D2.4 – BEYOND PMV Methodology Specifications

Grant Agreement n°	**957020**
Project Acronym	**BEYOND**
Project Title	**A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond**
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros STAMATOUKOS
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOELEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement n° 957020.

Deliverable No.	**D2.4**
Deliverable Title	**BEYOND PMV Methodology Specifications**
Work Package	WP 2 - Use Cases, Business Requirements and Architecture Design
WP Leader	CIRCE
Due Date	31/07/2021
Actual Date of submission	30/07/2021
Version	1.6
Status	Submitted
Dissemination Level	Public
Authors	Victor Ballestin (CIRCE) Juan Aranda (CIRCE) Tasos Tsitsanis (SUITE5)
Reviewers	Tasos Tsitsanis (SUITE5) Dionysios Pramangioulis (UBITECH)

Version	Modification(s)	Date	Author(s)
1.0	Overview of previous projects and methodologies	09/04/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE)
1.1	First PMV draft	05/05/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE)
1.2	Alignment with available data in the platform	24/05/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE), Tasos Tsitsanis (SUITE5)
1.3	Alignment with BEYOND platform algorithms and incorporation of PMV KPIs	01/07/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE), Tasos Tsitsanis (SUITE5)
1.4	Completion of introduction, conclusions and executive summary	19/07/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE), Tasos Tsitsanis (SUITE5)
1.5	Revised draft por reviewers with task partners input	26/07/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE), Tasos Tsitsanis (SUITE5)
1.6	Minor modifications after feedback from the reviewers	30/07/2021	Victor Ballestin (CIRCE), Juan Aranda (CIRCE), Tasos Tsitsanis (SUITE5)

EXECUTIVE SUMMARY

This deliverable sets the BEYOND Performance Measurement & Verification (PMV) methodology which will be applied during the validation activities of the project.

First, an analysis of existing PMV and baselining methodologies has been made to understand the state of the art and learn from previous experiences and barriers encountered. The BEYOND methodology builds from these existing methodologies adapting for the characteristic target and platform architecture of the BEYOND project.

BEYOND services are destined for buildings that already generate data streams and, therefore, much of the information gathered in other PMV methodologies (historical data, contractual aspects, system information) is already integrated in the platform at the start of the services. This simplifies the process but, in turn, the method has to be flexible enough to adapt to different services (smart automation actions, renovations, DR, cultural changes) and various data availability scenarios and define an accurate baseline for the model.

These efforts have resulted in a leaner PMV structure comprised of two phases and six total steps:

- M&V Implementation
    a) Characterisation of the event
    b) Analysis of event characteristics and data availability for algorithm calibration
    c) Definition of the demand baseline
- Ex post impact assessment
    a) Demand/generation flexibility assessment
    b) Energy savings assessment
    c) Definition of the PMV report

The methodology adapts to different services and events to be measured and to different information levels in terms of historic availability, granularity and metering level.

Finally, an adequate set of Key Performance Indicators has been defined to assess building energy performance and operation under the BEYOND framework solutions. This set of KPIs covers aspects such as energy, DR and flexibility indicators, comfort economic and environment.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

Acronym	Modification(s)
PMV	Performance Measurement & Verification
DR	Demand response
EEM	Energy efficiency measure
M&V	Measurement & Verification
ESI	Energy Saving Intervention

1 INTRODUCTION

1.1 Scope and Objectives

The aim of this deliverable is to state the performance measurement and verification methodology to assess the impact of the BEYOND services, both regarding energy savings and the demand flexibility. This methodology is based on accurate short-term forecasting algorithms that enable a dynamic and continuous baselining on real-time basis that takes into account internal building changes, external weather and season conditions, and respecting the users' comfort preferences.

The comparison of demand vs calibrated model, either short-term or seasonal depending on the nature of the event, will allow us to assess the impact of the BEYOND services for the customers.

Since the BEYOND services are destined for buildings that already generate data streams, information is integrated in the platform at the start of the services and therefore the PMV methodology is more data-driven and leaner than previous projects, skipping information gathering steps. In turn, the methodology has to be very versatile and adapt to different services (smart automation actions, renovations, DR, cultural changes) and various data availability scenario to define an accurate baseline for the model.

This deliverable also states a set of adequate KPIs in order to measure the resulting impact of the implemented actions in categories such as energy, DR and flexibility indicators, comfort economic and environment.

1.2 Relation to other tasks

This document feeds from the list of business scenarios and use cases described in D2.1. The PMV methodology defined will be connected to the activities of T2.5 and the platform architecture.

This methodology will also be a first step for the detailed pilot evaluation and impact assessment performed in T7.3 to validate specific scenarios and the characteristics of each pilot case.

1.3 Structure of the document

The document includes the following contents:

- Chapter 1 with the introductory section;

- Chapter 2 with an overview of the state of the art of PMV methodologies for energy efficiency and demand response;
- Chapter 3 with an overview of the state of the art of baselining methodologies;
- Chapter 4 with a pre-analysis of platform data available at the start of the services;
- Chapter 5 with the design of the PMV phases and steps;
- Chapter 6 with the definition of the KPIs;
- Chapter 7 with the main conclusions of the deliverable.

2 M&V OVERVIEW

M&V protocols are imperative when it comes to quantify the savings produced by an Energy Efficiency Measure (EEM). Thus, the early development of M&V protocols is closely linked to the development of ESCO business models. And that is why the growing use of energy savings performance contracts (ESPC) during the 1980s and 1990s in the US (Australasian Energy Performance Contracting Association, 2004), generated a response from different associations for the elaboration of guidelines and protocols. Figure 1, shows the evolution of these methodologies in the early stages of the M&V.

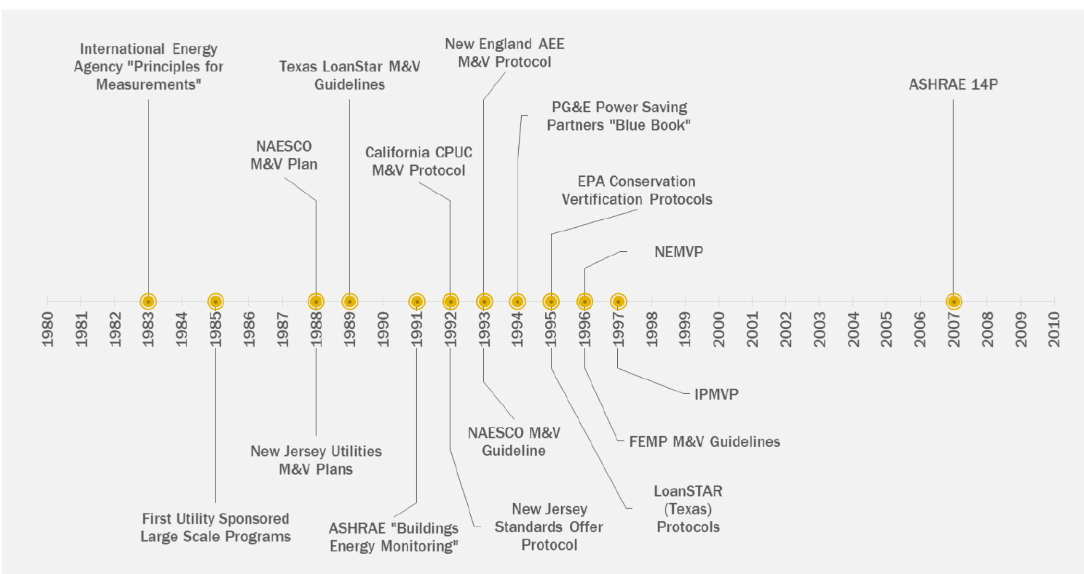

FIGURE 1. HISTORICAL EVOLUTION OF THE M&V PROTOCOLS (AUSTRALASIAN ENERGY PERFORMANCE CONTRACTING ASSOCIATION, 2004)

A defining moment occurred in 1994, when the US Department of Energy (DoE) started working with industries to address the lack of a unified and unanimous strategy for quantifying and verifying investments in energy efficiency. As a result of this joint work, in 1996, the North American Energy Measurement and Verification

Protocol (NEMVP) was published, which has been considered the first edition of a M&V protocol. Numerous companies from the USA, Canada and Mexico were implicated in the development of the methodology[1].

The international repercussion and interest on this protocol gave rise to a second edition, which was published in 1997 involving associations from twelve countries and professionals from more than 20 countries around the world. The document was renamed with the recognised title of International Performance Measurement and Verification Protocol (IPMVP) (Efficiency Valuation Organization (EVO), 2010). Although its multiple similarities with the previous version, contents related to efficiency opportunities in new construction projects and in the use of water were included on this second version.

In 2001, the third version appeared with two volumes:

- 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, it was decided to form an international non-profit organization: IPMVP Inc., to maintain and update the existing content, as well as to develop new content. In 2004, this organization was renamed as Efficiency Valuation Organization (EVO), which is the current name. So far, the published documents are continuously reviewed, and new ones are deployed. The latest English version dates from 2012 (Australasian Energy Performance Contracting Association, 2004) (EVO).

Although IPMVP is possibly the most widely used method, there are other protocols that either rely on it or share a large part of the methodology described. In 1973, with the aim of introducing a more efficient use of energy resources in government facilities, the US began a programme called the Federal Energy Management Programme (FEMP). With this as a base, in 1996, the FEMP M&V Guidelines (United States Department of Energy (DOE), 2015) was published, based on the recent NEMVP that later became the IPMVP. This methodology was thought as an IPMVP application especially oriented towards federal facilities. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) worked as well on the development of a methodology for the M&V, which result on the 2002 approved final document known as ASHRAE Guideline 14-2002 (American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2002). In this case, it was much more based on technical aspect, compared to the IPMVP.

In Europe, even though it was possible to apply the EVO's IPMVP protocol, the European Committee for Standardization (CEN) decided to publish, in 2012, the

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

standard EN 16212:2012: "Energy Efficiency and Savings Calculation, Top-down and Bottom-up Methods" (European Committee for Standardization (CEN), 2012). The main objective of this regulation is to harmonise the methods for monitoring and evaluating energy savings considering the numerous policies and actions carried out in recent years within the framework of the European Union in the field of reducing greenhouse gas emissions and energy efficiency. The document presents a general route for the calculation of energy savings in final energy consumption in buildings, cars, equipment and industrial processes, among others, to carry out *ex ante* and *ex post* evaluations in any chosen period.

The described methods, both top-down and bottom-up, were conceived within the framework of 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 an estimation of savings from indicators guessed with statistical data while the ascending method is based on actions and directives to end users to improve energy efficiency.

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), which complements the previous ISO 50001:2011 "Energy Management System" (International Organization for Standardization (ISO), 2011), in the context of M&V, key point for the energy management systems based on this standard.

In addition, this international organization has recently released the ISO 17741:2016 "General technical rules for measurement, calculation and verification of energy savings of projects" (International Organization for Standardization (ISO), 2016). In this case, energy savings are determined by comparing the consumption measured, calculated or simulated before and after the implementation of any energy-saving measure and by adjusting parameters in case of changes in relevant variables (routine adjustments) or in static factors (non-routine adjustments). In this way, the IPMVP's influence is evident in the realization of this international regulation.

In this framework the European Commission DG JRC (European Comission, s.f.) advise that performance-based projects should be subject to M&V protocols on the way of evaluating the efficiency of the energy management strategies. For these reasons and due to its international scope and its wide application within the BEYOND project, a detailed definition of a PMV methodology is required to verify the impact of the project's services.

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). The most extended and the basis for the development of the other existing protocols are briefly reviewed below as a summary of their key aspects and with the 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.

2.1 M&V methodologies used for energy efficiency assessment

2.1.1 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*. This volume explains the basic concepts and the methodology to be carried out. For this very reason it is the most important volume since it includes most of the information needed 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*. M&V methods in new building constructions and renewable energy systems are explained in detail. 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 from then on publishes the IPMVP Core Concepts, which 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 importance, the following review only addresses the most important concepts of the methodology's principles, framework and options, key to applying the M&V protocol. A principal step of the IPMVP is to define the principles of M&V on which it is based, they must be considered by any M&V plan based on this protocol:

- Accurate: the M&V reports should be precise, always taking into account the assigned budget.
- Broad: the demonstrating report of savings must take into account all aspects of a project.
- Conservative: when making estimates, it always has to underestimate the potential savings.
- Coherent: the reports must be reliable and coherent 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: measurement of the parameters of interest to determine savings and estimation of the least important or predictable.
- Transparent: all the M&V activities must be documented in detail.

Since energy savings cannot be measured directly because it involves the absence of energy consumption, the procedure to follow to estimate the savings achieved through an EEM is to compare the consumption between two periods of time. The first period is called *reference period* and is the one before the implementation of the EEM. In this period the *reference baseline* is defined, characterizing the consumption curve. Independent variables have a significant impact (e.g. outside temperature, hours of operation, occupancy, etc.). On the other hand, the period after the implementation of the EEM is called *reporting period*. Here the energy curve (called *adjusted baseline*) will be estimated based on the reference baseline found in the previous period and corrected according to some independent variables that will have a major impact (e.g., outside temperature, hours of operation, occupancy, etc.). The difference between the adjusted baseline and the actual measured consumption in the reporting period will identify the savings achieved. The IPMVP framework used to estimate energy/demand savings, is represented in the following figure.

