and savings calculation, top-down and bottom-up methods'.» 2012.

European Union. *EEmeasure.* s.f. http://eemeasure.smartspaces.eu/.

Federal Energy Management Program, US Department of Energy. «FEMP, M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0.» November 2015.

FEMP of the US Department of Energy. *M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0, Federal Energy Management Program.* November de 2015. https://www.energy.gov/sites/prod/files/2016/01/f28/mv_guide_4_0.pdf.

HOLISDER. *HOLISDER project.* s.f. http://holisder.eu/.

International Organization for Standardization (ISO). «ISO 50001:2011 Energy management systems – Requirements with guidance for use.» 2011.

—. «ISO 50015:2014 Energy management systems – Measurement and verification of energy performance of organizations – General principles and guidance.» 2014.

KEMA. «PJM Empirical Analysis of Demand Response Baseline Methods.» 2011.

Moeebius Project. *Energy performance assessment methodology.* 2016. http://www.moeebius.eu/333-report-moeebius-energy-performance-assessment-methodology.

North American Standard Energy Standards Board (NAESB). «WEQ Annual Plan Item 4(a) and 4(b) Final Action: Review and develop business practice standards to support DR and DSM-EE programs. .» 2010.

OrbEEt. «D1.2 Energy Measurements Methodology.» 2016.

OrbEEt project. s.f. http://orbeet.eu/.

Smith, John. «inteGRIDy sample reference.» *EU*, 2016: 1-2.

Southern California Edison . «Southern California Edison Company Demand Response Appendices to SCE-1, Volumes 1-4.» 2011.

The European Parliament and the Council of the European Union. «Directive 2006/32/EC of the European Parliament and of the Council on energy end-use efficiency and energy services and repealing council directive 93/76/EEC.» 2006.

U.S. Department of Energy. *Federal Energy Management Program.* Feb de 2016. https://www.energy.gov/sites/prod/files/2016/02/f29/fed_energy_mgmt_program.pdf.

United States Department of Energy (DOE). «M&V guidelines: measurement and verification for performance-based contracts version 4.0.» 2015.

Woodall, Gregg. *SMART 2011 / 0072: Methodology for energy-efficiency measurements applicable to ICT in buildings (eeMeasure): D1.5 Non-residential methodology.* 2011. http://eemeasure.smartspaces.eu/static/files/eemeasure_non_residential_methodology.pdf.

ANNEX 1: Background Literature Information for Measurement and Verification Methodology & Protocols

M&V protocols are key to accurately quantify the savings produced by an Energy Efficiency Measure (EEM). The first M&V protocols are intimately linked to the development of ESCO business models. Thus, the growing use of energy performance contracts (EPC) during the 1980s and 1990s in the US (Australasian Energy Performance Contracting Association 2004), led to the elaboration of the first guidelines on the matter. In Figure 1 the evolution of these methodologies in time since the early stages of the M&V is reflected.

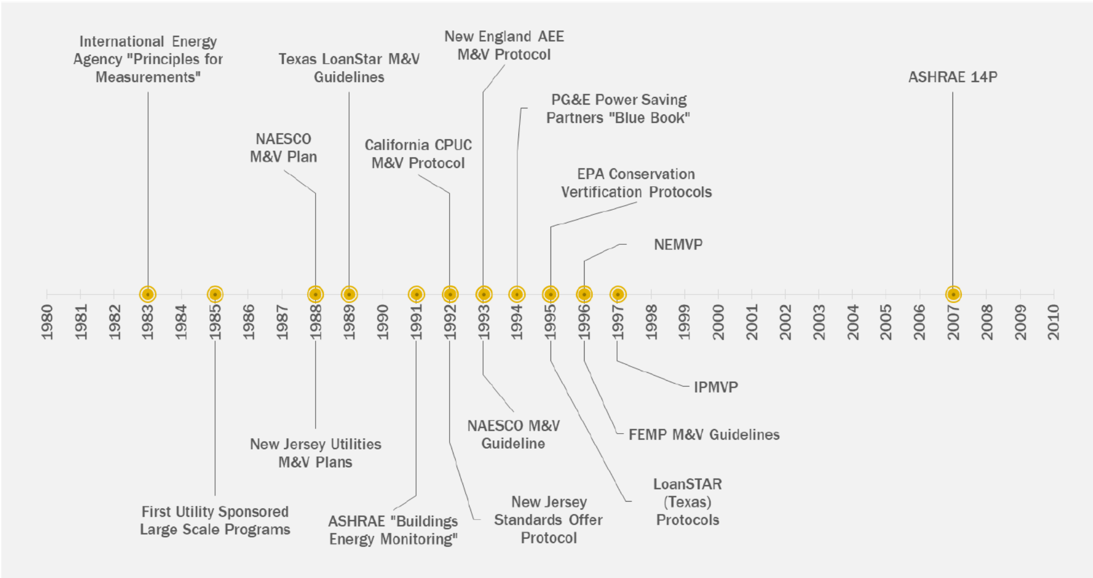

Figure 1. Chronogram of M&V protocols (Australasian Energy Performance Contracting Association 2004)

One of the key developments came in 1994 with the US Department of Energy (DoE) beginning to work with industries to find a unified and consensed method to measure and verify results in energy efficiency. From there, the North American Energy Measurement and Verification Protocol (NEMVP) came out in 1996, which could be considered the first version of a M&V protocol. Many companies from the USA, Canada and Mexico contributed to the development of the methodology[2].

Given the global interest on the matter, in 1997 a second edition was published involving organizations from twelve countries and professionals from more than 20 countries worldwide. The document was renamed International Performance Measurement and Verification Protocol (IPMVP) (Efficiency Valuation Organization (EVO) 2010). Although this version was very similar to the previous one, contents related to efficiency opportunities in new construction projects and in the use of water were included.

In 2001, a third version with two volumes was published:
Volume I: Concepts and Options for Determining Energy Savings

Volume II: Concepts and Practices for Improving Indoor Environmental Quality (Efficiency Valuation Organization (EVO), 2002).

At the same time, a non-profit organization was formed: IPMVP Inc., to maintain and update the current content, as well as to develop new guidelines. In 2004, this organization was renamed as Efficiency Valuation Organization (EVO), which is the current name. To date, the published documents have been continuously reviewed and new documents have been produced. The latest English version dates from 2012 (Australasian Energy Performance Contracting Association 2004) (EVO).

[2] https://evo-world.org/en/about-en/history-mainmenu-en

Even though IPMVP is possibly the most common procedure, there are other protocols that are based on the methodology described. In 1973, the US started the Federal Energy Management Programme (FEMP) with the goal of introducing a more efficient use of energy in government facilities. Therefore, in 1996, the FEMP M&V Guidelines (United States Department of Energy (DOE) 2015) were published, based on the recent NEMVP that after became the IPMVP. This method was especially designed for federal facilities. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) also worked on a methodology for M&V, resulting in 2002in the ASHRAE Guideline 14-2002 (American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) 2002). This document is much more technical than the IPMVP.

In Europe, the EVO's IPMVP protocol was being applied, but in 2012 the European Committee for Standardization (CEN) published the standard EN 16212:2012: "Energy Efficiency and Savings Calculation, Top-down and Bottom-up Methods" (European Committee for Standardization (CEN) 2012). The goal was to harmonise the methods for monitoring and evaluating energy savings considering the policies and actions within the framework of the European Union for the reduction of greenhouse gas emissions and the increase of energy efficiency. The document describes a procedure for calculating energy savings in final energy consumption in buildings, cars, equipment and industrial processes, among others, with *ex ante* and *ex post* evaluations in any period.

The top-down and bottom-up methods were designed within the European Directive 2006/32/EC on energy end-use efficiency and energy services (The European Parliament and the Council of the European Union 2006) (currently replaced by the European Directive 2012/27/EU on energy efficiency). The top-down method proposes the estimation of savings from large statistical data while the bottom-up method is based on summarising actions of end users.

Finally, in the international context, the International Organization for Standardization (ISO) published the standard ISO 50015:2014 "Energy management systems - Measurement and verification of energy performance of organizations - General principles and guidance" (International Organization for Standardization (ISO) 2014), complementing the previous ISO 50001:2011 "Energy Management System" (International Organization for Standardization (ISO) 2011), for M&V, important for ISO-based energy management systems.

Later, the ISO also published the standard ISO 17741:2016 "General technical rules for measurement, calculation and verification of energy savings of projects" (International Organization for Standardization (ISO) 2016). In this standard, energy savings are determined by comparing measured, calculated or simulated consumptions before and after the implementation of an energy-saving action and applying adjustments when relevant variables are modified (routine adjustments) or in static factors (non-routine adjustments). This document is clearly following similar procedures as IPMVP.

The European Commission DG JRC (European Comission s.f.) recommends that performance-based projects should be subject to M&V protocols in order to evaluate their results. Therefore, in the PARITY project a detailed definition of a PMV methodology is required in order to verify the response rate to flexibility signals.

Previously, other European Commission co-funded projects (e.g. eeMeasure, Moeebius, OrbEEt, HOLISDER, FLEXCoop) have developed or improved M&V methodologies for the verification and assessment of buildings energy performances mainly based on IPMVP (Efficiency Valuation Organization 2012) and FEMP (FEMP of the US Department of Energy 2015). Considering these international methodologies, the most extended and the basis for the development of the others existing protocols, a summary of their key aspects has been included below with a description of other existing methodologies and protocols. This includes guidelines such as the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) Guideline 14 and the US DOE’s Uniform Methods Project.

M&V methodologies used for energy efficiency assessment

International Performance Measurement and Verification Protocol

Until 2012, the IPMVP was divided into the following three volumes:

Volume I - *Concepts and Options for Determining Energy and Water Savings*. In this document the basic concepts are described and the methodology to be followed is developed. It is considered as the most essential volume, due to the fact that it includes most of the information required to apply the IPMVP.

Volume II - *Concepts and practices for improved indoor environmental quality (2002)*. This document addresses the environmental aspects of indoor air that are related to the design, implementation and maintenance of EEMs (Efficiency Valuation Organization (EVO) 2002).

Volume III - *Concepts and Options for Determining Energy Savings in New Construction*. It provides details for the M&V methods in the construction of new buildings and in renewable energy systems. It is divided in two parts:

- Part I - *Concepts and practices for determining savings in new construction (2006*) (Efficiency Valuation Organization (EVO) 2006).
- Part II - *Concepts and practices for determining energy savings in renewable energy technologies applications (2003)* (Efficiency Valuation Organization (EVO) 2003).