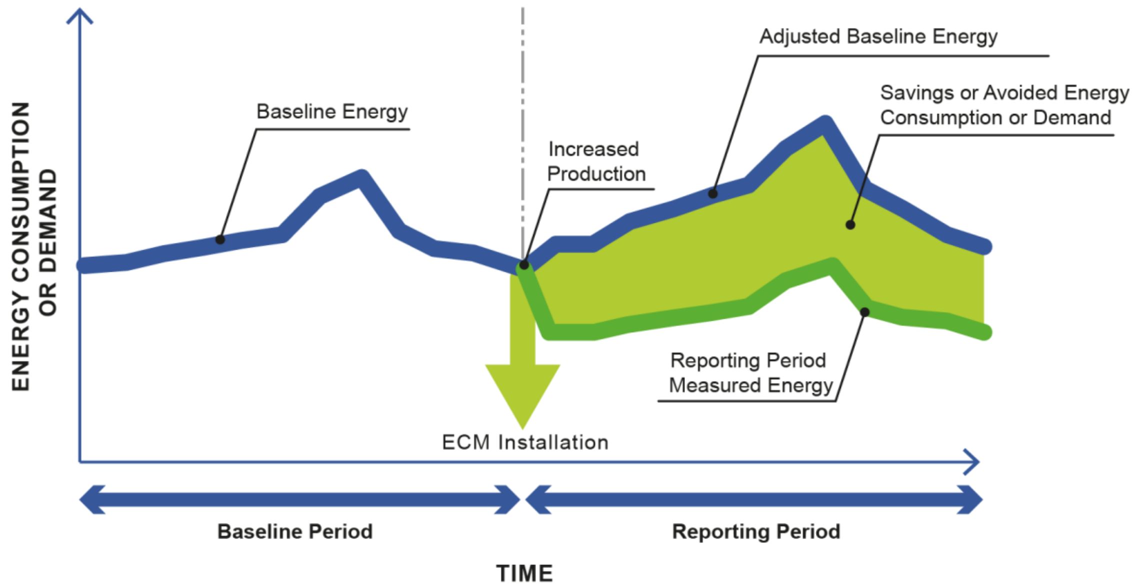

Figure 2. IPMVP framework (EVOIPMVP)

The amount of savings represented in the image above can be summarised by the following equation:

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

Aspects such as scope, available data, measurement equipment available, type of installation, budget for the implementation of M&V or the EEM itself; are key to determining potential savings, therefore, regarding their accessibility, IPMVP proposes four calculation methods:

- Option A. Retrofit isolation: key parameter measurement. It is the most economical option, but at the same time with the greatest uncertainty. Savings are determined by measuring one key parameter, 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 influence 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 whole installation 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 or energy uses in the facility.

- Option D. Calibrated simulation. The savings are determined by simulating the energy consumption of the entire installation or part of it. This simulation must be calibrated with information of the invoices or the measurement of some equipment. This option requires more advanced technical knowledge and therefore its cost is usually high. This option is designed for cases when real measurements are not available in the reference period.

A critical point to successfully develop a M&V plan is the correct selection of the measurement periods, both the reference and the reporting. For the first one, it must be guaranteed that it covers the complete operating cycle, and that it uses the period immediately before to the implementation of the EEM, since a distant period could distort the actual existing conditions. In the same way, for the second period, at least one complete operating cycle of the installation must be chosen in order to fully illustrate the effectiveness of the savings. The length of this period will depend on the user and of the savings reports. It must be taken into account that the measurement equipment has to be installed during the periods to provide the necessary data. On the other hand, if the savings based on the IPMVP serve as a basis for estimating future savings, outside the reporting period, these subsequent savings are not part of the IPMVP.

IPMVP needs to record the reference period data and all the influencing variables to have a successful determination of the savings. That is why data collecting is mandatory in a M&V Plan. This protocol collects the details of the M&V to allow a posterior consultation in a quick and simple way without risk of losing information. The M&V Plan should include the following points:

1. **Objective of the EEM**. Description of the EEM, objective pursued and the start-up procedure.

2. **Option of the IPMVP**. Definition of the IPMVP option. It will depend on the scope and the measurement limit. The date of publication, the version and the volume of the IPMVP edition should be referenced as well.

3. **Reference: period, energy and conditions**. Reference conditions and energy data is documented, including:

    - Reference period identification.

    - Data of reference consumptions.

    - Independent variables related to the energy data and their information.

- Static variables such as occupancy, operating conditions, equipment inventory, significant problems with equipment or power outages during the reference period, etc.

4. **Reference period**. Reference period should be identified.
5. **Base for adjustment**. Where the adjustment conditions for energy measurements will be defined. Both the independent variables with a significant impact on energy consumption and the static variables whose changes will need non-routine adjustments should be defined.
6. **Analysis procedure**. Analysis of the data as well as the algorithms and assumptions that will be used in the savings reports will be specified. All the elements used in the mathematical model and the validity range for the independent variables is also included.
7. **Energy prices**. The price of energy supply will be specified to assess economic results.
8. **Measurement specifications**. Where the measurement points and the characteristics of the equipment are specified, as well as the routine processes and the method against data loss.
9. **Monitoring responsibilities**. Assignment of the responsibilities of report elaboration and monitoring the energy data, independent variables and static variables recording during the reporting period.
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 uncertainty level of the measurements and the defined.
11. **Budget**. The budget and resources necessary to verify the savings is included.
12. **Report format**. The format and content of the savings report will be defined.
13. **Guarantee quality**. The quality procedures used in the saving report and during its preparation will be specified.

After having implemented the EEM, during the reporting period, the expected reports will be made with the format that was previously indicated in the M&V Plan. As an end result of the M&V, these savings reports will emerge, describing both the energy and economic savings achieved. The periodicity of the reports will be agreed in the M&V Plan and will be issued during the whole reporting period and will include saving results on a single, weekly or monthly according to the M&V Plan.

2.1.2 FEMP

The Federal Energy Management Programme (FEMP) is a program developed by the U.S. Department of Energy (DOE) focused on reducing the federal government's energy consumption that provides federal agencies with information, tools, and assistance to monitor the compliance with the requirements and energy-related objectives. FEMP establishes contracts with small companies that help in this effort (U.S. Department of Energy, 2016). FEMP (FEMP of the US Department of Energy, 2015) has identified six key actions to measure and verify savings:

1) Allocate Project Risks and Responsibilities: This is the essence of any project-specific M&V plan in such a way that, through this measure, project risk, financial, operational, and performance responsibilities are assigned between the ESCO and the customer involved.

2) Develop a Project-Specific M&V Plan: The M&V plan defines how savings will be calculated and specifies what activities should be carried out from the equipment installation. The project-specific M&V plan includes elements of the entire project and their details for each EEM.

3) Define the Baseline: It is decisive for savings to specify the baseline physical conditions (such as equipment inventory and conditions, occupancy schedule, nameplate data, equipment operating schedules, key energy parameter measurements, current weather data, control strategies, etc.). Information from this baseline is obtained through the development of surveys, inspections, spot measurements, and short-term metering activities. It is important to accurately define and document the reference conditions. It is also relevant to decide what should be monitored (and for how long) based on factors such as the measure complexity and the baseline stability, including the variability of equipment loads and operating hours as well as other variables affecting the load and data collection.

4) Install and Commission Equipment and Systems: Commissioning is achieved when it is guaranteed that the systems are designed, installed, functionally tested in all modes of operation. In such a way that they can be operated and maintained in accordance with the original design intent (appropriate lighting levels, cooling capacity, comfortable temperatures, etc.).

5) Conduct Post-Installation Verification Activities: Post-installation M&V follow-up activities are conducted to ensure that the proper equipment and systems have been installed, are working correctly and have the potential to generate the anticipated savings. Verification and monitoring methods include surveys, inspections, spot measurements, and short-term metering.

6) Perform Regular-Interval M&V Activities: M&V carries out an annual follow-up. This monitoring requires planning and coordination for the execution of M&V activities that verify the correct operation of an EEM (i.e., confirmation assumes that the EEM is running as planned). During the follow-up period, performance improvement activities can be established (e.g., recommissioning, retro-commissioning, or monitoring-based commissioning).

2.1.3 ASHRAE Guideline 14

ASHRAE Guideline 14: *Measurement of Energy, Demand and Water Savings*, is a reference for calculating energy and demand savings associated with performance contracts by taking measurements. Here, guidelines for data management and instrumentation are settle and methods for accounting for the uncertainty associated with models and measurements are described. Guideline 14 does not address other issues related to performance hiring. The ASHRAE guideline specifies three engineering approaches to M&V. Compliance with each approach requires that the overall uncertainty of the savings estimates be below prescribed thresholds. The three approaches presented are closely related to and support the options provided in IPMVP. On the other hand, Guideline 14 does not have a parallel approach to IPMVP/FEMP Option A (FEMP of the US Department of Energy, 2015).

2.1.4 The DOE Uniform Methods Project

Under the Uniform Methods Project3 (UMP), DOE developed a set of 24 protocols to determine the savings from different types of EEMs and programmes. The protocols are divided on four categories: commercial, residential, combined commercial and residential, and cross-cutting measures.

The protocols provide a simple method for evaluating gross energy savings for residential, commercial, and industrial measures offered in ratepayer-funded programmes in the United States. The measure protocols are based on a particular IPMVP option but include additional procedures necessary to add savings from individual projects to assess the impact of the entire program. For commercial measures, the FEMP guideline and the UMP are complementary. However, given that one of the objectives of M&V in a performance-based project is to ensure the correct functioning of the equipment in the long term, the FEMP guideline includes additional recommendations for annual inspection and measurements (FEMP of the US Department of Energy, 2015).

2.2 M&V methodologies used for DR assessment

M&V is the performance measurement process to quantify and validate the provision of a service in accordance with the specifications of a product. The primary role of M&V for DR is to determine the amount of energy or power that is "delivered" by a DR resource under the conditions imposed by a DR programme. The use of M&V for DR performance is the basis for determining a fair and transparent remuneration structure for market participants, a fundamental aspect that allows to generate confidence in the market. In fact, it is necessary to correctly determine the amount of flexibility delivered by a DR resource in order to provide an accurate payment according to their measured flexibility. On the other hand, a good prediction of the DR at individual and aggregated level (based on the reliability and guarantee of the DR performance measurements), allows the improvement of operational efficiency and the achievement of a more sustainable and efficient electricity system. Furthermore, measured DR performance is the main input for planning and designing a retail programme and ensure an evaluation that balances costs and benefits.

In summary, PMV for DR is used for:

- Establish resource eligibility or capacity: For most products and services that can provide DR the capability of the resource must be established before the resource can participate in the DR programme.
- DR settlement: DR settlement is the determination of DR quantities achieved, and the financial transaction between the programme or product operator and the participant, based on those quantities. For DR programmes that are incentivizing the DR provider for load reduction, the estimated load without event participation determines the calculated reduction amount that is the basis for the settlement and remuneration to each of the participants. More generally, different M&V methodologies can be used to settle between a retail programme operator and its clients or to settle that programme as an aggregated resource 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 estimation is valuable for evaluating program effectiveness and for ongoing planning.

There are a variety of arrangements that a retail operator can establish with its DR customers; many of these programme structures do not require demand reduction measurement as the basis for settlement with the retail customer or DR aggregator. However, when reduction at the programme- or segment-level is offered as a wholesale resource, the amount of the measured demand reduction for the programme or segment is usually required for wholesale settlement. For all programme types, if an impact estimation is made, its primary purpose is to determine the quantities of demand reduction achieved by the DR programme. Therefore, the application of a performance evaluation methodology to DR events consists of evaluating, against a baseline, the volume of demand variation that is sold into the market. This volume of demand flexibility is calculated as the difference between what users normally consume (the baseline) and the actual measured consumption during the dispatch event. The baseline cannot be directly measured. Thus, it must be estimated and calculated based on previously measured data combined with a robust methodology for assessing the baseline energy consumption. Consequently, measurement of any DR resource generally involves comparing observed load during the time of the curtailment to the estimated load that would otherwise have occurred without the curtailment. The difference is the load reduction (Figure 3).

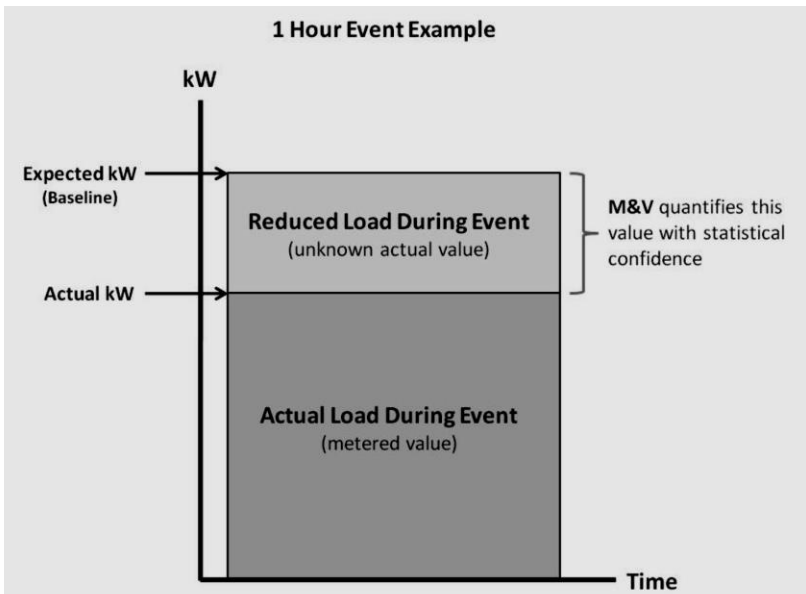

FIGURE 3. M&V QUANTIFIES LOAD REDUCTION VALUE (AEIC LOAD RESEARCH COMMITTEE, 2009)

The performance evaluation methodology used for settlement of the DR programme is crucial to the success of any DR programme. This should be useful to estimate the reduction capability and thus make the payment for the flexibility obtained from the reduction at the time of the event, which are key aspects of DR programmes where the frequency and deployment of events can lead to different types of baseline. On the other hand, in cases where pay-for-performance is measured against an absolute value, accurate measurement is essential and subsequent verification is straightforward. However, in cases where performance is measured relatively to a baseline, both defining the baseline and measuring energy are critical. The challenge is to obtain a simple but accurate estimate of reductions in a customer's energy usage relative to a baseline over a specific time interval (i.e., the DR deployment period) and make this calculation fair to all parties. Being estimates, the baselines are by definition imperfect.

However, according to NAESB (North American Energy Standards Board) recommendations, good baselines balance four main attributes:

1. **Accuracy**: giving credit to customers for the exact reduction achieved, no more, no less.
2. **Integrity**: a programme should not encourage irregular consumption and irregular consumption should not influence baseline calculations; in addition, a high level of integrity will protect against the attempts to “cheat” or defraud the system.
3. **Simplicity**: performance calculations should be easily understandable by all stakeholders, including end-users’ customers.
4. **Alignment**: DR programme designers should consider the ambitions of DR programme when choosing the baseline methodology.

It’s not easy to balance these attributes. In some situations, baselines that are resistant to manipulation are complex and difficult to calculate. In others, where the approaches are more basic, they may allow participants to exploit the baseline to their advantage. In addition, it is important to consider that baseline estimation should not reward or penalize the variation in natural load caused by system operations and generally related to variance in occupancy or local weather conditions.

In recent years, various M&V methodologies for DR have been implemented in the US context and in research projects in EU. The characteristics and specifications of these methodologies are presented in the following sections.

2.2.1 The eeMeasure methodology

As an extension of the IPMVP, the eeMeasure project evaluates two different M&V methodologies. Both are based on IPMVP and are developed from the experience of current and completed ICT PSP projects which included approximately 10,000 social dwellings and 30 public buildings (e.g. hospitals, schools) (European Union, s.f.). This is the first European project that has developed a methodology to measure and verify DR in the European context. These methodologies have been applied in three recognised H2020 projects and one FP7 project, such as NOBEL GRID, MOEEBIUS, ORBEET and Inertia, respectively.

The Residential Methodology (European Commission, 2012) is valid only to dwellings and generally accepts a monthly measurement period. In the residential sector, an assumption of constant demand (Option A of the IPMVP protocol) or a cycle of predictable demand (Option B) or another demand structure that can be accurately modelled (Option D of the IPMVP protocol) cannot usually be made. None of these assumptions applies to projects seeking to change the resident behaviour – i.e. change demand – as the means to reduce demand. Nevertheless, 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 of energy demand but is a before-after comparison instead.

Any property type can be characterized by the Non-Residential Methodology (Woodall, 2011) defined in eeMeasure (including residential) and it can be used with any data frequency. In this methodology, a process flow is defined, which directs projects to monitor appropriate variables and to create an accurate energy consumption model. A description of the underlying mathematical statistics is also included.

Option C for residential

The difference between the consumption after the Energy Saving Intervention (ESI) has been applied and the consumption under the same demand conditions without this intervention, will result in a before-after estimate of the energy savings (European Commission, 2012):

The estimation of consumption without the ESI is called *baseline data*. The baseline extension is the projection of consumption before the intervention into the period after the intervention.

The period after the intervention when the measurement of savings takes place is referred to as the *reporting period*. After the ESI intervention, energy consumption shall decrease.

The estimation of avoided consumption requires a model that is adjusted based on variations of independent variables, such as outside temperature, occupancy, household size etc. If no independent variables can be measured, the selection of a baseline period is essential for accuracy. The recommended approach is the generation of regression models that reproduce the energy consumption based on values of 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 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 necessary:

1. Nominate a time period for the creation of 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 estimated and measured consumption.

<u>Steps 1, 2 and 3. Baseline period estimation</u>

In order to compare energy saving at buildings level, it is necessary to consider the size-related energy savings of the considered units, which must also be the same for the baseline and for the reporting period. In addition, depending on the specific unit and the type of consumed energy, independent variables, such as ambient temperature, occupancy, and floor area; will significantly modify the energy savings. In cases of a considered impact in the baseline estimation, independent variables should be measured before the intervention, but if their measurement is not possible, the definition of a solid baseline period is a key step to perform an accurate M&V.

The duration of the baseline (day, week, month or year) will depend on the independent variables affecting the consumption, for instance different residential holidays' patterns or heat/cold periods. Since it is not possible to directly measure the "non-intervention consumption" that would have occurred without the event, the recommended approach is to develop regression models that reproduce the energy consumption based on values of the independent variables. Although, as well as the main dependent variable, consumption of energy, is precisely and constantly measured by smart meters, some independent variables, such as outside temperature, can also be measured automatically and reliably. Other independent variables that provide data on behaviour or attitudes related to energy and the social structure of households can be collected through tenant surveys and are subject to the GDPR legislation.

Step 4 and 5. Reporting period estimation

After the ESI and an appropriate adjustment time, the energy savings should stay stabilised for a certain period in the case where same behavioural and occupancy constrains have been satisfied. To monitor the increase or decrease of energy savings over time it is necessary to deploy the following steps:

- In the short term, energy savings can be compared weekly to check their continuity over time after the ESI, especially if the savings depend on social behaviour.
- In the long-term, it is very important to verify equipment renovations as the baseline estimations may vary.

2.2.2 Other EU Projects

The Residential eeMeasures appears in a variety of DR projects based on residential units. In the following subsections, these projects are introduced.

2.2.2.1 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 the modelling approaches (passive and active building elements) used to date and delivers innovative simulation tools. They fully understand and describe real-life building operation complexities in accurate simulation predictions that significantly reduce the "performance gap" and enhance multi-fold, continuous optimization of building energy performance as a means to further mitigate and reduce the identified "performance gap" in real-time or through retrofitting. The energy performance assessment methodology of this project is published on its website (Moeebius Project, 2016) and is based on the IPMVP and the FEMP methodologies (Federal Energy Management Program, US Department of Energy, November 2015). The Moeebius M&V consists of three phases: ex-ante analysis, implementation and M&V.

The ex-ante analysis compares the baseline and the simulation model. The baseline is characterised by:

- the analysis of the energy consumption over a sufficient period (about one year) and with sufficient resolution (hourly if possible) to identify variations in consumption;

- estimated breakdown in energy consumption according to use (e.g. lighting, heating office equipment, servers, etc.);
- independent and fixed variables that affect the energy consumption and the relevant values (i.e. degree days for heating or cooling, floor area for lighting, building opening hours, metering period length, etc.).

This data must be measured at the same time as the energy consumption data. And the calibrated simulation model that will be used for the evaluation of the gap between the expected (estimated by simulation) and the actual consumption must be defined as well.

The process first consists of identifying the energy sources and specifying the metering points, and then tracking the energy consumption (from real-time monitoring to time aggregation as day or month).

The M&V last phase calculates the KPIs' evolution and analyse & evaluates the final performance of the system to optimize energy at home/building level.

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

The innovative solution proposed by OrbEEt project seeks to facilitate public and social engagement in actions for energy efficiency by providing real-time evaluations of energy impact and organisational behaviour related to energy. The OrbEEt M&V uses Option C & D from the IPMVP and creates a methodology that combines annual bills and building sub-metering data (OrbEEt, 2016). This M&V establishes a continuous validation approach (different measurement periods) but in parallel for different loads (different load types). Periodic savings adjustments based on independent variables (as defined by the eeMeasure methodology) are needed to re-stablish the baseline demand for reported periods under a common set of conditions. Since at the beginning of the project, sub-metering information for all pilot zones, they simulate energy uses (Option D from IPMVP) when there was no data for the baseline period or when upcoming changes were expected. Energy simulation was calculated based on hourly or monthly utility billing data after installation of gas and electric meters.

Option B was applied at the next stage of measurement of energy consumption. Depending on the type of consumption which shall be compared, it is possible to have different time ranges (weekly, monthly, yearly) to define a baseline period. The definition of the baseline period for the different types of devices examined in the project is given below:

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

- Baseline period: a year period is required for baseline definition.

- Information to register: Monthly consumption.
- Independent variables (for routine adjustments): HDD or CDD and occupancy level.
- Static factors (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.

Electricity: NO HVAC systems (lighting and office equipment)

- Baseline period: a week period is required for baseline definition.
- Information to register: Week consumption (daily average).
- Independent variables (routine adjustments): Occupancy level.
- Static factors (non-routine adjustments): the facility size, the design and operation of installed equipment, the number of weekly production shifts, or the types of occupants.

The pilot sites had the information about environmental conditions (through external weather services) and occupancy levels (questionnaires to pilot representatives) and it is on these variables where routine adjustments (e.g. seasonal occupancy) are applied. Non-Routine adjustments are corrections for changes in parameters that cannot be predicted and for which a significant impact on energy use/demand is expected. Non-routine adjustments should be based on known and agreed changes to the facility:

- changes in the amount of space being heated,
- changes in the power or amount or use of equipment.
- changes in set-point conditions (lighting levels, set-point temperatures)
- changes in occupancy.

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

HOLISDER integrates a wide range of mature technologies in an open and interoperable framework, comprising the full range of tools necessary to develop the whole DR value chain. This guarantees consumer empowerment and their transformation into active market players, through the deployment of a variety of implicit and hybrid DR schemes, supported by a variety of end-user applications.

Being a combination of option B and C from IPMVP, the hybrid M&V approach for HOLISDER makes use of key methodological steps of Option B while extending it with features from option C to guard against unexpected events, such as the loss of sub-metering information, etc. Sub-metering is applied at the first stages of the baselining period of the project during the whole duration of the project; it facilitates the collection of fine-grained information from the pilot buildings. The eeMeasure methodology is enriched to follow a pooled baseline regression analysis model creating a variable relationship between event days and baseline consumption.

2.2.2.4 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 project pretends to create an end-to-end Automated Demand Response Optimization Framework, that enables the realization of new business models, allowing energy cooperatives to enter energy markets as aggregators. It prepares cooperatives with innovative and highly effective tools for the establishment of strong business practices to exploit their microgrids and dynamic VPPs as balancing and ancillary assets toward grid stability and alleviation of network constraints.

FLEXCoop brings together 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.

Although the FLEXCoop PMV methodology cannot be strictly associated to the Options offered by IPMVP, it could be said that it captures aspects from the approaches of Options B and D. In fact, it continually measures individual loads and parameters to define the baseline, as it happens in the Option B approach. On the other hand, since in FLEXCoop PMV approach the information from measurements is used to generate forecasting models and to continuously calibrate them, 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. The latter corresponds to the reporting period. The reference period is the one allowing the creation and calibration of FLEXCoop models with the minimum required data possible. In particular, the reduced amount of data needed for baseline construction and calibration is an added value of FLEXCoop PMV method since it addresses a common issue of IPMVP that is the requirement of large amount of data during a long period to achieve an accurate baseline.

2.2.2.5 PARITY project - Flexibility market platform based on blockchain and IoT paves the way for smart energy grids

PARITY is working on the integration of IoT and blockchain technologies in a local flexibility market platform. The solution also includes active network management tools to address the present ‘structural inertia’ of the distribution grid. PARITY’s solution aims to increase grid durability and efficiency, favouring the penetration of renewable energy sources in the electricity energy mix above 50 %.

In the same way as the FLEXCoop, the PARITY PMV, bases its method on the continuous measurement of individual loads and parameters that define the baseline, always contemplating the minimum reference period possible for its calibration. This baseline will be defined with algorithms specific to each DR system considered in the project.

The minimum comfort conditions considered for each event are tailored to each participant considering the different characteristics of the buildings partaking in the project.

For the elaboration of the PARITY PMV, different user-acceptance scenarios were considered in each of its steps. The level of participation in a demand response event for each user will consider both the parameters defined in their contract and the current grid conditions (normal, critical or emergency).