Starting in 2014, EVO decided to reorganize the IPMVP documents and now publishes the IPMVP Core Concepts. The document defines the terminology and principles for applying M&V. It describes the project framework, the contents and requirements, and saving reports:

- Principles
- IPMVP Framework
- IPMVP Options
- IPMVP Adherent M&V Plan and Report
- Adherence with IPMVP

Due to its significance, the following review only addresses the most important concepts of the methodology's principles, framework and options, the key to be able to apply the M&V protocol. One of the initial steps is to define the principles of M&V on which the IPMVP is based, and that must be considered by any M&V plan based on this protocol:

- Accurate: the M&V reports should be as accurate as possible, always taking into account the allocated cost.
- Broad: a report that demonstrates the savings must consider all aspects of a project.
- Conservative: when making estimations, the savings should be underestimated.
- Coherent: the reports must be consistent with the different energy efficiency projects, the professionals responsible for energy management, the time periods of a project as well as projects for energy supplies.
- Relevant: to calculate the savings, the parameters of interest must be measured, while the least important of them or the predictable ones can be estimated.
- Transparent: detailed documentation of all the M&V activities must be provided.

Considering that energy saving is impossible to be directly measured, since it is the absence of energy consumption, the way to estimate the savings achieved through an EEM is to compare the consumption in two periods of time. The first or *reference period* comes before the implementation of the EEM. In this period the *baseline* is established, characterizing the consumption curve. Independent variables have a significant impact (e.g. outdoor temperature, hours of operation, occupancy, etc.). On the other hand, the period after the implementation of the EEM is called *reporting period*. In it the energy curve (called *adjusted baseline*) will be projected based on the reference baseline identified period and corrected according to some independent variables that have a substantial impact (e.g. outdoor temperature, hours of operation, occupancy, etc.). The difference between the adjusted baseline and the actual measured consumption in the reporting period will represent the energy savings achieved. The IPMVP framework is exemplified in the following figure.

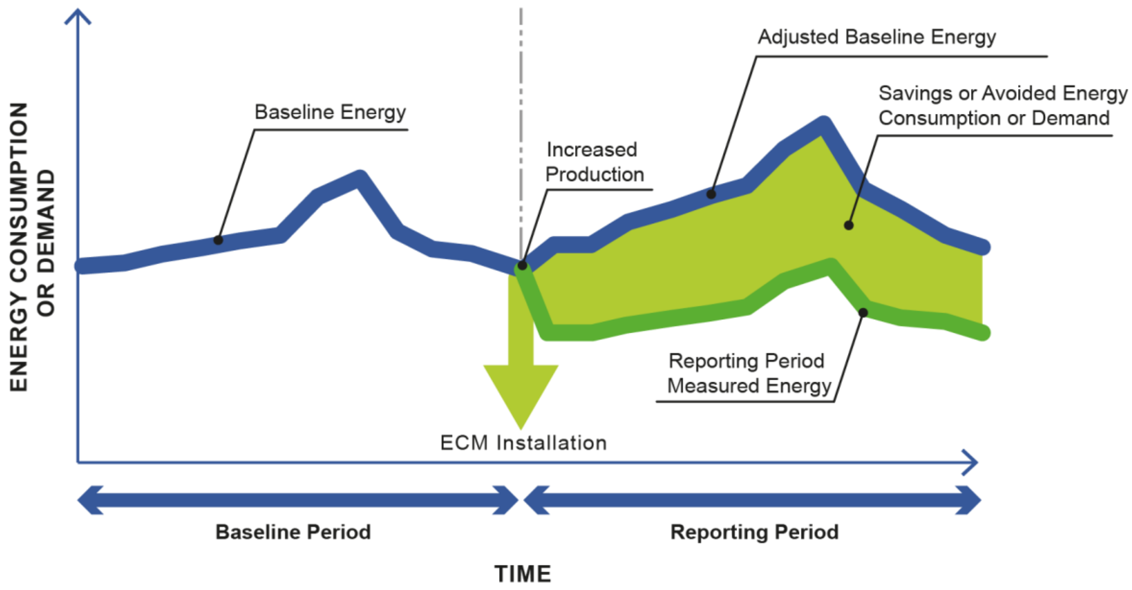

Figure 2. IPMVP framework (EVOIPMVP)

The amount of savings illustrated in the figure above in green colour, can be calculated by the following equation:

Savings = (Baseline Period Energy – Reporting Period Energy) ± Adjustments

Depending on factors such as scope, available data, measurement equipment available, installation characteristics, budget for the M&V or the EEM; to calculate the savings the IPMVP proposes four options:

- Option A. Retrofit isolation: key parameter measurement. It is the option of the lowest cost, however it yields the greatest uncertainty. Savings are determined by measuring a key parameter and by estimating the rest based on historical data, manufacturer specifications or technical assumptions. The measurement made can be continuous or punctual depending on the expected variation of the key parameters.
- Option B. Retrofit isolation: all parameters measurement. The saving is determined by measuring all the parameters that may affect energy consumption. Like the previous option, the measurement can be carried out in a timely or continuous manner depending on the expected variation of savings.
- Option C. Whole facility. The savings are determined by measuring the energy consumption of the installation (whole or a part of it). The measurement is carried out continuously throughout the reporting period. This option is recommended when, for example, the EEM affects several equipment in the facility.
- Option D. Calibrated simulation. The savings are determined by simulating the energy consumption of the installation (whole or a part of it). This simulation must be calibrated with information of the invoices or the measurement of some equipment. The cost of this option is usually high, as it requires deeper technical knowledge. This option is proposed to be used in cases when there are no real measurements available in the reference period.

An important point to effectively develop a M&V plan is the correct selection of the reference and reporting measurement periods. Regarding the reference period, it must be ensured that it covers a complete operating cycle, and is immediately before the implementation of the EEM, since a more distant period could alter the conditions. Similarly, the reporting period has to cover at least one normal operating cycle to assess the effectiveness of the savings. The duration of the period will depend on the user and of the savings reports. The measurement equipment must be installed before the periods in order to gather the necessary data.

IPMVP needs to record the reference period energy data and all influencing variables to have an accurate determination of the savings. Data gathering must be an integral part of the M&V Plan. This document collects the details of the M&V to allow an easy consultation without losing information. The M&V Plan should include the following points:

1. **Objective of the EEM**. Description of the EEM, objective pursued.
2. **Option of the IPMVP**. Definition of the IPMVP option that will be used depending on the scope and the measurement. The date of publication, the version and the volume of the IPMVP edition should be referenced.
3. **Reference: period, energy and conditions**. Reference conditions and energy data in this period will be documented, including:
    - Identification of the reference period.
    - Data of reference consumptions.
    - Information about the independent variables related to the energy data.
    - Static variables such as occupancy, operating conditions, equipment inventory, significant problems with equipment or power outages during the reference period, etc.
4. **Reference period**. The reference period should be identified.
5. **Base for adjustment**. The conditions under which the energy measurements will be adjusted in the reporting period. Both the independent variables that may have a significant impact on energy consumption and static variables (changes that trigger non-routine adjustments) should be defined.
6. **Analysis procedure**. The procedure for analysing the data as well as the algorithms and estimations used in the analysis will be specified. All the elements that have been used in the mathematical model and the validity range for the independent variables will be also included.
7. **Energy prices**. The price of energy supply will be specified to assess economic results.
8. **Measurement specifications**. The measurement points will be described as well as their characteristics, routine calibration processes and the method to deal with potential data gaps.
9. **Monitoring responsibilities**. The responsibles for report elaboration and energy data, independent and static variables gathering should be assigned.
10. **Expected accuracy**. Evaluation of the expected accuracy of the measurement, data collection, sampling and data analysis, including qualitative and quantitative assessments according to the quality of the measurements and the adjustments defined.
11. **Budget**. The budget and resources required to verify the savings will be included.
12. **Report format**. The format and substance of the savings report will be specified.
13. **Guarantee quality**. Quality procedures used in the report and its preparation.

After the EEM’s implementation, during the reporting period, the expected reports will be made with the defined format in the M&V Plan. These reports will be the final output of the M&V, describing the energy and economic savings achieved. The periodicity of the reports will be agreed in the M&V Plan.

FEMP

The Federal Energy Management Programme (FEMP) is a U.S. Department of Energy (DOE) programme focused on reducing the federal agencies energy consumption by offering information, tools, and support toward tracking and achieving energy related obligations and targets. FEMP pursues contracts with small businesses to help in this effort (U.S. Department of Energy 2016). FEMP (FEMP of the US Department of Energy 2015) defines six steps to measure and verify savings:

1) <u>Allocate Project Risks and Responsibilities</u>: The basis of any project-specific M&V plan is determined by the distribution of key project risks regarding financial, operational, and performance problems and obligations between the ESCO and the customer.

2) Develop a Project-Specific M&V Plan: The M&V plan specifies how savings will be calculated and stipulates any activities that will take place after equipment installation. The project-specific M&V plan contains project-wide elements as well as particulars for each EEM.

3) Define the Baseline: Baseline physical conditions are determined through surveys, examinations, spot measurements, and short-term metering activities. Properly defining and documenting the baseline conditions is key in the process. Monitoring needs (and for how long) depend on factors such as the action's complexity and the baseline stability, including the variability of equipment operating cycles operating hours, and the other variables affecting the load.

4) Install and Commission Equipment and Systems: Commissioning confirms that systems are designed, installed, functionally tested in all operation modes, and able to be operated and maintained according to the design scope (appropriate lighting levels, cooling capacity, comfortable temperatures, etc.).

5) Conduct Post-Installation Verification Activities: Post-installation M&V activities are performed to make sure that proper equipment/systems were installed and operational, and have the capability to produce the predicted savings. Verification methods include surveys, inspections, spot measurements, and short-term metering.

6) Perform Regular-Interval M&V Activities: M&V is required an annual basis. With proper coordination and planning, M&V activities that provide operational verification of an EEM (i.e., confirmation that the EEM is operating as intended) through the performance period can help with ongoing commissioning activities (e.g., recommissioning, retro-commissioning, or monitoring-based commissioning).