3 BASELINE ESTIMATION IN M&V METHODOLOGIES

Based on 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 M&V methodologies change. In fact, the fundamental aspects in the design and implementation of an M&V are mostly related to a correct definition of a baseline estimation methodology which also includes the definition of methodologies for historical data analysis, baseline adjustments and the assessment of baseline accuracy. In this part, the most diffused methodologies are collected before presenting practical experiences (and associated recommendations) from their application.

3.1 Baseline estimation methods

In Northern America organised electricity markets have acquired significant experience with explicit DR testing several PMV methodologies in many of these cases. To promote harmonization and remove market barriers for new suppliers, The North American Energy Standards Board (NAESB) (North American Standard Energy Standards Board (NAESB), 2010) has defined five types of methodologies:

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

Depending on the particular case, the most suitable method will be chosen to evaluate the performance of the end user during a DR event.

3.1.1 Maximum Base Load

This is considered the easiest way for defining performance in DR events. It refers to the ability of a resource to operate at an electrical load level or below a specified level. Using data, usually from the previous year, this static technique draws a line at a certain power level below which the customer must maintain demand when requested. Many times, this level of demand is not representative of current load conditions, as the customer makes changes within their facilities. Therefore, this technique often bases the maximum base load (MBL) on previous year peaks, even if they do not coincide with system peaks. According to PJM (KEMA, 2011), this type of baseline method is the most accurate to assess the contribution of DR in the capacity market.

3.1.2 Meter Before/Meter After

This method refers to performance measured against a baseline defined by meter readings prior to deployment and similar readings during the sustained response period. It is generally used only for fast-response programmes, as it reflects load changes in real-time, collecting the meter data before and after response with the intention of measuring the change in demand. In this way, it is the most appropriate method, according to PJM and NAESB, to evaluate load reduction in ancillary services such as frequency regulation and reserve events when individually interval meters are available. Nevertheless, it requires demand resources with flatter load profiles. If a resource has periods of high variability, the meter Before/Meter After approach might over or underestimate the actual level of load reduction even for the shorten period.

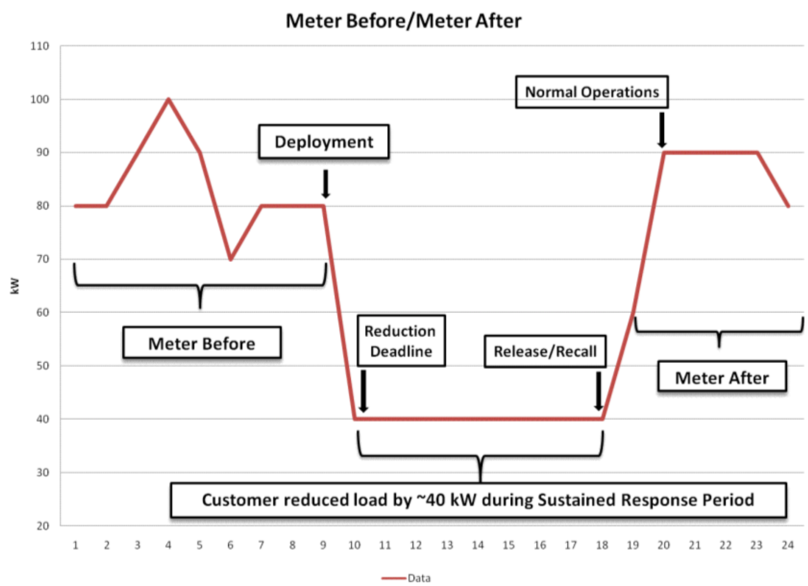

FIGURE 4. METER BEFORE/METER AFTER METHODOLOGY

3.1.3 Baseline Type I

In this case, the baseline is generated by using historical interval meter, weather and/or calendar data. The use of techniques such as rolling averages, matching day values, and period averages (see Section 3.2) are classic for this data analysis. Moving averages frequently use historical meter data weighted towards more recent data and rely on having enough data to reflect representative conditions. Matching day methods are developed by identifying a representative day in the past, but these methods suffer from: 1) a lack of objective criteria for selecting a specific day and 2) they rely upon after-the-fact identification. Period averaging methods create baselines by averaging historical energy data to estimate load for specific time intervals that are "representative" of the load. These are also known as 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 mean that are taken the highest four values of the last five days (see Section 3.2). According to PJM, it is interesting to use this method to measure and verify the contribution of DR in day-ahead or in real-time energy markets when all individual intervals metered are available. For a DR program that permits the aggregation of individually metered end users, the calculation of the aggregate baseline is feasible from the aggregate of the individual end users' interval load data and compared with the aggregate observed load to determine the demand reduction. Otherwise, the aggregate demand reduction may be calculated as the sum of individual end user reductions, each calculated from its own baseline and own actual load.

3.1.4 Baseline Type II

This method is based on the use of statistical sampling. It is often used in those scenarios where only aggregated meters are available. The baseline is created using aggregated historical meter data and it is appropriately distributed to individual sites or loads that could not be measured. This method is typically more appropriate for residential DR as commercial and industrial facilities can meter energy usage in a cost-effective way. At this point, it should be noted that the Type II methods are often more complex and may not produce suitable results leading to a lack of real-time visibility. This method is only recommended by the NAESB as an alternative to Baseline Type I when all individual intervals metered are not available or in 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 for a statistical sample of the end users. The baseline is calculated from the interval load data for the sample.

3.1.5 Experimental design

In recent years, the experimental design has also been used an impact evaluation method, through the random assignment of eligible participants to treatment and control groups, so it could be considered as an application of Baseline II method. Using experimental design means that during each DR event, a randomly selected subset of participants is not dispatched, thereby serving as a control group. This approach can be useful for programmes with large numbers of relatively homogeneous customers, primarily residential and small commercial. And it makes sense to use it when individual customer impact measuring is too expensive or time-consuming. Impact estimation is achieved by aggregating all participating customers and comparing the resultant load shapes against similar non-participating customers. To generate these load shapes a well-defined target market for the DR program is required. Target markets are segments of larger customer classes defined by specific characteristics. Customers in the target market that accept the program offer are classified as program participants, whereas those customers who reject the program offer are classified as non-participants (control group). Another way is a random assignment of customers into the two groups, one of which is "treated" and the other remains as a "control" group. The average demand reduction per participant is calculated as the difference between the averages of the two groups. An alternative calculation with this design is a difference of differences method, i.e. a baseline calculation or load model constructed for each participant, in both the dispatched and un-dispatched groups (treated and control groups, respectively). The impact is then calculated as the difference between the dispatched group's modelled and observed load, minus the corresponding difference for the control group. With this approach, the departure of the control group from its modelled load essentially provides an estimate of how the treatment group's actual load would have been higher or lower than its model, absent a DR event.

In many contexts, it is not possible to randomly assign customers to different rates or different dispatch regimes. When this happens, comparison groups of customers identified as similar to the participants after the fact are sometimes used for impact estimation. Although, without true randomization there are always unknown underlying differences between participants and nonparticipants, and these differences can bias any estimate based on comparing the groups. The randomized control experimental design is conceptually the best way of evaluation but has been limited in its practical applications until recently. The practical limitations are due to the fact that most full-scale programme applications and regulatory contexts do not allow for random assignment of customers to participate in a programme or not. A recent exception in the energy efficiency context is behaviour-based programmes offering information to large numbers of randomly selected residential customers. In feasible applications, experimental design will produce the most accurate results for estimating load reduction. This method would virtually eliminate any systematic difference between treatment and control, providing an unbiased estimate, and with a sufficiently large sample to provide the best precision, becoming thus a very valuable method. On the other hand, it would be less effective for evaluating smaller numbers of customers or large commercial or industrial customers, because the treatment-control differences will have too much random error to be reliable.

The experimental design offers many advantages when most participants have interval metered data available, including the following:

- First, since the M&V is performed separately for each event day, participants do not have to be assigned to treatment or control permanently. In fact, it is more appropriate to have the control group be a different, randomly selected set of participants for each event. Because it is ensured that the treatment and control group are the same in all ways other than being dispatched on a particular day, including that they have otherwise equivalent programme experience.

- Second, for a large-scale program, large control samples will be able to provide highly accurate results without substantially reducing the total dispatched resource. It has already been shown using measurement samples that load control programmes for samples on the order of a few hundred (depending on the level of granularity desired) were enough to provide adequate precision for the estimated reductions. So, a programme, say, with 50,000 customers enrolled could easily have a control sample of 1,000 customers for each event day producing accurate estimates of programme load reductions.
- Third, for ex-post estimation or for settlement directly based on the metering sample, it has been shown that without requiring an explicit weather modelling, a randomly assigned treatment-control difference provides a highly accurate estimate of the savings. If weather modelling is used, the difference of differences method ensures that any systematic bias in the modelling can be corrected by subtracting the difference between the modelled and actual load of the control group from the difference between the modelled and actual load of the control group of the dispatched group.
- Fourth, for ex-ante estimation, observing large numbers of both dispatched and non-dispatched customers during each event provides a much more accurate basis for modelling event effects as functions of weather or other conditions. This type of modelling can be very challenging in particular if all participants are dispatched on the few hot days.
- Fifth, as an extension of the last point, with a random control group as the basis for settlement and evaluation, calling events on every hot day does not create a problem for M&V.
- Finally, the experimental design approach can allow good load reduction estimates to be developed for a wide range of conditions, while exposing any individual customer to a limited number of control events.

3.1.6 Metering Generator Output

The generator output data in this method will determine the demand reduction, taking into account 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.

3.2 Exploratory data analysis

Pre-research studies have shown that the baseline estimation is a critical aspect in M&V protocols in particular for customers with highly variable and weather-sensitive load, such as residential. Although M&V protocols exist since 1993, those protocols have been mainly used for the M&V of energy savings derived from the implementation of an EEM and not from evaluating the energy or power reduction produced in response to a DR event. The main difference lies in the impact timeframe of these cases. While the implementation of an EEM is permanent, the effect of a DR event is temporary (a DR event affects energy loads only during few minutes or hours). This difference, on one hand has the advantage that energy measurement can also be done after the DR event, on the other has the disadvantage that the energy reduction can be measured only during few intervals (those available during the event). Furthermore, since DR events are usually called when a demand peak is foreseen (e.g. on very hot or cold days), the use of historical data for the baseline estimation should be carried out considering that historical data could be misleading because the energy behaviour during the DR event could corresponds to special conditions that do not usually happen. That is why baselines techniques for DR event prefer using recent historical data (e.g. from last 10 days prior to DR event) to make estimates instead of longer periods as in the case of energy savings assessment generated by the implementation of EEM where at least one cycle (i.e. one season or one whole year) should be deemed. In DR context, having a longer period of measurements available for estimating the baseline has the advantage that in case of lack of monitoring data due to errors or malfunctions, data from similar days in other months can be used to replace those are missing. In case of DR events since measurements are referred to unusual conditions, it is difficult to found energy values that can replace those are missing. This difficulty is typical in matching day methods where a critical aspect is to found occurrences similar to event day for their use in baseline generation. This method, together with regression analysis is the most common technique for data handling. Both are presented in the following sections.

3.2.1 Day matching

For day matching, a short historical period is taken (which can be anywhere from one week to sixty days in length) and attempted to be matched with the usage for an event day that would have been based on the usage during the historical period chosen. Typically, this involves choosing a subset of days from the historical period and averaging them, often with an adjustment for the current day's conditions applied to the calculated baseline. For example, if the DR event day occurs on a weekday, hourly data from weekdays are used in the calculation of the baseline. Common bases for identifying 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.

Each participating customer will then have a baseline or reference load that will correspond to that customer's load on the match day (or the average of the match days if there are multiple). Demand reductions are calculated 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 examined 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 best for baseline construction. However, to define the selection of the number of days it depends upon many factors and requires 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: some days are excluded from consideration such as holidays, previous DR event days, weekends, thresholds and scheduled shutdowns (as these are not representative of "normal" operation).
3) Ratio of X to Y: the selected subset of X days in the range of Y days relates to the characteristics of the DR programme and the customer's general energy usage patterns.
4) Time intervals: more frequent data capture provides greater detail about load behaviours.
5) Baseline adjustments: adjustments are based on day-of-event load conditions to improve baseline accuracy. Adjustments may also be made based upon weather, calendar days, etc.
6) Adjustment Duration: if the time period associated with the adjustment is either too short or too long, it may not be representative.
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 done at the facility level vs. at a portfolio level.

A crucial advantage of day matching is simplicity and transparency. In addition, if there are no well described variable loads by hourly or weather models, day matching may be more accurate than regression models, as long as the matching criteria include characteristics of the individual customer's load. On the other hand, for loads that can be reasonably well described in terms of hourly loads and weather patterns, regression methods will tend to be more correct. Another disadvantage of Day Matching is the reliance on historical data. When these data are not enough, the accuracy would be low. As will be seen in the next section, in those cases, regression models are recommended since effectively interpolate and extrapolate loads from particular observed conditions (e.g. from weather conditions). Assessing the accuracy of a match-day estimate is more problematic than assessing the precision of a regression model. Testing for lack of fit or systematic bias is not as straightforward with a matching procedure as with an explicit model and is not commonly included in match-day analysis. Measuring the precision or level of random variability of a match-day estimate is also not as clear-cut. It is possible to calculate a standard deviation across match-day estimates from multiple event days, but it is not clear to what extent this variability reflects differences in event-day conditions. If the analysis is done for a sample of customers rather than for the full population, changeability across different match days does not reflect the sampling errors (that is, the differences would not be equivalent to what would appear if different random samples were selected) and thus, determining the true uncertainty based on those approximations is challenging.

3.2.1.1 Proxy Day Approach

The proxy day approach chooses a single selected day, called proxy, to represent the user's hourly loads during the DR event day. The proxy day must have similar characteristics to the DR event day. Features typically used to select a proxy day include maximum temperature, day-of-week, weekday vs. weekend, etc. Most methods currently in use limit the period that may be considered when selecting the proxy day to the prior sixty days.

3.2.1.2 Previous Days Approach

This approach calculates a baseline for a DR event day by averaging hourly load data from a subset of days included in an historical period prior to the DR event. The selection of subset of days must be of the same type as the DR event day (e.g. weekend days if the event-day is during weekend, etc.). In this way, the baseline load curve is the result of the hourly values' average calculated from user's previous actual loads. In Figure 5 below, is shown an example of hourly baseline construction from average hourly loads of three equivalent days prior to the DR 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 5. EXAMPLE OF HOURLY BASELINE CONSTRUCTION FROM AVERAGE LOADS (ASSOCIATION OF EDISON ILLUMINATING COMPANIES (AEIC), 2009)

3.2.1.3 Average Daily Energy Usage Approach

By using daily loads (the sum of the 24-hourly load values for a day), this method selects the most appropriate days to be included in the baseline calculation. Suitable days are selected based on their daily load, which should be compared to the daily load of a *selected day,* prior to the DR event day (to be comparable days, each daily load should represent between the 75-100% of the daily load of the selected day). The *selected day* is chosen because it is the most recent non-DR event day and the same type of day as the DR event day. Additionally, for the selection of comparable days is also taken in consideration the ratio between the daily load of the suitable days and the selected day.

Taking the same values of the previous approach, in this one, the last days of the same type prior to the event day are selected. Once selected, as shown in the Figure 6, the daily ratio among them is calculated.

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 6. EXAMPLE OF DAYS' SELECTION FOR BASELINE CONSTRUCTION (ASSOCIATION OF EDISON ILLUMINATING COMPANIES (AEIC), 2009)

Then, the hourly baseline is calculated as shown in Figure 8. It follows a parallel way to PJM methods (High 5 of 10), by averaging the hourly load of the days with the five highest daily ratios (represented in yellow in the figure above).

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 7. EXAMPLE OF BASELINE CONSTRUCTION FROM AVERAGE LOADS (ASSOCIATION OF EDISON ILLUMINATING COMPANIES (AEIC), 2009)

3.2.2 Regression analysis

Regression analysis is another widely used technique to derive the user's load shape during an event day. From an accuracy point of view, this allows a DR program to use advanced statistical tools to calculate a baseline, leading the highest degree of accuracy. Furthermore, taking into account the complexity, the ways to game the system are reduced, which leads to greater integrity. Unfortunately, the complexity argument also makes the regression less welcoming to stakeholders, and therefore, understanding the relationship between their actual curtailment efforts and the performance for which they are credited is progressively more difficult. Added to this, appears the difficulty to calculate the baseline until the end of the events, due to the data requirements of this regression approach, which limits the ability of understanding event performance in near real-time. This point of view, based on precision versus simplicity can create significant performance issues as incentives become increasingly diffuse. The data collected to develop the baseline could be grouped 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.

3.2.2.1 Individual regression

Individual regression analysis fits a regression model to an individual customer's load data for a season or year. A basic model describes loads at each hour of the day (or perhaps the average for an event window) as a function of a variable (e.g. weather variables such as cooling degree-days). If the model is more elaborate, the cooling degree-day base can be determined by a better regression fit, which can include calendar and day of week effects, lag terms that inform about the temperature over multiple hours, and humidity. Normally, the individual regression models are fit to loads on non-event days and is applied with the conditions of each event day to give an estimate of the customer's load that would have occurred on that day without the DR event. To calculate the impact, the difference between the modelled and measured load is performed for each hour of the event period. If load data is only available for a sample of participating customers, the decrease in total load is approximated by expanding the sample that comes from the individual customer impacts. When load data are available for all participating customers, load reduction is the sum of the individual customer impacts. This individual regression model can also incorporate event-day terms and adjust to all types of them (event days and non-event days). However, unless there are multiple event days spanning a wide range of the other terms in the model, including event-day terms in individual regressions will provide no more information than the average over event days of the modelled versus observed approach explained above. Comparing individual regression models with pooled regression, in the first ones a higher level of estimation error can be given because the dispersion of the results reflects both the spread of individual responses and the estimation “noise” or random errors. On the other hand, if event-day effects are estimated for an individual customer, these individually estimated effects can often be lost in the noise even if across all customers there is an effect. The opposite can also occur, that is, statistically significant effects are found for large numbers of control group customers who had no event to respond to. In this case, there is a systematic modelling error, which would affect the 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 approximately the average coefficients of the individual fits. This equality will be even more true if both models (individual and pooled) use the same variables (e.g. degree-day base) and if the observations are carried out in the same hours and have equal weights. And the biases that could arise will be present in both models. Furthermore, the application of the individual regression method also has other advantages:

- The information is personalized for each customer. These results provide a database that contributes to richer analysis since it provides information from the observation of results distributions rather than averages only. Individual customer results can also be related to other customer information.
- Meaningful results can more easily be developed for groups of customers whose load patterns are dissimilar, since each is modelled separately.
- Results can be added into any segments that may be of interest after that initial analysis is completed.
- Customers for which the basic regression structure is not a good description can be identified by model diagnostics and treated separately.
- Weather response terms such as the best degree-day base can be determined separately for each customer, achieving better and more meaningful overall fits.
- Ex-ante results can be derived by fitting individual regressions to design or extreme temperature data and then aggregating the resulting estimates.
- Results serve to understand the relative customer engagement in programmes that promote behavioural changes and how them occur.

3.2.2.2 Pooled regression analysis

Pooled regression analysis structure is similar to the individual regression analysis model structure, with the difference that creates a single model for large group of participants and hours. In this case, to describe the average load pattern of all clients, a single set of coefficients is used and it is common to include variables from the day of the event in the regression model. With the larger pooled sample, it is possible to recognise indicators that might not be well determined for an individual customer. When compared with an individual, a pooled model approach is more complex as there will be correlations between series and patterns in the regression errors that, if are not appropriately accounted for, can finally result in estimates that seem much more precise than they really are, especially if many thousands of customers are included in the regressions. Thus, the calculated standard errors for the regression terms and associated savings estimates may be understated. Nevertheless, there are several advantages in using pooled regression method:

- The coefficients utilize information across all customers, so those effects that might be poorly estimated by each individual regression can be well determined.
- Segment level effects can be obtained by including segment indicators in the model, or by fitting the model separately by segment.
- Overall results are provided even if there are some customers for which the basic regression structure is not a good description.
- Ex-ante estimates can be obtained directly from the event-day terms in the model.
- On the other hand, the disadvantages of the pooled regression method include:
- Segments of interest need to be identified in the model development stage and cannot be easily estimated after the fact from the basic results.
- Weather response terms are estimated only in aggregate, which can reduce the model accuracy.
- The method works best when pooling is across a group of fairly similar customers, such as residential or small commercial.

Table 1 shows a summary of all the data analysis techniques for baseline estimation, as well as advantages and disadvantages of each of them.

Exploratory analysis	PRO	CON
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 takes out the variability in load for the days to create the average day.	An average load shape created from multiple day load shapes will not totally capture the usage pattern for an event day. The 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. Selecting the correct variables to use the model.

TABLE 1: RESUME OF DATA ANALYSIS TECHNIQUES FOR BASELINE ESTIMATION

3.3 Baseline adjustments

To improve and specify the baseline in relation to the observed conditions of the DR event day, new calculations should be applied to the initial estimate. Traditionally, these consist of determining the difference between the calculated baseline and the actual customer load for some pre-event period. Once characterized, the calculation that makes the pre-event period estimated load equal to the pre-event period observed baseline is applied to the event period. The factors on which these settings may be based could be temperature, humidity, calendar data, sunrise/sunset time, event day operating conditions. The types of loads participating in the DR programme will make some baseline adjustments more effective than others, and will affect issues that need to be addressed when designing the programme rules (e.g. event notification). The two basic 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 that to the baseline for the duration of the event period. The amount is applied hour by hour to the provisional baseline load, so the adjusted baseline will equal the observed load at a time just before the start of the event period (e.g. If the observed demand during an adjustment period is 20 kW above the estimated baseline, 20 kW is added to the estimated baseline for each time interval during the event).
- Multiplicative or scalar: this approach applies the ratio between the pre-event estimated load and the pre-event observed 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) is the period of time for which usually the adjusted baseline matches the measured load and it could be related to the same day of the event or the day before. Nevertheless, NAESB guidance indicates that the adjustment window shall begin no more than four hours preceding to event deployment. Examples of adjustment windows include:

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

Furthermore, it is recommended to have adjustments based on the observed load prior to the time of event notification for those cases of weather-sensitive loads (e.g. heating or cooling loads), that are common for residential customers. It is also recommended to use the weather as the basis for adjustment system or weather characteristic to avoid the effects of the DR event. By providing a day-ahead notification, attractive to those participants who want more time to respond to events, the program makes any day-of-event adjustment subject to preparatory effects, both legitimate and manipulative. The extent and nature of these effects is difficult to measure, but conceptually depends on the timing of the notification along with the specification of the adjustment window and method. Event effects during the adjustment window can occur in several ways including the following:

- <u>Preparatory increase in response to the notification</u>: From the time of event notification up to just before the event, the building is pre-cooled to a cooler than usual level. This is a legitimate, reasonable response that makes programme participation more viable for the building. However, if the adjustment window includes hours between notification and the event, the baseline will be inflated, and load reduction overstated.
- <u>Anticipatory increase prior to the notification</u>: Whenever a very hot day is forecasted, the building is pre-cooled to a cooler than usual level in the early morning, which makes probable a DR event. If 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 there will be in the adjusted baseline.
- <u>Manipulative increase</u>: A DR asset deliberately ramps up load during the adjustment window after event notification or based on its determination that an event is likely. The baseline is artificially inflated. This behaviour may be difficult to distinguish from appropriate preparatory or anticipatory increases. However, if the adjustment window is set to end before notification, the opportunities of this method can be limited. On the other hand, the earlier the adjustment window, the less effective it may be in adjusting the baseline to estimate day-of load conditions.

Another possibility could be adjustments based on weather conditions of the event day without allowing pre-event responses to distort the baseline. Through simple regression of load on weather, this method compares event-day weather conditions during the event window to the conditions during a window prior to the event at the same hours. The ratio of the regression-based load estimates for the two periods provides the adjustment. The approach has the advantage of adjusting to the event day weather conditions without requiring a pre-event load to be informative. The disadvantage is that it adjusts only for weather and does not adjust for an asset's natural, non-distorting operations on the event day.

However, both additive and multiplicative adjustments can be limited. A paradigmatic case is that of asymmetric adjustment, which only applies if the value of adjustment increases the baseline (doesn't working with decreases). In the same way, another limit to the magnitude of any adjustment, is the use of a cap. For instance, a customer with a 100 kW baseline exhibits demand of 130 kW prior to event notification. Using an additive adjustment, the customer baseline throughout that day's event would be increased by 30 kW. However, in the presence of a limit, that additive adjustment would be restricted: if the cap were 20%, then the additive adjustment would be 20 kW. In a very changeable climate or with unforeseen changes, this type of adjustment may not work well due to peak demand on a hot day after a stretch of cooler weather. In this case, if the costumer has a weather-sensitive variable demand it is reasonable to assume that actual demand is significantly higher than demand observed during the pre-event window. However, in the face of a cap, such a customer may receive little or no credit despite taking curtailment action and delivering real value to the grid. In conclusion, for residential customers with substantial weather sensitivity, baselines based on averages of recent days have been found to perform poorly, even with day-of-event adjustments. To calculate program-level reductions for programmes with large numbers of homogenous customers, effective alternatives with higher accuracy are experimental design (see Section 3.1.5), or use of unit savings calculations determined from prior studies using regression analysis.

3.4 Uncertainty

As measurement instruments are not 100% accurate, the measurement of any physical quantity includes errors. In this case, the errors appear as the difference between the observed and the true energy use, and managing them is important because in a savings-determination process, errors prevent the exact determination of savings. The uncertainty of a savings report can be managed by controlling random errors and data bias. Random errors are affected by the quality of the measurement equipment, the measurement techniques, and the design of the sampling procedure; while data bias is affected by the quality of measurement data, assumptions, and analysis. Reducing errors usually increases M&V cost so the need for improved uncertainty should be justified by the value of the improved information. For this reason, the M&V Plan should include the method of quantifying errors, to ensure that the resultant error (uncertainty) is acceptable to the users of a savings report. 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. In the vast majority, the magnitude of such errors is defined by manufacturer's specifications and is managed through periodic re-calibrations.