ASHRAE Guideline 14

ASHRAE Guideline 14: *Measurement of Energy, Demand and Water Savings*, is a reference for determining energy and demand savings related to performance contracts using measurements. In addition, it sets guidelines for instrumentation and data management and describes methods for coping with uncertainty associated with models and measurements. Guideline 14 does not discuss other issues related to performance contracting. The ASHRAE guideline identifies three approaches to M&V. Compliance with each approach requires that the general uncertainty of the savings estimations be below stipulated thresholds. The three approaches offered are strongly related to and support the options provided in IPMVP, except that Guideline 14 has no similar method to IPMVP/FEMP Option A (FEMP of the US Department of Energy 2015).

Overview of the Uniform Methods Project (UMP) by DOE

Under the Uniform Methods Project3 (UMP), DOE created 24 different protocols for verifying savings from different sorts of EEMs and programmes. The protocols are classified on four groups: commercial, residential, combined commercial and residential, and cross-cutting measures.

The protocols offer a direct method for evaluating gross energy savings for residential, commercial, and industrial measures usually offered in ratepayer-funded programmes in the US. The measure procedures are based on a particular IPMVP option, but involve additional processes needed to aggregate savings from individual projects to evaluate impacts on a program level. For commercial actions, the FEMP standard and the UMP are complementary. Nevertheless, since one of the goals of M&V in a project is to ensure long-term performance, the FEMP guideline includes further proposals for annual inspection and measurements where applicable (FEMP of the US Department of Energy 2015).

M&V methodologies used for Demand Response verification

The main role of M&V in DR is to determine the quantity of energy or power that is "delivered" by a DR resource under the conditions imposed by a DR programme. The use of a significant M&V for DR performance is needed for a fair and clear financial flow to and from market actors and ultimately establishing market confidence. In fact, correctly determining the quantity of demand provided by a DR resource is needed to ensure an accurate payment based on their measured flexibility. On the other hand, an accurate forecast of the DR at an individual and aggregated level (reliant on the consistency of the DR performance measurements), can improve operational efficiency and the realization of an efficient and sustainable electricity system. Moreover, measured DR performance is the main input to design a business model and guarantee a cost-effective evaluation.

In summary, PMV for DR is used for:

- Determining the eligibility or capacity of resources: For most products and services that DR can deliver, the potential of the resource needs to be assessed before it can participate in the DR programme.
- DR settlement: DR settlement is the determination of DR amounts achieved, and the financial operation between the programme or product operator and the user, based on those quantities. For DR programmes that offer incentives for load reductions, the estimated load without curbing determines the calculated reduction quantity that the contract with each DR resource is based on. More generally, different M&V may be used to resolve between a programme operator and its end users, and it is used to settle that programme as an aggregated reserve in the wholesale market. However, even if measured reductions are not required for settlement either with retail participants or with the wholesale market, DR M&V via impact assessment is important for evaluating programme effectiveness and for future planning.

There are many types of agreements a retail operator may have with its customers; many of these structures do not need measurement of demand reduction for settlement with the customer or aggregator. However, when the programme- is offered as a wholesale resource, the calculated demand reduction total for the programme or segment is usually needed for settlement. For all programme types, if impact estimation is performed, its primary goal is to establish the quantities of demand reduction accomplished by the DR programme. Thus, using a performance evaluation system in DR events involves the evaluation against a baseline of the volume of demand variation that is sold into the market. This volume of demand flexibility is determined as the difference between what the consumers typically consume (the baseline) and the real measured consumption during the dispatch event. Since the baseline cannot be measured directly, it has to be estimated and calculated from other measured data using a robust method. Therefore, measurement of any DR resource normally entails comparing observed load during the time of the reduction to the estimated load that would otherwise have consumed without the curtailment. The resulting difference represents the actual load reduction (Figure 3).

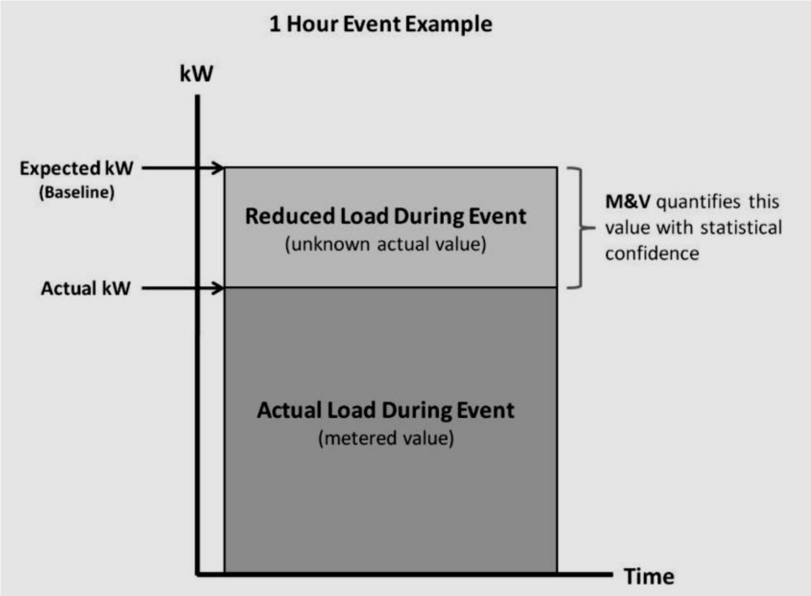

Figure 3. M&V Quantification of Load Reduction Value (AEIC Load Research Committee 2009)

The performance evaluation methodology that is utilized for settlement of the DR programme is crucial to the success of any DR programme. Being able to assess the available reduction capability and linking payment to the amount of reduction on the event, are key features of DR programmes where event frequency and implementation can need different types of baseline. When pay-for-performance is measured by comparison to a certain value, precise measurement is vital, and verification is simple. In cases where performance is evaluated in relation to a baseline, the definition of the baseline and actual energy measurement are crucial. The task is to find a simple but accurate valuation of a customer's energy usage reductions comparative to a baseline during a specific timeframe (i.e., the DR deployment period) that is fair to all sides. As estimates, baselines are intrinsically flawed.

Nevertheless, according to NAESB (North American Energy Standards Board) recommendations good baselines search for four main attributes:

1. **Accuracy**: giving credit for no more and no less than the curtailment achieved.
2. **Integrity**: discouraging irregular consumption and making sure that it does not influence baseline calculations. A high level of integrity will protect against efforts to trick the system.
3. **Simplicity**: performance calculations should be clear and understandable for all actors, including end-users' customers.
4. **Alignment**: DR programme designers select the baseline methodology taken into consideration the goals of DR programme.

Balancing these attributes is difficult. In some cases, baselines impervious to manipulation will be complex and difficult to be calculated an understood. In others, cleaner approaches could allow participants to take advantage of the baseline in their favour. Moreover, it is important to contemplate that baseline estimation should not reward or penalize normal load variation caused by system operations and generally related to changes in occupancy or weather conditions.

Various M&V methodologies for DR have been developed in recent years in the US and in previous research projects in EU. In the following pages, specifications of these methodologies are presented.

The eeMeasure methodology

As an expansion of the IPMVP, the eeMeasure project examines two different methodologies for M&V. Both are based on IPMVP and are built from the experience of current and past ICT PSP projects which includes approximately 10,000 social homes and 30 public buildings, such as hospitals and schools (European Union s.f.). This was the first European project that developed a methodology to measure and verify DR in the European context. These methodologies have been utilised in three EU H2020 projects and one FP7 project, such as NOBEL GRID, MOEEBIUS, ORBEET and Inertia.

The Residential Methodology (European Commission 2012) is valid only to dwellings and normally accepts a monthly measurement period. In the residential sector, a hypothesis of constant demand (Option A) or a cycle of predictable demand (Option B) or another demand structure that can be accurately modelled (Option D) cannot usually be made. None of these assumptions can be applied to projects seeking to change the resident behaviour as the key for the intervention to take effect. Therefore, the approach offered in IPMVP as Option C is certainly the only applicable in this context. Option C verifies energy savings annually or even in a shorter time period based on energy use measurements at the whole facility or sub-facility level. This option does not presume constant energy demand or any modelled variation but is a before-after comparison instead.

Non-Residential Methodology (Woodall 2011) can be used for any property type (including residential) and with any data frequency. In this methodology, a process flow is defined, which monitors appropriate variables in order to build an accurate model. A description of the core mathematical statistics is also provided.

6.1.1.1 Option C for residential

The before-after comparison of energy savings is estimated from the difference between consumption after the Energy Saving Intervention (ESI) and the consumption estimated to have taken place under the same conditions without the ESI (European Commission 2012):

- The estimation of consumption without the ESI is called *baseline data*. The baseline extension is the forecast of consumption in the period after the intervention.
- The period after the intervention during which measurement of saving takes place is described as the *reporting period*. After the ESI intervention, energy consumption should be reduced.

In order to estimate avoided consumption, it is required a model that changes based on the variations of independent variables such as outside temperature, occupancy, household size etc. If no independent control variables can be measured, the selection of a baseline period is key for accuracy. The recommended approach is to create regression models to mimic the energy consumption based on the independent variables. Climatic changes are of the main link to variability in residential consumption profiles. Average temperature or heating degree days (HDD) and cooling degree days (CDD) are often used. For regression models, an adequate accuracy of modelling of the dependent variable is necessary to correctly estimate the extended baseline in the reporting period. One metric to assess this accuracy is the squared multiple correlation coefficient R^2, which reflects the proportion of variance explained in the model. If R^2 is low (less than 0.7), additional independent variables must be included to improve predictions. If R^2 remains low, only very large energy savings will be reliably identified.

In the before-after comparison approach of eeMeasure, six steps are required:

1. Nominate a time period for the baseline which captures all variation of immeasurable independent variables and can yield an average which can reasonably be expected to be repeated in the future.
2. Gather data for the energy consumption (dependent variable) and for all accessible independent variables (baseline period).
3. Perform a regression analysis to establish the coefficients for each independent variable.
4. Nominate a time period for the reporting period which is again long enough to capture all variation of immeasurable independent variables.
5. Gather data for the energy consumption (dependent variable) and for all accessible independent variables (reporting period).
6. Apply the coefficients estimated in the baseline to the reporting period, yielding the result: energy saving as the difference between the estimated and actual energy consumption.