Modelling: Modelling errors can be due to inappropriate functional form, inclusion of irrelevant variables, or exclusion of relevant variables that result in the inability of the mathematical forms to explain variations in energy use.

Sampling: using a sample of the full population is error-inducing due to variation in values within the population or biased sampling. Sampling may be done 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). One of the most typical requirement for the definition of sampling precision 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 (outside the measurement boundary) which are not totally included in the savings computation methodology.

Levels of *confidence* and *precision* are necessary when reporting savings in a statistically valid manner. *Confidence* refers to the probability that the estimated savings will fall within the *precision* range. Thus, when it is correct, the savings estimation process will enable claims 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 (the ±20% portion) without a *confidence* level (the 90% portion) is meaningless. In this way, the *M&V* process may generate extremely high *precision* with low *confidence*. For example, the *confidence* level can drop from 95% to 35% even if the *savings* are stated with a *precision* of ±1%. Furthermore, savings are deemed to be statistically valid if they are large relative to the statistical variations. Specifically, the savings need to be larger than twice the standard error of the baseline value. If the variance of the baseline data is excessive, the unexplained random behaviour in energy use of the facility or system is high, and any single savings determination is unreliable. To address these criteria, the following possible solutions appear:

- more precise measurement equipment
- more independent variables in any mathematical model

- larger sample sizes
- an IPMVP Option that is less affected by unknown variables.

3.5 Application of baseline methodologies

At the international level, in North America, the precision of the different baseline estimation methodologies has been evaluated by empirical verification of the baseline estimation in DR. At the following, the main studies and the corresponding recommendations are presented.

3.5.1 California Energy Commission

The full range of possible baseline accuracies were compared by the California Energy Commission (CEC) in the report “Protocol Development for Demand Response Calculation – Findings and Recommendations” (California Energy Commission , 2002) using actual data. Interval load data were provided from several parts of the U.S., for both curtailed and uncurtail accounts. An amount of 646 accounts were used in the analysis and for some of them, multiple years of data were collected. Methods tested were organized based on the three key characteristics of any 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 restrictions based on average load (e.g. five days with the highest average load out of most recent ten);
- 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.

After testing 146 combinations of data selection criteria, estimation methods and adjustments, and detailing specific findings for each of the three characteristics of a baseline methodology, the overarching conclusion was that no single approach offered a comprehensive solution across all kinds of account load characteristics and conditions. However, some recommendations were made:

- The best and most practical default baseline is the one made as a rolling ten-day window with an additive adjustment based on the two hours prior to event start.
- 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, but 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.
- Highly variable loads are a challenge regardless of the baseline methodology employed.

3.5.2 ERCOT Demand Side Working Group

ERCOT (ERCOT Demand Side Working Group (Freeman, Sullivan & Co.), 2012) supported an analysis of the settlement options for baselines for weather-sensitive loads with short curtailments. The analysis compared eleven baseline calculation methods across four different levels of data aggregation. 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 different specification types
- Randomly assigned comparison group (means and difference in difference)
- Pre-calculated load reduction estimate tables

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

- The best performing methods are those with randomly assigned control groups and large sample sizes.
- Day matching approaches were the least effective approach when working with 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 err for individual days, especially if they are cooler.
- Complex methods offer limited improvement.
- It is not always true that a finer interval data improves the accuracy of demand reduction measurement.

3.5.3 Southern California Edison - Methods for Short-duration events

In the period between 2007 and 2011, Southern California Edison (SCE) (Southern California Edison , 2011) investigated the viability of integrating short-duration dispatch events (fewer than 30 minutes) of its residential and commercial air conditioner cycling programme into the California ISO market for non-spinning reserve ancillary services. This project demonstrated the value of short-term direct load control programmes and also the technological barriers to be overcome for aggregations of small DR resources to meet the requirements and need of an ancillary service market to supply electricity. The load impact evaluation and the analyses of dispatch events were made using end use and feeder level SCADA data. In general, the main conclusions found are the following:

- Short duration events had a minimal impact on customer comfort and reduced post-event snapback.
- As there was no pre-event notification of dispatch to participating customers and virtually no snapback, baseline modelling approaches that used both pre and post event load information were shown to be effective.
- Although ex-ante forecast accuracy improved at the same time as the calibration to realized ex-post impact estimates, inherent variability in the measurable load impact of the aggregate resources continues to be a challenge to wholesale market integration. Telemetry of the aggregate resource through technological developments in AMI deployment present the most promising opportunity for this barrier to be overcome.

3.5.4 PJM

In 2011, PJM[2] sponsored an analysis of baseline options for PJM DR programmes (KEMA, 2011) ranking its performance by measuring the relative error and variability as well as expected administrative costs. Where baselines delivered similar levels of accuracy, preference was given to baselines with a lower expected cost to administer.

The available 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. In the study, an attempt was made to cover and test the full range of baselines used by ISOs today, and therefore, a variety data selection criteria and estimation methods were represented. Four of the baselines were based on the average load of a subset of a rolling window (e.g. high 5 of 10), there were two kinds of match-day baselines, two flat baselines and two regression-based baselines.

Four different adjustment types were applied 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

[2] 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

same day load-based adjustments common across the industry. The PJM WS adjustment approach provides an adjustment based on event day weather rather than event day load. This approach avoids concerns related to same day load-based adjustments (e.g. early shutdown, pre-cooling) but uses a regression-based characterization of weather sensitivity that requires additional data and computational complexity while only explicitly addressing weather as a source of variability. As a summary of the found conclusions from the analysis of all these baseline methods:

- Those methods that use an average load over a subset of a rolling time period (10 of 10, high 5 of 10, high 4 of 5, middle 4 of 6) with the same day additive or multiplicative adjustment executed better than any unadjusted baselines or those adjusted with the PJM WS adjustment.
- In contrast with the rest of the baselines, the variable load customer didn't work as well as expected across all segments, time periods and weather conditions. Variable load customers should be segmented for purposes of applying a different performance evaluation methodology and/or market rule.
- Among the methods without load adjustment, the PJM weather-sensitive adjustment applied to the PJM economic programme high 4 of 5 baseline provided the best 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 and required no additional administrative cost to implement.
- While other baseline methods demonstrated slightly better accuracy (e.g., 10 of 10,), PJM found that the incremental benefits could not justify the incremental costs, and no changes were made to the baseline method. Under a different scenario with a different existing baseline method and a different range of cost considerations, it is possible a different conclusion would be met.

3.5.5 eeMeasure methodology

The eeMeasure methodology considers four specific baseline methodologies to estimate the degree of peak shaving achieved in a DR scenario (European Commission, 2012).

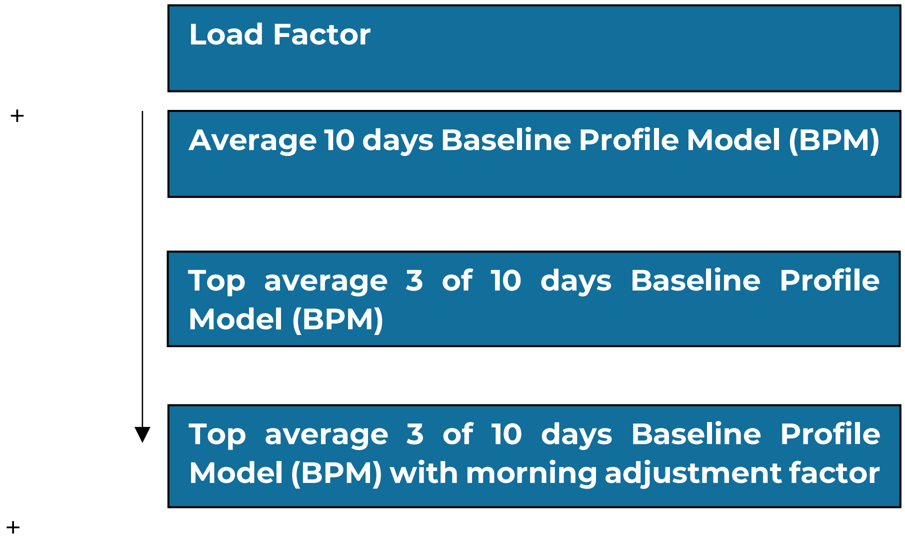

FIGURE 8. DR BASELINE METHODOLOGIES (EUROPEAN UNION, S.F.)

Load factor

The load factor (LF) is the value obtained by dividing the minimum power demand by the maximum power demand of a building:

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

If the value of the load factor is close to 1, this means that there are fewer demand curve peaks. When the building load curve peaks correspond to the electricity network peaks, shifting the LF towards 1 can represent useful peak shaving for the utility.

10 days Baseline Profile Model

Baseline profile models (BPM) are used to estimate the shaving of peaks, which arbitrarily occur on singular days, the peak “event”. To estimate non-intervention consumption at the peak event, it is generally accepted that a baseline period of 10 business days directly prior to the event reasonably represents consumption for normal operations. The reporting period is typically the 24 hours of the event day.

In this model, the average represents the non-intervention reporting period (event day) estimate and it is compared with the actual consumption on the event day to quantify the peak shaving. The 10 days of consumption are averaged as follows:

b:(d1+d2+d3+d4+d5+d6+d7+d8+d9+d10)/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 averages the 3 highest consumption figures from the previous 10 days, excluding other event days, holidays etc. The estimator 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

The model captures day-of realities in a customer load profile through an adjustment based on day-of event conditions. The estimator 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

4 Pre-analysis of platform data

The main difference between the BEYOND PMV and the previous works described above comes in the data gathering phase. Since the BEYOND services are targeted for buildings that already generate data streams before the implementation, the basis for the impact assessment will come from the big data platform developed in BEYOND.

Historical data, system information and contractual aspects will be available in the platform and feed the BEYOND PMV methodology.

Historical data

The platform will register not only consumption data but also all available variables that may affect energy demand in order to facilitate the assessment of energy reduction. These variables will also be used for the creation and auto-calibration of BEYOND forecast models' and are typically related to interior and exterior climate conditions (e.g. temperature, humidity, etc.) and to user behaviour (e.g. occupancy, schedule of electrical equipment, etc.). The forecasting algorithms will assess their dependency with energy uses affected by EEMs and DR events.

Contractual aspects

- Information on services provided and events in which the end user will participate in the two main areas of work in the project: energy efficiency and demand response.
    - Energy efficiency services: The BEYOND platform will have information on which types of energy efficiency measures may be applied on the end user, as defined in the use cases description of D2.1.

      Remuneration information for each type (i.e. if it will be done monthly, yearly and the unit price) and potential notifications for smart automation capabilities will need to be agreed and reflected in the platform.
    - Demand response: The BEYOND platform will have information on which types of DR event the end user will potentially participate, including information about their frequency or foreseen schedule during a year or along the duration of the contract between end user and service provider. 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 BEYOND solution may provide automated response to DR events

(without requirements of users' interaction), sending a notification to the users before the beginning of the event to inform them that a DR event will start is not needed, but it is recommended in order to address potential issues about user confidence and friendliness in BEYOND models.

- Minimum comfort conditions (if any): in case of participation in DR events, for each type of use that will be affected by DR events, an agreement on minimum comfort conditions that must always be maintained and stated in the platform, should be taken between the service provider and end users in order to avoid any future dissatisfaction. 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 comfort conditions defined by the users and/or inferred by the BEYOND comfort profiling engine will feed the BEYOND model to optimize the consumptions as well as the demand shifts during DR events. In addition, since comfort conditions can vary along a year, BEYOND 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 BEYOND solution on dwelling's systems. This information will be collected by BEYOND models that will automatically learn, which the optimal comfort conditions are at any time.

System information

- Systems involved: list of the main systems included in the contract (lighting, HVAC, Domestic Hot Water, Electric Vehicle charger, etc.), historic changes. The BEYOND platform will also collect information on electrical systems that will be used for participation such as nominal power, efficiency, type of technology, etc. in order to assess flexibility potential.

  A history of technological changes applied on each system will be collected. This aspect will be key for determining a stable reference period for forecasting without major changes that affect accuracy.

- Metering information: specification of data available both regarding scope (dwelling level metering or system level) and specifications.

- Actuation systems available: the platform includes information on the regulation capabilities of the consumption systems (on/off or regulation, manual or automatic, etc.), to assess flexibility and/or energy savings potential.

5 Design of the BEYOND PMV

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

The methodology proposed takes into consideration the methodologies and protocols applied in the last years, both in European projects and in the American energy markets. As seen in the previous sections, IPMVP and FEMP are the international pillars in the field of M&V protocols used to assess the impact of EEM. In the last years, EU financed projects, such as OrbEEt and Moeebius, have merged these methodologies together to obtain a new hybrid approach. On the other hand, the North American Energy Standards Board (NAESB) has been the first entity that has listed recommendations for M&V applied to DR with two main objectives, among others. First, identifying for each type of DR event the best M&V methodologies for the determination of demand flexibility quantities. Second, providing a standard terminology for the definition of measurement methods and DR events.