Step 1, 2 and 3. Baseline period estimation

To determine energy saving at building level, they have to be associated to the size of the units. The considered units must be the same for the baseline and for the reporting period. Depending on the particular unit and the kind of consumed energy, energy savings may depend on different independent variables such as ambient temperature, occupancy, and floor area. However, the effect of independent variables can sometimes be deemed as insignificant. If an impact is considered in the baseline estimation, independent variables should be part of the measurement plan before the intervention, but if their measurement is either not possible or financially viable, the definition of a consistent baseline period is crucial for accuracy.

The length of the baseline will vary depending on the independent variables affecting the consumption. For instance different holidays' patterns or climatisation periods. Since the "non-intervention consumption" cannot be directly measured, the suggested approach is to develop regression models that replicate the energy consumption based on independent variables data. The dependent variable, consumption of energy, is accurately and continuously monitored and reported by smart meters. Some independent variables like outside temperature, can also be measured and registered reliably. Energy-related behaviour or the social structure of households are both independent variables that provide energy consumption patterns data and therefore, have a direct implication on energy savings. Such data may be collected through surveys to residents and are subject to the GDPR.

Step 4 and 5. Reporting period estimation

After the ESI and a following period with improvements/adjustments, the energy savings should stabilize for a definite time period. To monitor the variation of energy savings in time the following steps are necessary:

- In the short term, energy savings can be compared weekly to check their stability over time after the ESI, especially if they are influenced by social behaviour.
- In the long term, it is critical to take into account equipment renovations as the baseline estimations may vary.

6.1.1.2 Demand Response benchmarking methodologies according to the eeMeasure methodology

The eeMeasure methodology considers four baseline methodologies to assess the peak shaving achieved in a DR scenario (European Commission 2012).

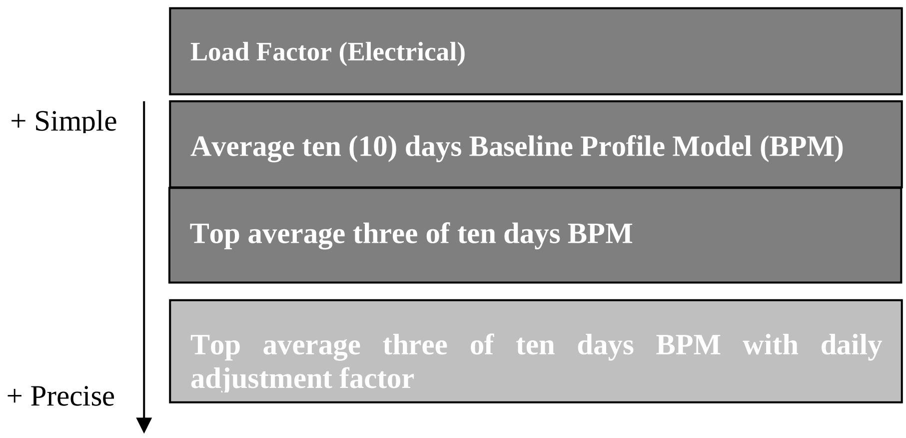

Figure 4. Demand Response benchmarking methodologies (European Union s.f.)

Load factor

The load factor (LF) is defined as the outcome of the division of the minimum power demand by the maximum power demand of a building:

LF = (min power demand)/(max power demand)

The closer the value of the load factor is to 1, the less the demand curve peaks. If the building load curve peaks run parallel to the electricity network peaks, changing the LF towards 1 can signify useful peak shaving for the utility.

10 days Baseline Profile Model

Baseline profile models (BPM) are employed to estimate the shaving of peaks, which unpredictably occur on particular days, the peak "event". To estimate non-intervention consumption at the peak event, it is usually agreed that a reference period of 10 business days directly prior to the event fairly represents consumption for normal operations. The reporting period is normally the whole duration of the event day.

In this model, the average represents the non-intervention reporting period (event day) estimate. Real consumption on the event day is compared to this average to calculate the peak shaving. The consumption over the 10 days is averaged according to the equations below:

b:(d1(t,h)+d2(t,h)+d3(t,h)+d4(t,h)+d5(t,h)+d6(t,h)+d7(t,h)+d8(t,h)+d9(t,h)+d10(t,h))/10 for the number of hours of the event   or

DR consumption= Demand event day (day 11) - Baseline (average 10 days)

Top 3 of 10 days Baseline Profile Model

This model is an average of the three highest consumption records from the previous 10 days, excluding other event days, holidays etc. The estimation for the non-intervention event day consumption is:

b: max (1,3) (Σdn(t,h))/3   or

DR consumption= Demand event day (day 11) - Baseline (average high 3 of 10 days)

Top 3 of 10 days Baseline Profile Model with morning adjustment factor

In cases where consumption is higher on event days, this model captures day-of realities in a customer load profile by adjusting with day-of event conditions. The estimation for event day (reporting period) non-intervention consumption is:

b': max (1,3) (Σdn(t,h))/3

P: (d(t,h-1) – b(t,h-1) + d(t,h-2) – b(t,h-2))/2

DR consumption= Demand event day (day 11) - Baseline (average high 3 of 10 days) + morning adjustment factor

Other EU Projects

There have been DR projects focused on residential units in which the Residential eeMeasure methodology is used. In the next pages we delve into these projects.

6.1.1.3 Moeebius project - Modelling Optimization of Energy Efficiency in Buildings for Urban Sustainability (European Commission s.f.)

Moeebius introduces a Holistic Energy Performance Optimization Framework that improves modelling approaches and provides innovative simulation tools. They describe real-life building operation difficulties in accurate simulation forecasts that significantly lower the "performance gap" and enhance multi-fold, continuous performance optimization of building energy to further reduce the identified "performance gap" in real-time or through retrofitting. The energy performance assessment approach of this project was published on its website (Moeebius Project 2016) and feeds from the IPMVP and the FEMP methodologies (Federal Energy Management Program, US Department of

Energy November 2015). The Moeebius M&V is organized in three phases: ex-ante analysis, implementation and M&V.

The ex-ante analysis compares the baseline and the model. The baseline is described by the following:

- the examination of the energy consumption over an adequate period (around one year) and with adequate granularity (e.g. hourly) to identify alterations in consumption;
- estimated analysis in energy consumption by usage type (e.g. lighting, heating equipment, office machinery, etc.);
- independent and constant variables that influence the energy consumption and the relative values for their measurement (i.e. building opening hours, degree days for heating or cooling, floor area for lighting, duration of metering period, etc.).

This data need to be registered simultaneously with the energy consumption data. A calibrated simulation model will be needed for the evaluation of the difference between the projected (estimated) consumption and the actual consumption.

The implementation entails identifying the energy sources, stipulating the metering points, and the energy consumption (from real-time monitoring to time aggregation).

The last phase calculates the KPIs' evolution evaluates the final performance of the system to optimize energy at home or building level.

6.1.1.4 OrbEEt project - ORganizational Behaviour improvement for Energy Efficient administrative public offices (OrbEEt project s.f.)

The OrbEEt project introduced an innovative solution to accelerate public and social engagement to action for energy efficiency by providing real-time evaluations of the energy impact and energy-related organisational behaviour. The OrbEEt M&V applies Option C & D from the IPMVP, and creates a procedure that unites annual bills and building sub-metering data (OrbEEt 2016). This M&V creates a continuous validation approach (different measurement intervals) but in parallel for different loads (different systems). The periodic savings need to be adjusted to re-state the baseline demand of the periods for a common set of conditions. These adjustments are based on independent variables (weather conditions, building occupancy, etc.), as the eeMeasure methodology defined. Because at the beginning of the project, sub-metering information for all pilot zones was lacking, they simulated energy uses (Option D from IPMVP) when data was not available for the baseline period or when future changes were expected. Energy consumption was simulated based on hourly or monthly utility billing data after the installation of gas and electric meters.

Option B was used at the next phase of measurement of energy consumption. Subject to the type of consumption which shall be compared, it is possible to have different timeframes (weekly, monthly, yearly) to define a baseline period. The definition of the baseline period for the different types of devices examined was:

<u>Fuel/Gas: HVAC systems</u>

- Baseline period: one year period needs to be used for baseline definition
- Information to gather: Monthly energy consumption
- Independent variables for routine adjustments: HDD or CDD and occupancy level
- Static information (non-routine adjustments): the facility size, the design and operation of installed equipment, the number of weekly production shifts, or the types of the occupant.

<u>Electricity</u>: NO HVAC systems (lighting and office equipment)

- Baseline period: one week period needs to be used for baseline definition
- Information to gather: The daily average of week consumption
- Independent variables for routine adjustments: Occupancy level
- Static information (non-routine adjustments): the facility size, the design and operation of installed equipment, the number of weekly production shifts, or the types of the occupant.

Information about external conditions (received through external weather service) and occupancy levels (surveys to pilot representatives) was also available from the pilot sites. Routine adjustments, such as seasonal occupancy, are made to these independent variables. Non-Routine adjustments are adjustments in parameters that cannot be predicted and for which a substantial effect on energy use/demand is anticipated. Non-routine adjustments have to be based on known and agreed changes to the building:

- changes in the volume of space being heated,
- changes in the characteristics of the equipment (power, amount, use)
- changes in equipment set-points (lighting levels, set-point temperatures)
- changes in occupancy

6.1.1.5 HOLISDER project - Integrating Real-Intelligence in Energy Management Systems enabling Holistic Demand Response Optimization in Buildings and Districts (HOLISDER s.f.)

HOLISDER integrates many different fully developed technologies in an open framework, a fully-fledged suite of tools to address the needs of the whole DR value chain. To goal is to ensure consumer empowerment to become active market players, through a variety of implicit and hybrid DR schemes, with support from end-user applications.

The hybrid M&V approach of HOLISDER project combines option B and C from IPMVP, making use of methodological steps of Option B but applying features from option C to manage unexpected events, such as sub-metering data gaps, etc. Sub-metering is employed at the first stages of the reference period of the project during the whole duration of the project; collecting high-resolution information from the pilot buildings. The eeMeasure methodology is enhanced to follow a pooled baseline regression analysis model establishing a variable relationship between event days and baseline consumption.

6.1.1.6 FLEXCoop project - Democratizing energy markets through the introduction of innovative flexibility-based demand response tools and novel business and market models for energy cooperatives

FLEXCoop introduces an end-to-end Automated Demand Response Optimization Framework. It enables the realization of novel business models, allowing energy cooperatives to introduce themselves in energy markets under the role of an aggregator. It equips cooperatives with innovative and highly effective tools for the establishment of robust business practices to exploit their microgrids and dynamic VPPs as balancing and ancillary assets toward grid stability and alleviation of network constraints.