The BEYOND PMV methodology introduces a data-driven approach based on the BEYOND Big Data platform. Previous projects such as FEMP and Moeebius have been structured in three main phases: Ex ante analysis, Implementation and Ex post assessment. Since the BEYOND services are destined for buildings that already generate data streams, this information will be integrated in the platform at the start of the services. This data availability provided by the BEYOND platform already covers the data gathering work usually done in the Ex-ante analysis phase, resulting in a simpler, more automatic PMV process.

In this case, thanks to the algorithms and forecasting models (mainly addressing energy behaviour, comfort profiles, demand and generation forecasts of varying temporal granularity) developed in the project, the users' actions, energy behaviour and associated demand are modelled and can be predicted in a very accurate way, especially in the short-term, while increased accuracy in the prediction outcomes is expected to be achieved in longer-term predictions. Such models will be dynamic, allowing to accommodate behaviour changes or other alterations happening in buildings. A continuous learning and calibration process will allow the data-driven calibration and update of pre-trained models and profiles, along with their re-training in case of observed deviations (mainly in occupants' comfort). In more detail, BEYOND will deliver a bundle of pre-trained models (based on sampling information from the BEYOND pilot buildings or historical information available in open datasets) which can be easily adjusted and calibrated to the actual context of the BEYOND demo buildings, once data collection activities commence and information from actual buildings become available for further analysis. In this context, information flowing from sensors (such as temperature, humidity, etc.), meters and actuators in buildings will be the main inputs for calibrating the pre-trained models and monitoring their performance and accuracy. This data-driven approach is mainly based on implicit user feedback provided in the form of control actions performed over specific loads that can deliver significant energy savings and flexibility capacity and in this sense it is considered as a non-intrusive process of better understanding the constituents of building occupants' energy behaviour. As such, occupants' comfort and energy behaviour are profiled in detail and are used as inputs for forecasting demand in a bottom-up and very accurate manner. The implementation of this procedure avoids the need for control and evaluation groups of end users needed in other methodologies.

The BEYOND PMV has been structured in two main phases: M&V Implementation and Ex post impact assessment. Each of these phases is composed of several steps that are defined at following.

M&V Implementation

a) Characterisation of the event

The first step is to classify the type of event to be measured in order to assess the data needs to define a baseline. The nature of each event can need different reference periods. For instance, a retrofitting of windows in a dwelling changes energy performance both in the winter and in the summer and would need a year-long seasonal approach to characterize the initial scenario. Meanwhile, switching off lighting systems through smart automation would need a short-term approach relating consumption and occupancy, since larger periods can hinder accuracy by introducing uncontrolled variables in the environment.

In the same manner, the metering scope needs vary depending on the characteristics of the event. Impact from technological and cultural changes can be assessed with consumption data on a dwelling level, but smart-automation actions and demand response are short-term results that need system consumption data.

From the use case description started in D2.1, four categories for the classification of events to be measured in the PMV methodology have been defined. The service provider should categorize each action in one of these groups:

- Energy efficiency measures (EEM): depending on the nature of the measure, the following subclassification is defined.
    - Cultural changes: long-term behaviour changes based on BEYOND platform recommendations. The stand in comparison with consumers of similar profiles allows to identify weak points of their energy performance and make informed (and guided) decisions to alter their

energy behaviours and through a progressive interaction process as stated in UCs 6.1 and 6.6.

Implicit DR initiatives would be included in this category. Even though the savings in these price-based actions would be in cost but not energy, the effects of long-term schedule changes in the user demand would be cultural and require long-term assessment and baselining.

- Smart automation actions: short-term actions such as switching off unused systems or regulating the use of HVAC or DHW to minimize energy consumption in a human-centric manner without compromising the comfort and daily schedules of building occupants (UCs 6.5 and 6.7).
- Technical improvements (UCs 6.1, 6.3): any other technology-based energy efficiency measure like control systems, retrofitting, etc. suitable to be controlled in an Energy Performance Contract (UC 6.3)

- Demand response events: either providing demand or generation flexibility. These events are related to use cases 6.9, 6.10, 6.11 and 6.12.

Even though EEM involves some automation actions that can be measured short-term, since, for example, an EPC may involve both behavioural/technological measures (that require seasonal forecast to assess different demand scenarios) and automation actions to be implemented, the PMV will need to measure the efficiency of the measures in a longer period to verify the savings achieved (e.g. 1 year).

Short-term forecasts may be utilized to measure the efficiency of specific actions in order to assess their effectiveness for the service and utilize them even more to achieve larger energy savings, but the overall business case and performance verification shall be based on the long-term forecast performed prior to putting in action the EPC.

In order to ensure accuracy, the reference periods for baselining and minimal metering scope in each type of event would be the following:

Event type	Baselining period	Minimal metering scope/ Data source
Energy efficiency measures - Cultural	Seasonal forecast (up to 1 year)	Dwelling level / Smart Meter
Energy efficiency measures – Smart automation	Seasonal forecast (up to 1 year) *	Device-system level / Sub-metering

Energy efficiency measures - Technological	Seasonal forecast (up to 1 year)	Dwelling level / Smart Meter
Demand response	Short-term forecast (15 min to 1 hour ahead, based on data granularity)	(DR) device-system level/ Sub-metering

TABLE 2: RESUME OF DATA NEEDS FOR EACH TYPE OF EVENT

** Short-term forecast will be used internally to finetune the smart automation services and assess their effectiveness and potential for increasing energy savings, but not for verification of savings as part of an EPC*

In the end, for the definition of the baselining period and the forecasting model there will be two main types of events: energy efficiency measures and demand response events. Specific elaboration on the algorithms to be utilized for the various forecasting approaches will be performed as part of the activities of WP4 for the delivery of the pre-trained analytics models referring to demand (and generation) forecasting.

b) Analysis of event characteristics and data availability for algorithm calibration

A period for the calibration of the BEYOND models applying either at dwelling level or device/ system level (HVAC, DHW and Artificial Lighting flexibility) estimation, is needed before starting the participation in EE or DR services.

Consumption forecasts, and therefore flexibility/savings verification will be verified in the lower measurement level possible, prioritizing system level information if available and relying on dwelling level data if not available.

One of the key aspects to establish an accurate baseline is the definition of the reference period for the forecasting algorithm. This period has to be representative, covering all operation modes of the installation and immediately prior to the activation of the event.

There are three major aspects that influence the reference period selection:

- Length of historical data available
- Demand pattern changes during that period: any major modifications or outliers detected in demand patterns (as identified by analyzing the historical data from the building and the individual systems) may trigger the restart of the training process (especially if such modifications have been observed for a long period) for the forecasting models or the more detailed definition of a specific baselining period starting from the date/ period the modification has been observed (and consistently applied from that point forward). This will allow the reliance on up-to-date data and conditions with regards to the building and the systems installed or occupancy involved, so that baselining (and corresponding verification of savings) reflects as accurately as possible the latest context of the building, without considering data and information that may be outdated or not relevant to the updated context.
- Nature of the event: as stated in Table 2, energy efficiency measures need seasonal reference period to ensure the analysis of different operation modes and demand flexibility events require short-term forecast on a dwelling level since the reporting period for verification will also be very short.

In the end, events that need short term forecast will select the shortest reference period possible that ensures an accurate prediction. This is relevant to the available data granularity and is expected to range from 15 minutes to 1 hour ahead forecasting (considering also market requirements and bids placed with regards to flexibility provision), though more frequent granularity can be accommodated (even at the level of 1 minute). Events that need seasonal forecast should select the longest period available (up to 1 year) considering data availability and lack of major changes in the related systems.

The calibration of this forecast relies on this historical data to perform the analytics needed for the baselining prior to the verification period. In cases without availability of this data, the algorithms will rely on pre-trained models that can be adjusted in local contexts to perform the initial baselining.

In the first case (data availability) calibration of algorithms will be a process of a really short duration, while some additional time may be required in the second case (non-available historical data) to adjust the pre-trained algorithms.

c) Definition of the demand baseline

As mentioned before the forecasting model used for baseline definition will be defined by the type of event/service taking place.

BEYOND short-term forecasting 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 using machine learning and deep learning algorithms. Ideally the models will provide 15-min ahead forecasts. But in case such granular data is not available, the system could rely in 1-hour ahead forecasts.

This way of estimating the baseline follows the same philosophy of the approaches analysed in the State-of-the-Art 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, since it is performed continuously in the BEYOND Analytics Toolkit.

In addition, as part of the data cleaning process in the platform, setting exclusion rules within the PMV process is not needed since the BEYOND models automatically exclude outliers and erroneous values, while specific data quality mechanisms and rules are applied to fill in missing data according to the data provider instructions and expertise/ knowledge. This exclusion process is performed not only to avoid considering values representative of extraordinary users' behaviour but also to exclude from baseline estimation (especially when it comes to long-term forecasts for EEM baseline estimation) values of demand affected by the DR event. In this way, BEYOND models are able to understand when measurements should not be considered for baseline construction being not representative.

Since this approach is based on calibrated forecasting models, it is similar to the Option D of the IPMVP protocol, with the main difference, that in BEYOND PMV, the energy loads are analysed individually and not at building/dwelling level for DR purposes.

On the other hand, energy efficiency measures that require long-term assessment such as cultural and technological changes will follow a more traditional approach if historical data is available. Similar to option C of the IPMVP, the baseline will be determined by the BEYOND's deep learning and machine learning algorithms based on all the available parameters that may influence energy consumption (without though relying on simulation, but mainly on advanced Machine/ Deep Learning techniques, according to the data-driven approach followed in BEYOND to avoid generalizations and rely on real data flowing from each individual building).

Historical consumption data is a must for long-term forecasting in order to assess the building and user behaviour in different seasons. If the granularity of this data is not sufficient to feed the deep learning algorithms, the seasonal variations will be taken into account in the calibration of the pre-trained models.

Ex post impact assessment

a) Demand/generation flexibility assessment

The demand or generation flexibility provided by the prosumer will be measured simply by comparing the actual demand/generation registered during the event and the expected values predicted by the short-term baseline.

As previously mentioned, the short-term forecast models will use a very limited number of days prior to the event for the demand forecast, therefore, adjustments for potential demand pattern changes are not expected.

b) Energy savings assessment

Energy savings will be measured simply by calculating the difference between the actual energy consumption taken place in the reporting period and the expected consumption prior to the energy efficiency actions given by the energy baseline.

Potentially, ex-post non-routine adjustments may be need in the case of major demand pattern changes or force majeure events the adapt the baseline to the new operating conditions.

c) Definition of the PMV report

A PMV report will be issued for each end user after their participation in EEMs and/or DR events. It will include the explanation of the demand flexibility assessment made through the BEYOND PMV. The detailed information that the report will provide to the end user should be defined at this step of the methodology. The minimum details for each action should be:

- Energy Efficiency measures: type of action, systems involved, schedule and duration, amount of reduced consumption (kWh), comfort conditions during the event (temperature, humidity, etc.), increased amount of self-consumption rate.
- Demand response events: type (e.g., aFRR, RR, etc.), systems involved, schedule and duration, amount of reduced demand (kW or kWh), unitary price (€/kW or €/kWh), comfort conditions during the event (temperature, humidity, etc.), remuneration information.

This reporting phase will be issued to the end user through the Personalized energy analytics application in the BEYOND platform according to its preferences. In DR scenarios, sending remuneration information with high frequency should guarantee higher transparency to the programme.

6 KPIs definition

This section provides adequate Key Performance Indicators (KPIs) needed to assess the results of energy efficiency improvement and demand flexibility achieved by BEYOND.

The KPIs defined are classified in four major categories:

- Energy KPIs: focused on the results from energy efficiency measures
- Demand response and flexibility KPIs: assessing the participation in DR events and the flexibility contribution
- Comfort KPIs: measuring any non-conformity with the stablished minimum comfort conditions
- Economic KPIs: evaluating revenue, both in terms of savings and remuneration from DR participation and return of investment

The following pages detail each of these KPIs thoroughly. For each KPI, the calculation method with the formulas used and monitoring needed to determine the performance of the parameters.

Category	KPI ID	Name
Energy	EN1	Self-consumption ratio
	EN2	Energy saving
	EN3	(Buildings) Final consumption
	EN4	Total renewable energy consumption
DR and flexibility	DR1	Peak load reduction
	DR2	Aggregated flexibility provided
	DR3	Amount of flexibility requested
	DR4	Prosumer compliance per request
	DR5	Flexibility provided vs. flexibility requested ratio
Comfort	COM1	System average interruption duration
	COM2	Thermal discomfort factor
	COM3	Visual discomfort factor
Economic	EC1	Energy cost savings
	EC2	DR revenue
	EC3	Return on Investment
Environmental	ENV1	CO_2 emissions reduction
	ENV2	Indoor air quality (VOC reduction)

Table 3: List of KPIs

6.1 Energy indicators

The following tables provide Energy KPIs for quantifying the (renewable) electricity consumption/production as well as energy saving.