FLEXCoop makes use of a wide range of baseline technologies to build an open and interoperable DR optimization framework, including a fully-fledged tool suite for energy cooperatives (aggregators) and prosumers involved in the DR value chain.

The FLEXCoop PMV methodology cannot be strictly associated to the IPMVP’s options but it has common points with Option B and Option D approaches. In particular, it is based on continuous measurement of individual loads and parameters that define the baseline, thus, being very close to the Option B approach. On the other hand, since measurement data are used to generate and continuously calibrate forecasting models in the FLEXCoop PMV approach, it is also similar to Option D. In this case, the difference is that the models are not created at building level, but for each electrical use participating in DR events.

The difficulties in the selection of the reference and reporting period, in the case of FLEXCoop PMV method can be overtaken both thanks to the methodology itself and to the different duration of EEM implementation, that in case of DR events is limited to a short period, which corresponds to the reporting period. The reference period is the one allowing the creation and calibration of FLEXCoop models with the minimum amount of required data. In particular, the reduced amount of data needed for baseline construction and calibration is an advantage of FLEXCoop PMV method since it addresses a common issue of IPMVP that is the requirement of large amount of data during a long time period to create an accurate baseline.

ANNEX 2: BASELINE ESTIMATION IN M&V METHODOLOGIES

M&V methodologies are adapted to the type of programme (e.g. energy, reserve, etc.), load (e.g. weather-sensitive, flat load, etc.) and customer (e.g. residential or commercial). The most important features required in their design and implementation are focused on achieving a correct definition of a baseline estimation methodology that also incorporates the definition of methodologies for historical data analysis, baseline adjustments and the evaluation of its accuracy. In this section, the most well-known methodologies are collected before introducing practical experiences (and associated recommendations) learned in their application.

Baseline estimation methods

In Northern America organised electricity markets have experience with explicit DR testing numerous PMV methodologies. The North American Energy Standards Board (NAESB) (North American Standard Energy Standards Board (NAESB) 2010) has defined five types of methodologies to promote coordination and eliminate market barriers for new providers:

- Maximum base load,
- Meter before / meter after,
- Baseline type-i
- Baseline type-ii
- Experimental design
- Metering generation output.

Depending on each case, one of these methods could be considered as the most suitable to assess the performance of the end user during a DR event.

Maximum Base Load

This is the simplest way of estimating performance in DR events, based on the ability of a system to operate at an electrical load level or below a specified level. It is a static technique that uses data, usually from the previous year, to set a target at a certain power level below which the user must maintain demand when an event is called. This demand level does not take into account current load conditions due to changes within the customer's facility. Therefore, this technique often bases the maximum base load (MBL) on previous year peaks either coincident or non-coincident with system peaks. According to PJM (KEMA 2011), this method is the suitable to evaluate the contribution of DR in the capacity market.

Meter Before/Meter After

This approach refers to performance measured versus a baseline defined by demands collected prior to deployment and similar readings during the response period. It is normally used only for fast-response programmes and shows actual load changes in real-time by reading the meter before and after response to calculate the change in demand. This method, according to PJM and NAESB, is the most apt to calculate load reduction in ancillary services such as frequency regulation and reserve events when individually interval meters are available. However, it needs demand resources with relatively flat load profiles during the time of dispatch. In case a resource has periods of ramping up or down or overall variability, this approach can misinterpret the actual level of load reduction.

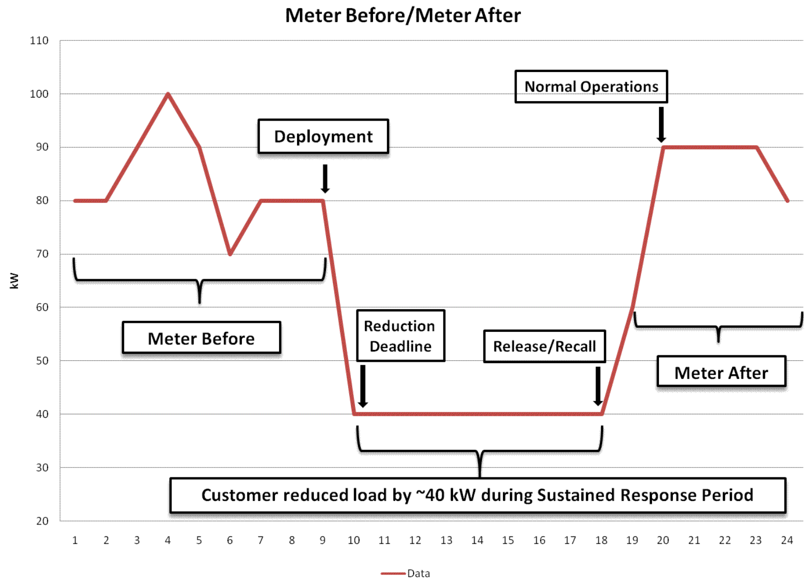

Figure 5. Smart Sensing Before/Smart Sensing After benchmarking methodology

Baseline Type I

The baseline, in this case, is produced by using historical interval meter, weather and/or calendar information. Techniques for data analysis, such as rolling averages, matching day values, and period averages are used. Rolling averages typically use historical meter data weighted on recent data and are dependant on having sufficient data to reflect representative conditions. Matching day methods find a representative day in the past, but these methods have limitations: 1) a lack of objective criteria to choose a specific day and 2) they rely upon ex-post identification. Period averaging methods generate baselines by averaging historical energy data to approximate load for specific time intervals that are "representative". These are also named High/Mid X of Y baselines where Y is the number of most recent days with X of those days having the highest load for High X of Y baselines or middle load for Mid X of Y baselines. As examples, High 4 of 5 baselines take the four highest values of the last five days. This method, according to PJM adoption is more specific to measure and verify contributions of DR in day-ahead or in real-time energy markets when all individual intervals metered are available. For a DR system that allows the aggregation of individually metered end users, an aggregate baseline can be calculated aggregating individual end users' interval load data and comparing with the aggregate observed load to establish the demand reduction. Alternatively, the aggregate demand reduction can be totalled as the sum of individual reductions, each calculated with its own baseline and registered load.

Baseline Type II

Statistical sampling is utilized for creating this baseline. It is usually applied in scenarios where aggregated meter data is available but individual site data is not. Aggregated historical meter data is the basis to model a baseline that is properly assigned to individual sites or loads that are not metered. This method is more suitable for residential DR since energy metering in commercial and industrial facilities are more cost-effective. Type II methods are often complicated and may not produce timely results leading to a lack of real-time visibility. NAESB commends the use of this method as an alternative to Baseline Type I when not all individual intervals are measured or in the case of aggregate loads. In fact, for a participant that is an aggregate of individual end users who are not all on interval meters, interval metering may be required only for a statistical sample of the end users. The baseline is calculated from the interval load data for the sample.

Experimental design

According to the experimental design, eligible participants are randomly assigned to treatment and control groups. It has been applied as an impact evaluation method and is considered akin to an application of Baseline II method. Using experimental design implies that during each DR event, a randomly selected segment of participants does not participate, remaining as a control group. This approach is interesting for programmes with large numbers of fairly homogeneous customers, primarily residential and small commercial. This method is cost-effective when the individual measurement by customer is too expensive or time-consuming. Impact estimation is calculated by aggregating all participating customers and comparing to aggregations of similar non- participating customers. To generate these load shapes the target market for the DR program needs to be specified very well. Target markets are segments of larger customer classes that share specific common characteristics. Customers in the target market that participate in the program offer are classified as program participants, while customers declining event participation are classified as non- participants (control group). Another approach is to randomly assign customers into the two groups. The average demand reduction per contributor is determined as the difference between the averages for the two groups. An alternative calculation with this design is a difference of differences method, using a baseline calculation or load model constructed for each participant, in both groups. The impact is then assessed as the difference between the participating group's modelled and observed load, minus the difference for the control group. With this method, the departure of the control group from its modelled load serves as an estimate of how the treatment group's load would have changed without a DR event.

In many situations, randomly assigning customers to different dispatch regimes is unfeasible. In these cases, comparison groups of customers considered similar to the participants after the fact are occasionally used for impact estimation. Nevertheless, without actual random assignment there will be non-considered differences between participants and nonparticipants, and these differences can alter any estimate based their comparison. The randomized control experimental design is conceptually the best evaluation approach but has been limited in its practical applications until recently. The main limitation is that most full-scale programmes regulatory frameworks do not permit random assignment of customers to participate in a programme or not. A recent exception for energy efficiency services are behaviour-based programmes presenting information to large numbers of randomly selected residential customers. Where viable, experimental design can potentially produce the most accurate results possible. The method is valuable because it could eliminate any difference between treatment and control, avoiding any bias, and with a sufficiently large samples can be very accurate. On the other hand, it is not that effective for smaller numbers of customers or heterogeneous large commercial or industrial customers bases.

When data is available on most participants, the experimental design offers many advantages:

- Since the M&V is conducted separately for each event day, participants do not have to be constantly in either treatment or control. The control group can be a different, randomly selected set of participants each time. This approach ensures that the treatment and control group are similar in all ways other than being this particular dispatch and have equivalent experience with the programme.
- For a large scale program, large control samples can offer in highly accurate results without significantly lowering the total dispatched resource. When load control programmes had to be assessed using metering samples deployed only for that purpose, samples on the order of a few hundred were enough to provide acceptable accuracy for the estimated reductions. A programme with 50,000 customers registered could need a control sample of only 1,000 customers for each event day to deliver accurate estimates of programme load reductions.
- For ex-post estimation or for settlement directly based on the metering sample, treatment-control difference accurately measures reduction without requiring explicit weather modelling. If weather modelling is used, the difference of differences method ensures that any bias in the modelling can be corrected by comparing the difference between the modelled and real load of the control group and the difference between the modelled and actual load of the dispatched group.
- For ex-ante estimation, observing event data for large numbers of both dispatched and undispatched customers provides a better modelling base than functions of weather or other conditions. This type of modelling can be very difficult in particular if extreme weather days.

- The experimental design approach can operate for a wide range of conditions, while exposing any individual user to a fairly limited number of control events.

Metering Generator Output

This method establishes the demand reduction based on generation data, assuming that all load taken served by the generator would otherwise have been on the system. It is applicable to behind-the-meter onsite generation and in combination with another performance evaluation methodology when the DR resource reduces the load in addition to its behind-the-meter generation.

Exploratory data analysis

All the literature consulted during this research underscored that the baseline estimation is a crucial aspect in M&V protocols, in particular with consumers with highly variable and climate-sensitive load, such as residential consumers. Although M&V protocols exist since 1993, those protocols have been mostly tailored for M&V of energy savings produced by an EEM and not for determining the energy or power reduction provided in a DR event. The main distinction between these two service areas is that the effect of a DR event is temporary (only a few minutes or hours) and not permanent like the results of the implementation of an EEM. This offers an advantage by permitting energy measurement also after the DR event, but on the other has the disadvantage that the impact can only be measured during few intervals (during the event). Moreover, since DR events are typically called when a demand peak is expected (e.g. on low or high temperature days), the historical data used for the baseline estimation could not be representative because the energy behaviour during the DR event corresponds to special conditions that are not common. For this reason, baselining methods for DR event prefer using recent historical data (e.g. from last 10 days prior to DR event) for estimations instead of longer periods as in energy savings assessment from EEM where at least one cycle (i.e. one season or one whole year) should be considered. In DR, having a longer period of measurements available for estimating the baseline has the advantage for replacing missing data with similar days. However, since DR events since measurements are referred to unusual conditions, it is difficult to found energy values that can replace those are missing. This problem is usual in matching day methods where a crucial goal is to find circumstances similar to event day for baseline generation. This method, along with regression analysis, is the most widespread technique for data handling. Both methods are presented in the following sections.

Day matching

Day matching takes a short historical period (from one week to sixty days) and attempts to match what the demand for an event day would have been based on the usage during the period chosen. This usually requires selecting a subset of days from the reference period and averaging them, often with an adjustment for the event day’s conditions applied to the calculated baseline. For instance, if the DR event day is a weekday, hourly data from weekdays only are utilized for calculating the baseline. Common bases for spotting match days for a given event day include:

- Similar temperature or temperature-humidity index;
- Similar system load; or
- Similar customer load at non-event hours for the individual customer.

For each participant, that customer’s load on the match day (or the average of the match days if there are multiple) operates as the baseline or reference load. Demand reductions are determined as the difference between the (average) match day and event day load at each hour. This method, also called High X of Y method, has been analysed and is recommended by the EnerNOC “Demand Response Baseline” White Paper (Enernoc 2009) and the KEMA “PJM Empirical Analysis of Demand Response Baseline Methods” (KEMA 2011) as the optimal for baseline creation in most cases. The selection of the number of days in X of Y baselines depends upon many factors and require the definition of the following aspects:

1) Look-back Window: the range of days prior to the event that is considered (i.e. the value Y).

2) Exclusion rules: certain days such as holidays, previous DR event days, weekends, thresholds and scheduled shutdowns are excluded and not considered at all (as these are not representative of "normal" operation).
3) Ratio of X to Y: the chosen subset of X days in the range of Y days is related to the properties of the DR programme and the customer's general energy usage patterns.
4) Time intervals: more frequent data capture describes load behaviours in higher detail.
5) Baseline adjustments: adjustments are based on day-of-event load conditions to improve baseline accuracy. Adjustments may also be made based upon other parameters such as weather conditions, calendar days, etc.
6) Adjustment Duration: the time period associated with the adjustment may not be representative if it is either too short or too long.
7) Multiplicative vs. additive adjustments: multiplicative reflects percentage demand comparisons and additive reflects actual differences. Additive and multiplicative adjustments both use the difference between the baseline and observed load but the additive adjustment is constant across the entire event period while the multiplicative adjustment adjusts as a percentage of loads during the event period. This can produce an adjustment more appropriate for a load shape that changes during the event period.
8) Capped vs. uncapped adjustments: a higher or lower limit set to adjustments.
9) Symmetric vs. asymmetric adjustments: symmetric adjustments can increase or decrease the baseline while asymmetric adjustments only allow adjustment in one direction.
10) Aggregation level: calculations can be performed at different aggregation levels (facility level vs. a portfolio level).

The key benefits of day matching are its simplicity and transparency. Besides, for variable loads that are not well simulated by hourly or weather models, day matching may be more precise than regression models, if the matching criteria includes characteristics of each customer's demand. On the other hand, for loads that can be quite well defined in terms of hourly loads and weather patterns, regression methods will have a tendency to be more accurate. Another drawback of Day Matching is that they are constrained to actual registered days and averages of those days. If actual historical data is not enough, accuracy would be low. As will be expanded in the next section, when availability of historical data is limited, regression models are advised since they are able to interpolate and extrapolate loads from particular observed conditions (e.g. from weather conditions). Evaluating the accuracy of a match-day estimate is more difficult than assessing the accuracy of a regression model. Testing for the fit or systematic bias is not as simple with a matching procedure as with an explicit model and is not usually included in match-day analysis. Measuring the exactitude or level of random unpredictability of a match-day estimate is also not as clear-cut. It is possible to determine a standard deviation across match-day estimates from multiple event days, but it is not apparent how much of this variability comes from differences in event-day conditions. If only a sample of customers rather than the complete population, variability across match days does not show the sampling errors. As a result, determining actual uncertainty based on those estimates is difficult.

6.1.1.7 Proxy Day Approach

The proxy day method uses the hourly loads of a specific selected day (proxy day) to exemplify the user's hourly demand during the DR event day. This proxy day needs to have the same characteristics as a DR event day. Elements typically used to choose a proxy day are maximum temperature, day of the week, weekday or weekend, etc. The majority of methods currently used limit the selection to one of the prior sixty days.

6.1.1.8 Previous Days Approach

This method determines a baseline for a DR event day with an average of hourly demand data from a subset of days prior to the DR event. The selection of those days must be searching for a similar type as the DR event day (e.g. weekend days). This way, the baseline load curve is the average per hour calculated from the user's previous loads. Figure 6 shows an example of hourly baseline constructed from average hourly demand of three equivalent days before the event day.

Hour	Days Averaged to Create Baseline			Hourly
	Day 1	Day 2	Day 3	Baseline
1	1.81	1.20	1.14	**1.38**
2	1.64	1.08	0.98	**1.23**
3	1.49	0.97	0.92	**1.13**
4	1.41	0.91	0.88	**1.07**
5	1.34	0.93	0.83	**1.03**
6	1.30	0.96	0.83	**1.03**
7	1.29	1.02	0.89	**1.07**
8	1.45	1.05	1.04	**1.18**
9	1.53	1.10	0.99	**1.21**
10	1.59	1.31	1.09	**1.33**
11	1.75	1.52	1.10	**1.46**
12	1.86	1.58	1.14	**1.52**
13	2.06	1.83	1.23	**1.71**
14	2.11	1.98	1.39	**1.83**
15	2.21	2.16	1.47	**1.95**
16	2.29	2.22	1.62	**2.04**
17	2.30	2.25	1.76	**2.11**
18	2.41	2.37	1.75	**2.17**
19	2.41	2.43	1.89	**2.24**
20	2.29	2.24	1.75	**2.09**
21	2.26	2.24	1.71	**2.07**
22	2.37	2.34	1.71	**2.14**
23	2.27	2.24	1.65	**2.05**
24	1.99	1.88	1.45	**1.77**

Hourly baseline = Average of Day 1, Day 2, Day 3

Figure 6. Indicative exemplary samples of hourly baseline construction from average electrical loads (Association of Edison Illuminating Companies (AEIC) 2009)

6.1.1.9 Average Daily Energy Usage Approach

The third methodology uses daily loads (24 hours sum) to choose the most appropriate days to include in the baseline calculation. Appropriate days are chosen based on their daily load that should be equivalent to the daily load of a *selected day*, prior to the DR event day (to be considered comparable, each daily load should be 75-100% of the daily load of the selected day). The *selected day* is selected as the most recent non-event day and the same type of day as the event day. Additionally, the ratio between the daily load of the similar days and the selected day is also considered for the selection of comparable days.

Using the same values of the previous methodology example, the last days of the same type before the event day are selected. The daily ratio among this group and the selected day is calculated as shown in the following Figure 7.

Date	Day Of Week	Daily Energy	Ratio	Acceptable Day
7/31/2006	Monday	39.792	1.307	Yes
7/28/2006	Friday	31.226	1.026	Yes
7/27/2006	Thursday	30.511	1.002	Yes
7/26/2006	Wednesday	30.647	1.007	Yes
7/25/2006	Tuesday	29.899	0.982	Yes
7/21/2006	Friday	28.995	0.952	Yes
7/20/2006	Thursday	29.373	0.965	Yes
7/19/2006	Wednesday	28.798	0.946	Yes
7/18/2006	Tuesday	32.707	1.074	Yes
7/17/2006	Monday	40.264	1.323	Yes
	Average	**32.221**		
	Selected Day	**30.445**		

Figure 7. Example of days' selection for baseline construction (Association of Edison Illuminating Companies (AEIC) 2009)

Then, like PJM methods (High 5 of 10), by averaging the hourly demand of the days with the five highest daily ratios (represented in yellow in the figure above), the baseline is determined as shown in Figure 8.