KPI ID	EN1
KPI Name	Self-consumption ratio
Category	Energy
Description	Measuring the efficiency of load shifting mechanisms and energy storage by quantifying the amount of electricity produced and consumed locally relative to the total electricity generation
Formula	$EN1 = \frac{EP}{TEP} \times 100$
Unit	%
Metrics / Data	EP: amount of electricity produced and consumed locally [kWh] TEP: total electricity consumption [kWh]

TABLE 4: SELF-CONSUMPTION RATIO KPI

KPI ID	EN2
KPI Name	Energy saving per building
Category	Energy
Description	Quantifying the difference between measured and reference consumption data in a building within a predefined period
Formula	$EN2 = MEC - REC$
Unit	kWh
Metrics / Data	MEC: measured energy consumption in the period for the building [kWh] REC: reference energy consumption in the period as defined by the baseline for the building [kWh]

TABLE 5: ENERGY SAVING PER BUILDING KPI

KPI ID	EN3
KPI Name	Energy saving per system
Category	Energy
Description	Quantifying the difference between measured and reference consumption data in a system within a predefined period
Formula	$EN3 = MEC - REC$
Unit	kWh
Metrics / Data	MEC: measured energy consumption in the period for the system [kWh] REC: reference energy consumption in the period as defined by the baseline for the system [kWh]

TABLE 6: ENERGY SAVING PER SYSTEM KPI

KPI ID	EN4
KPI Name	(Buildings) Final consumption
Category	Energy
Description	Quantifying the total amount of energy consumed in a building (or in a part of it) within a predefined period
Formula	$EN4 = MEC$
Unit	kWh
Metrics / Data	MEC: measured energy consumption in the period for the system [kWh]

TABLE 7: FINAL CONSUMPTION KPI

KPI ID	EN5
KPI Name	Total renewable energy consumption
Category	Energy
Description	Quantifying the total amount of renewable energy (electricity) consumed in a building (or in a part of it) within a predefined period
Formula	$EN5 = \sum_{source} EPs$
Unit	kWh
Metrics / Data	EPs: measured energy production per source [kWh]

TABLE 8: FINAL CONSUMPTION KPI

6.2 DR and Flexibility indicators

The following tables provide DR and Flexibility KPIs for tracking customers participation in flexibility programs, quantifying the aggregated flexibility and measuring the flexibility provided, as well as tracking the participation in each of these programs both in terms of percentage of requests and percentage of flexibility requested that was finally provided.

KPI ID	DR1
KPI Name	Peak load reduction
Category	DR and flexibility
Description	Calculating the demand peak reduction in comparison to the baseline value, for a period/event
Formula	$DR1 = PLMe - PLRe$
Unit	kWh
Metrics / Data	PLMe: maximum peak load measured in the event [kW] PLMe: maximum peak load expected as reference energy for the event [kW]

TABLE 9: AGGREGATED FLEXIBILITY PROVIDED KPI

KPI ID	DR2
KPI Name	Aggregated flexibility provided
Category	DR and flexibility
Description	Calculating the total flexibility provided in comparison to the baseline value, for a period/event
Formula	$DR2 = MECe - RECe$
Unit	kWh
Metrics / Data	MECe: measured energy consumption in the event [kWh] RECe: reference energy consumption in the event [kWh]

TABLE 10: AGGREGATED FLEXIBILITY PROVIDED KPI

KPI ID	DR3
KPI Name	Amount of flexibility requested
Category	DR and flexibility
Description	Registering the sum of flexibility requested to the prosumers for a period/event
Formula	$DR3 = FRe$
Unit	kWh
Metrics / Data	FRe: amount of flexibility requested per event and customer [kWh]

TABLE 11: AMOUNT OF FLEXIBILITY REQUESTED KPI

KPI ID	DR4
KPI Name	Prosumer compliance per request
Category	DR and flexibility
Description	Percentage of compliance with the flexibility requests for a customer
Formula	$DR4 = \frac{TSA}{TSD} \times 100$
Unit	%
Metrics / Data	TSA: total number of signals where the flexibility request was accepted TSD: total number of signals dispatched

TABLE 12: PROSUMER COMPLIANCE PER REQUEST KPI

KPI ID	DR5
KPI Name	Flexibility provided vs. flexibility requested ratio
Category	DR and flexibility
Description	Percentage of the flexibility requested that was finally provided for a period/event and specific consumer
Formula	$$DR5 = \frac{DR2}{DR3} \times 100$$
Unit	%
Metrics / Data	DR2: see Table 10 DR3: see Table 11

TABLE 13: FLEXIBILITY PROVIDED VS. FLEXIBILITY REQUESTED RATIO KPI

6.3 Comfort indicators

The following tables present Comfort KPIs for quantifying events where minimum comfort conditions are not met, and restoration time if minimum comfort conditions are not met.

KPI ID	COM1
KPI Name	System average interruption duration
Category	Comfort
Description	Measuring the average outage duration that any given customer would experience (average restoration time)
Formula	$$COM1 = \prod RT$$
Unit	seconds
Metrics / Data	RT: restoration time [seconds]

TABLE 14: SYSTEM AVERAGE INTERRUPTION DURATION KPI

KPI ID	COM2
KPI Name	Thermal discomfort factor
Category	Comfort
Description	Assessing incidences outside the comfort conditions regarding the thermal environment in an event and specific prosumer
Formula	$$COM2 = \frac{TTD}{TED} \times 100$$
Unit	%
Metrics / Data	TTD: time that a prosumer has been in thermal discomfort conditions on automation event TED: time of the event duration

TABLE 15: THERMAL DISCOMFORT FACTOR KPI

KPI ID	COM3
KPI Name	Visual discomfort factor
Category	Comfort
Description	Assessing incidences outside the comfort conditions regarding visual discomfort in an event and specific prosumer
Formula	$$COM3 = \frac{TVD}{TED} \times 100$$
Unit	%
Metrics / Data	TTD: time that a prosumer has been in visual discomfort conditions on automation event TED: time of the event duration

TABLE 16: VISUAL DISCOMFORT FACTOR KPI

6.4 Economic indicators

Here the Economic KPIs for all revenue generated and provides an evaluation of the economic efficiency of the measurements are described.

KPI ID	EC1
KPI Name	Energy cost savings
Category	Economic
Description	Summing up all economic savings derived from energy efficiency measures per customer
Formula	$$EC1 = 1 - \frac{CSc}{CRe} \times 100$$
Unit	%
Metrics / Data	CSc: The actual operational cost post EEM implementation CRe: The baseline operational cost (before BEYOND EEM implementation)

TABLE 17: ENERGY COST SAVINGS KPI

KPI ID	EC2
KPI Name	DR revenue
Category	Economic
Description	Summing up all the revenue from the participation in DR markets per customer
Formula	$$COM3 = \sum DRR$$
Unit	euros
Metrics / Data	DRR: DR event remuneration

TABLE 18: DR REVENUE KPI

KPI ID	EC3
KPI Name	Return on Investment
Category	Economic
Description	Evaluating the economic efficiency of energy measures for the whole building
Formula	$$COM3 = \frac{EC1 \times TEC + EC2 - SPC}{TCO}$$
Unit	years
Metrics / Data	EC1: see Table 17 TEC: total energy cost [€] EC2: see Table 18 SPC: Service Provision Cost (equipment and service subscription cost) TCO: total cost of operation [€]

TABLE 19: RETURN ON INVESTMENT KPI

6.5 Environmental indicators

The next tables list and describe the KPIs proposed for evaluation of the environmental impact.

KPI ID	ENV1
KPI Name	CO_2 emissions reduction
Category	Environmental
Description	Summing up all economic savings derived from energy efficiency measures
Formula	$ENV1 = \sum_{energy\ source} TECs \times CDRs - \sum_{energy\ source} RECs \times CDRs$
Unit	tCO_2 equivalent
Metrics / Data	TECs: total energy consumption per energy source [kWh] CDRs: carbon dioxide emissions ratio for the energy source [tCO_2/kWh] RECs: reference energy consumption per energy source [kWh]

TABLE 20: CO2 EMISSIONS REDUCTION KPI

KPI ID	ENV2
KPI Name	Indoor air quality (VOC reduction)
Category	Environmental
Description	Percentage of reduction of volatile organic compounds present indoors
Formula	$EC1 = 1 - \frac{VOCm}{VOCr} \times 100$
Unit	%
Metrics / Data	VOCm: measured volatile organic compounds concentrations after BEYOND EEM implementation [ppm] VOCr: reference volatile organic compounds concentrations before BEYOND EEM implementation [ppm]

TABLE 21: INDOOR AIR QUALITY (VOC REDUCTION) KPI

7 Conclusions

The deliverable defines a data-driven PMV methodology adapted to the target services and buildings of the project, that is, energy efficiency and flexibility solutions for buildings that already generate data streams.

In contrast with previous methodologies analysed ex-ante analysis and data gathering steps are not needed since information (historical data, contractual aspects, system information) is already available in the platform for the energy assessment and baselining. They key aspect of this methodology is being able to adapt the baselining process to different data availability scenario (historical data, metering level, granularity) and also to measure impact on both energy efficiency and flexibility solutions.

In the end a two-phase methodology was defined with six total steps. This first phase is the M&V Implementation in which the event to be measured and its characteristics are analysed to calibrate the baseline algorithm accordingly. This phase is comprised of three steps:

a) Characterisation of the event
b) Analysis of event characteristics and data availability for algorithm calibration
c) Definition of the demand baseline

The second phase is the ex-post impact assessment where the actual demand is compared to the baseline reference to determine the impact and the reporting characteristics for the customer are defined. It is also comprised of three steps:

a) Demand/generation flexibility assessment
b) Energy savings assessment
c) Definition of the PMV report

This methodology gives a fair and accurate assessment of the impact provided by the BEYOND services and can serve as a basis for remuneration schemes. A set of KPIs has been developed in terms of energy., flexibility, comfort, economic and environmental impacts to measure the PMV results.

The procedure will be integrated in the platform architecture defined in T2.5, feeding from the data available and providing the impact assessment. This task will also serve as a basis for a global evaluation framework for the BEYOND validation activities in T7.3 concerning the evaluation of energy related project aspects.

8 References

(n.d.). Retrieved from https://evo-world.org/en/about-en/history-mainmenu-en

(n.d.). Retrieved from https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp

(n.d.). Retrieved from https://evo-world.org/en/about-en/history-mainmenu-en

(n.d.). Retrieved from https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp

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

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

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

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

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

Australasian Energy Performance Contracting Association. (2004). A best practice guide to measurement and verification of energy savings. A companion document to ‘A best practice guide to energy performance contracts’.

California Energy Commission . (2002). Protocol Development for Demand Response Calculation. Draft and Findings. .

Diao, Lu, Elizondo, Mayhorn, Zhang, & Samaan. (2012). Electric Water Heater Modeling and Control Strategies for Demand Response.

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

Efficiency Valuation Organization (EVO). (2003). International performance measurement & verification protocol. Concepts and practices for determining energy savings in renewable energy technologies applications. Volume III Part II.

Efficiency Valuation Organization (EVO). (2006). International performance measurement & verification protocol. Concepts and practices for determining energy savings in new construction. Volume III Part I.

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

Efficiency Valuation Organization (EVO). (2018). Uncertainty Assessment for IPMVP - Release of a New IPMVP Application Guide.

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

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

Enernoc. (2009). The Demand Response Baseline.

European Comission. (n.d.). Retrieved from http://iet.jrc.ec.europa.eu/energyefficiency/european-energy-service-companies/monitoring-and-verification

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

European Commission. (2016). *REGULATION (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

European Commission. (n.d.). *Moeebius project*. Retrieved from http://www.moeebius.eu/

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

European Union. (n.d.). *EEmeasure*. Retrieved from http://eemeasure.smartspaces.eu/

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

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

Goldberg, M., & Agnew, G. (2013). *Measurement and Verification for Demand Response.* U.S.A.

HOLISDER. (n.d.). *HOLISDER project*. Retrieved from http://holisder.eu/

Holmberg, D., Hardin, D., & Koch, E. (2013). Towards Demand Response Measurement and Verification Standards.

Holmberg, D., Hardin, D., & Koch, E. (n.d.). *Towards Demand Response Measurement and Verification Standards.* 2013.

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

International Organization for Standardization (ISO). (2014). ISO 50015:2014 Energy management systems – Measurement and verification of energy performance of organizations – General principles and guidance.

KEMA. (2011). PJM Empirical Analysis of Demand Response Baseline Methods.

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

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

OrbEEt. (2016). *D1.2 Energy Measurements Methodology.*

OrbEEt project. (n.d.). Retrieved from http://orbeet.eu/

Sepulveda, A., Paull, L., Morsi, W., Li, H., Diduch, C., & Chang, L. (2010). A Novel Demand Side Management Program using Water Heaters and Particle Swarm Optimization.

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

The European Parliament and the Council of the European Union. (2006). 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.

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

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

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