Hour	Days Averaged to Create Baseline					Hourly Baseline
	07/17/06	07/31/06	07/18/06	07/28/06	07/26/06	
1	1.49	1.20	1.34	1.14	1.12	**1.26**
2	1.46	1.08	1.18	0.98	1.01	**1.14**
3	1.29	0.97	1.07	0.92	0.95	**1.04**
4	1.21	0.91	1.00	0.88	0.87	**0.97**
5	1.11	0.93	0.97	0.83	0.86	**0.94**
6	1.08	0.96	0.97	0.83	0.88	**0.95**
7	1.10	1.02	1.02	0.89	0.90	**0.99**
8	1.18	1.05	1.06	1.04	1.03	**1.07**
9	1.29	1.10	0.99	0.99	1.15	**1.10**
10	1.46	1.31	1.12	1.09	1.26	**1.25**
11	1.61	1.52	1.22	1.10	1.24	**1.34**
12	1.65	1.58	1.23	1.14	1.33	**1.39**
13	1.68	1.83	1.39	1.23	1.40	**1.51**
14	1.94	1.98	1.63	1.39	1.50	**1.69**
15	2.00	2.16	1.62	1.47	1.50	**1.75**
16	2.01	2.22	1.74	1.62	1.50	**1.82**
17	2.02	2.25	1.80	1.76	1.63	**1.89**
18	2.23	2.37	1.80	1.75	1.66	**1.96**
19	2.22	2.43	1.87	1.89	1.68	**2.02**
20	2.29	2.24	1.82	1.75	1.56	**1.93**
21	2.03	2.24	1.60	1.71	1.42	**1.80**
22	2.18	2.34	1.59	1.71	1.55	**1.87**
23	2.07	2.24	1.46	1.65	1.45	**1.77**
24	1.64	1.88	1.22	1.45	1.23	**1.48**

Hourly baseline = Average of Day 1, Day 2, Day 3, Day 4, Day 5

Figure 8. Example of baseline construction from average loads (Association of Edison Illuminating Companies (AEIC) 2009)

Regression analysis

Regression analysis is another common technique to generate the user's load during an event day. Considering accuracy, a regression model offers the DR program advanced statistical tools to calculate a baseline, resulting in high precision. Moreover, because a regression analysis is so intricate, gaming the system becomes very difficult, promoting integrity. Unfortunately, this complexity also makes the regression less transparent to stakeholders, since they may not understand the link between their actual reduction efforts and the performance measured in their remuneration. Furthermore, the data requirements of this approach, can sometimes mean that a baseline cannot be calculated until after an event's completion, limiting knowledge of event performance in near real-time. This unbalance between accuracy and simplicity can create substantial performance issues as incentives become harder to understand. The data for the development of the baseline can be gathered in two ways:

1) Including only non- event day data for an individual customer,
2) Using a pooled data series that distinguishes between the event and non- event days.

6.1.1.10 Individual regression

Individual regression analysis looks for a regression model to each individual customer's load data for a season or year. A basic model describes demands at each hour of the day (or perhaps the average for an event timeframe) as a function of a variable (like weather data such as heating degree-days). More detailed models can allow the degree-day base to be defined by the regression best fit, and may involve calendar and day of week influence, lag terms that consider temperature over multiple hours, and humidity. Typically, the individual regression models are defined referencing loads on non-event days considering the conditions of each event day to provide an assessment of the customer's load without the DR event. The impact then is the difference between the model and the real measured load for each hour of the event period. In the case that data is available only for a sample of participating customers, the total load reduction is extrapolated from the individual customer impacts. When load data are accessible for all participating customers, there is merely the sum of the individual impacts. The individual regression model can also take into account event-day terms and be fitted across event days and non-event days. However, unless there are various event days covering a wide range of the other terms in the model, including event-day terms will not provide more insight than the average over event days of model versus observed approach explained before. Compared with pooled regression, individual regression models can lead to a higher estimation error since the spread of observed results is affected by the spread of individual responses and also by the estimation “noise”. On the other hand, if event-day impacts are assessed for each individual customer, effects can often be lost in the noise at an individual level even though effects are clear if we look across all customers. The opposite can also happen, where statistically significant effects are observed for large numbers of control group customers with no event to respond to. This pattern is proof of a systematic modelling error that would affect a pooled model just as much as it would affect the average of individual models. In general, if the same model structure is applied as an individual and as pooled, the coefficients of the pooled fit will be more or less the average coefficients of the individual fits. This equality will be precisely true if both models (individual and pooled) use the same set of control variables (e.g. degree-day base) and if the observations are carried out in the same time period with equal weights. Any bias in the individual fits will appear also in the pooled fit. Furthermore, there are other benefits in case an individual regression method is applied:

- Results are determined for each customer, which provides a foundation for deeper analysis, such as studying distributions of results rather than averages only. Individual customer results can also be linked to other customer information.
- Significant results can more easily be measured for heterogeneous groups of customers with different load patterns, since each of them is modelled separately.
- Results can be aggregated into segments that may be of interest after that initial analysis.
- When the basic regression structure cannot describe well customers’ load structure, then the customers can be identified by diagnostics and treated separately.
- Weather response conditions such as the best degree-day base can be decided separately for each customer, resulting in improved overall fit.
- Ex-ante results can be derived by fitting individual regressions to design or extreme temperature data and then aggregating the resulting estimates.
- Results can be analysed to get insight on customer engagement in programmes that stimulate behavioural changes.

6.1.1.11 Pooled regression analysis

Pooled regression analysis applies a similar model structure as individual regression analysis but creates a unique model across a large group of participants and hours. In this method, a single set of coefficients for each variable is used to describe all customers’ load pattern. With pooled analysis, it is more frequent to include event-day conditions in the regression model. As a benefit of a larger pooled sample, terms that are hard to be established for an individual customer can be estimated. When compared with individual approach, a pooled model approach has an added layer of difficulty since there will be serial correlations and patterns in the regression errors that, if are not properly accounted for, can cause estimates where precision is overestimated, especially large sets of customers are included in the regressions. Therefore, the calculated standard errors for the regression terms and associated impact estimates may be understated. Nevertheless, there are several benefits for this method:

- The coefficients use information across all customers, so influences that may be poorly estimated by each individual regression can be better determined.
- Segment level effects can be found by including segment indicators in the model, or by fitting a model for each segment.

- Overall results are obtained even if there are some customers for which the basic regression structure is not a good description.
- Ex-ante estimates can be acquired directly from the event-day terms in the model.

However, there are some drawbacks of the pooled regression method, which are the following:

- Segments of interest need to be selected in the model development stage are not easy to estimate after the fact from the basic results.
- Weather response terms are considered in aggregate, reducing model accuracy.
- The method performs best when pooling is across a more homogeneous groups of customers, such as residential or small commercial.

To summarise and compare all data analysis methods for baseline estimation analysed in this section, advantages and disadvantages for each of them are resumed in Table 8.

Table 8. List of collective data processing & analysis techniques for baseline calculation

Exploratory analysis	Advantages	Disadvantages
Previous day	Most likely the same usage pattern as the event day. Easy method for costumer to understand.	Does not take into account the effects of weather on load. The need for a baseline adjustment.
Average daily usage	Easy method for costumer to understand. Averaging eliminates the variability in load for the days to generate the average day.	An average load shape created from multiple day load shapes will fail to fully capture the usage pattern for an event day. There is need for a baseline adjustment.
Proxy day	Matches a day based on defined variables uniform with event day.	Finding a day based on the defined variables. The need for a baseline adjustment. There might not be a day to use as the proxy day.
Regression model	Concept of variable relationship is easy to understand.	Costumer understanding of the process used. The need to select the correct variables to use the model.

Baseline adjustments

To associate the calculated baseline with particular conditions of the DR event day, additional adjustments are needed to improve accuracy. Traditionally, these involve evaluating the difference between the calculated baseline and the real customer demand for some pre-event period. Once specified, the calculation that makes equals the pre-event period estimated load and the pre-event period baseline is applied to the event period. These adjustments can be based on influencing variables such as temperature, humidity, calendar data, sunrise/sunset time, event day operating conditions. The search for baseline adjustments that can be effective has to consider the DR systems participating in the programme. This set of systems have to be analysed looking for issues to be addressed in designing the programme rules (e.g. event notification). The two fundamental kinds of pre-event period baseline adjustments are:

- Additive: this approach measures the magnitude of the pre-event period load difference (positive or negative) and adds it to the baseline throughout the event period. The amount is applied to the provisional baseline load in each hour, such that the adjusted baseline will equal the observed load at a time shortly before the start of the event period. For instance, if the observed demand during an adjustment period is 10 kW above the estimated baseline, this difference of 10kW is added to the estimated baseline for each time interval during the event.
- Multiplicative or scalar: this approach uses the ratio between the pre-event estimated load and the pre-event real load to the baseline throughout the event period (e.g. If the observed demand during an adjustment period is 20% above the estimated baseline, the estimated baseline for each time interval during the event is multiplied by 120%).

The pre-event period (adjustment window) can be the same day of the event or the previous day and is the reference period to adjust the baseline matches the measured load. According to the NAESB guidance the adjustment window must begin in the four hours prior to event dispatch. Some examples of adjustment windows are:

- The hour before the event (hour -1).
- The two hours before the event (hours -1 to -2).
- The two hours that end two hours before the event (hours -3 to -4)

Moreover, in weather-sensitive DR systems (e.g. heating or cooling loads), common for residential customers, it is advised to adjust the baseline considering the registered load before the time of event notification or use as the basis for adjustment system or weather characteristic to avoid the effects of the DR event. A programme gives a day-ahead notification is more attractive to participants since it offers more time to respond to events, but in this case any day-of-event adjustment can be affected to preparatory actions, be them legitimate or manipulative. The extent and nature of these actions is difficult to measure, but should depend on the timing of the notification and the specifics of the adjustment window and methodology. Event effects during the adjustment window can occur in several ways including the following:

- Preparatory increase in response to the notification: A building is pre-cooled to a cooler than usual level from the time of event notification up to right before the event. This is a legitimate, reasonable response that makes participation easier for the building. Nevertheless, if the adjustment window uses as reference hours between notification and the event, the baseline will be inflated, and load reduction overestimated.
- Anticipatory increase prior to the notification: A building is pre-cooled to a cooler than usual level in the early morning whenever a very hot day is forecasted, making likelier a DR event. As long as some hot days do not have DR events, the pre-cooling can be expected to be reflected in at least some of the non-event days used to calculate the baseline. The more routine the pre-cooling is, and the more the baseline window and exclusion rules select for similarly hot days, the less bias will be introduced to the adjusted baseline.
- Manipulative increase: A DR asset deliberately ramps up load during the adjustment window after being notified or thinking that an event is likely. This way the baseline is artificially inflated. This behaviour is difficult to differentiate from proper preparatory or anticipatory increases.

Setting the adjustment window to end before notification limits deliberate manipulation in preparation. However, on the other hand, the earlier the adjustment window, the less accurate may be to day-of load conditions.

An alternative could be to adjust to weather conditions of the event day without allowing pre-event actions that distort the baseline. This method applies a simple regression of load on whether to compare event-day conditions during the event dispatch to the conditions during a reference before the event at the same hours. The ratio of the regression-based load estimates for the two periods gives the adjustment. This approach has the benefit of adapting to the event day weather conditions without information from a pre-event load. The drawback is that it adjusts only for weather and does not adjust for an asset’s natural behaviour or other operations on the event day.

Additive and multiplicative adjustments can be limited. For instance, asymmetric adjustments apply only if the adjustment increases the baseline (not in case of decreases). Another restriction to the extent of an adjustment is the use of a cap. For example, a customer with a 100 kW baseline exhibits demand of 130 kW before notification. Using an additive adjustment, the customer baseline would be increased by 30 kW for the whole event day. But, in the presence of a cap, that additive adjustment is limited: if the cap were 20%, then the addition would only be 20 kW. This type of adjustment might not work well for instance in peak demand for a hot day following a period of cooler weather. In this case, if the costumer has demand influenced by weather it is reasonable to presume that actual demand is substantially higher than the levels during the pre-event window. However, with a cap, the real curtailment that the customer provided would be underestimated. In conclusion, residential customers with significant weather sensitivity (very common), baselines calculated as averages of recent days have been found to be ineffective, even with day-of-event adjustments. To calculate aggregate-level reductions for programmes with considerable numbers of homogenous customers, alternatives like experimental design, or use of unit savings calculations from prior studies by regression analysis are more accurate.

Uncertainty

The measurement of any physical value contains errors since no measurement instrument can be 100% accurate. There will always be differences between observed and true energy use. In a savings-verification process, these errors prevent an exact determination of savings. This uncertainty in the savings report can be managed by controlling random errors and limiting data bias. Random errors are influenced by the quality of the measurement equipment, the measurement techniques, and the sampling procedure. Data bias is related to the quality of measurement data, assumptions made and the analysis of the information. Reducing errors usually comes with an increased M&V cost so the need for improved uncertainty should be justified by the value provided from the improved information (cost-benefit analysis). To make sure that the resultant error (uncertainty) is acceptable to the recipients of a savings report, the method for their quantification should be included in the M&V Plan. According to EVO10100 – 1:2018 (Efficiency Valuation Organization (EVO) 2018), characteristics of a savings determination process which should be carefully reviewed to manage accuracy or uncertainty are:

- **Instrumentation**: measurement equipment errors are due to the accuracy of sensors, calibration, inexact measurement, or improper meter selection installation or operation. The magnitude of such errors is largely given by manufacturer's specifications and managed by periodic re-calibration.
- **Modelling**: lack of ability to find mathematical forms that fully account for all variations in energy use. Modelling errors can be due to inappropriate functional form, exclusion of relevant variables, or use of irrelevant variables.
- **Sampling**: use of a sample of the full population of items or events to represent the entire population introduces error due to the variation in values within the population or biased sampling. Sampling may be performed in either a physical sense (i.e., only 2% of the lighting fixtures are measured) or a temporal sense (instantaneous measurement only once per hour). A common sampling precision requirement is that the load should be estimated so as to have a confidence interval that is ±10% of the estimate at a 90% confidence level.
- **Interactive effects** (beyond the measurement boundary) that are not fully included in the savings computation methodology.

In order to communicate savings in a statistically acceptable manner, *savings* need to be stated along with their *confidence* and *precision* levels. *Confidence* expresses the probability that the estimated savings will fall within the *precision* range. For example, the savings estimation process may lead to a statement such as: "the best estimate of savings is 1,000 kWh annually with a 90% probability (*confidence*) that the true-average savings value falls within ±20% of 1,000". A statistical *precision* statement without a *confidence* level has no practical use and value. The *M&V* process may yield extremely high *precision* with low *confidence*. For example, the *savings* may be stated with a *precision* of ±1%, but the associated *confidence* level may drop from 95% to 35%. Additionally, savings are regarded as statistically valid if they are large compared to the statistical variations. Specifically, the savings have to be larger than twice the standard error of the baseline value. If the variance of the baseline data is too high, the unexplained behaviour in energy use of the facility or system is excessive, and those single savings determination are not reliable. Where these criteria are not met, candidate solutions can be:

- more precise measurement equipment
- more independent variables considered in the mathematical model
- larger sample sizes
- an IPMVP Option that is less affected by unknown variables.

Application of baseline methodologies

In North America, different methodologies for baseline estimation in DR events have been empirically evaluated to measure the accuracy of baseline estimation methodologies. In the following pages, the main studies and their recommendations are presented.

California Energy Commission

The California Energy Commission (CEC) in the report "Protocol Development for Demand Response Calculation – Findings and Recommendations" (California Energy Commission 2002) compared baseline accuracy in the full range of possible baselines using real data. Interval load data was gathered from several parts of the U.S., for both curtailed and uncurtailed accounts. A total of 646 accounts were used in the analysis. For some accounts, multiple years of data were available. Methods tested were classified based on the three key characteristics of each baseline methodology:

- Data selection criteria: short, rolling windows (5 to 10 prior eligible business days) to full prior seasons of data. The rolling windows can include further constraints based on average load (e.g. five days with the highest average load out of the ten previous days);
- Estimation methods: simple averages to regression approaches using either hourly or daily temperature, degree days or temperature-humidity index (THI); and
- Adjustments: additive and multiplicative approaches based on various pre-event hours as well as a THI-based adjustment not dependent on event day load.

The analysis examined 146 combinations of data selection criteria, estimation methods and adjustments, and provided specific findings for each the three characteristics of a baseline methodology. The main conclusion was that no single method offered a comprehensive solution for all kinds of load characteristics and conditions. Nevertheless, some recommendations were indicated:

- A rolling ten-day window with an additive adjustment based on the two hours prior to event start provides the most useful and practical default baseline.
- For weather-sensitive loads, limiting the rolling window to the five highest average load days is not as effective using a baseline adjustment. THI-based adjustment is the only adjustment that avoids the distortions of pre-cooling or gaming.
- Weather regression can be effective, however, the increased data requirements, processing complexity and potential for changes at the site make these options less practical. Furthermore, simple averages with adjustments are nearly as good as weather regressions.
- Loads that are highly variable pose a challenge regardless of the baseline methodology utilized.

ERCOT Demand Side Working Group

ERCOT (ERCOT Demand Side Working Group (Freeman, Sullivan & Co.) 2012) supported an analysis of the settlement alternatives for baselines for weather-sensitive loads with short reductions. The analysis contrasted eleven methods for baseline calculation across four different data aggregation levels. The baseline methods included:

- Adjusted Day-matching approaches with and without adjustment caps (10 of 10 and 3 of 10)

- Adjusted Weather-matched baseline without adjustment cap
- Regression-based baselines (four specification types)
- Randomly assigned comparison group (means and difference in difference)
- Pre-calculated load reduction estimate tables

Baselines were evaluated on Individual AC, Aggregate AC, Household-level and Feeder data and the following recommendations were provided:

- Methods with randomly assigned control groups and large sample sizes have better performance.
- Day matching approaches were the least effective for weather-sensitive loads.
- Pre-calculated load reduction tables can produce results that on average are correct if based on estimates created using randomly assigned control groups and large sample sizes. May get it wrong for individual days, especially if they are cooler.
- Complex methods offer only slight improvement.
- Higher resolution data do not necessarily improve the accuracy of demand reduction measurement.

Southern California Edison - Methods for Short-duration events

Between 2007 and 2011, Southern California Edison (SCE) (Southern California Edison 2011) studied the feasibility of integrating short-duration dispatch events (less than 30 minutes) of its residential and commercial air conditioner cycling programme into the California ISO market for non-spinning reserve ancillary services. The demand impact evaluation, and related analyses of dispatch events using end-use and feeder-level SCADA data, demonstrated the value of short-term direct load control programmes. The study also determined that there are still technological barriers to be overcome for aggregations of small DR sources to meet the requirements of ancillary service market. The main conclusions that were reached are the following:

- Short duration events were discovered to have negligible impact on customer comfort and reduced post-event snapback.
- Since there was no pre-event notification of dispatch to participating customers and snapback was minimal, both baseline based on pre and post event load information were effective.
- While ex-ante forecast accuracy improved concurrently with calibration to realized ex-post impact estimates, inherent variability in the measurable load impact of the aggregate resources remains a barrier to wholesale market integration. Telemetry of the aggregate resource through technological developments in AMI deployment provide the most promising opportunity to deal with this barrier.

PJM

In 2011, PJM[3] sponsored an analysis of baseline methods for PJM DR programmes (KEMA 2011). This analysis graded baseline performance based on relative error and variability as well as its administrative costs. Where baselines delivered similar levels of accuracy, preference was given to baselines with a lower operation costs.

The sample of DR customers represented 39% of the total number of DR customers across PJM territory and 54% of Peak Load Contribution (PLC, the load of the customers at the time of PJM's system peak). The evaluation tested a range of baselines intended to represent the types of baselines used by ISOs today. The baselines represented a wide range of data selection criteria and estimation methods. Four of the baselines used the average load of a subset of a rolling window (e.g. high 5 of

[3] PJM Interconnection is a regional transmission organization (RTO) that coordinates the movement of wholesale electricity in all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia and the District of Columbia

10). In addition, match-day baselines, two flat baselines and two regression-based baselines were tested.

Four separate adjustment types were employed to all the baselines (where feasible and reasonable) including additive, ratio (multiplicative) and an additive, regression-based PJM weather-sensitive (WS) adjustment. The additive and ratio adjustments were the same day load-based adjustments across the industry. The PJM WS adjustment methodology gives an adjustment based on event day weather rather than event day load. This method avoids troubles related to same day load-based adjustments (e.g. early shutdown, pre-cooling) but uses a regression model characterization of weather sensitivity that entails additional data and computational complexity but only explicitly addresses weather as an independent control variable. The analysis of all these baselines methods offered the following conclusions:

- Baselines using an average load over a subset of a rolling time interval (10 of 10, high 5 of 10, high 4 of 5, middle 4 of 6) with the same day additive or multiplicative adjustment played better than any unadjusted baselines or those adjusted with the PJM WS adjustment.
- All these baselines have similar results and performed well across all segments, time periods and weather conditions except in the case of customers with variable load. Variable load customers should be segmented in order to apply distinct performance evaluation methodologies and/or market rules.
- The PJM weather-sensitive adjustment applied to the PJM economic programme high 4 of 5 baseline provided the best non-load adjusted results. This approach has the additional cost and complexity of the regression-based adjustment approach.
- PJM’s existing high 4 of 5 baseline with additive adjustment was consistently among the most accurate baselines without requiring any additional administrative cost to implement.

Even though other baseline methods showed slightly higher accuracy, PJM found that the added benefits could not justify the corresponding costs, and therefore no changes were made to the baseline method. A different conclusion would be possibly reached under a different scenario with a distinct existing baseline method and different cost considerations.