Ref. Ares(2021)7369529 - 30/11/2021

H2020-LC-SC3-EE-2019

EUROPEAN COMMISSION

Executive Agency for Small and Medium-sized Enterprises

Grant agreement no. 893857

frESCO

New business models for innovative energy services bundles for residential consumers

Project acronym	**frESCO**
Full title	New business models for innovative energy service bundles for residential consumers
Grant agreement number	893857
Programme	H2020-EU.3.3.1. - Reducing energy consumption and carbon footprint by smart and sustainable use
Topic identifier	LC-SC3-EE-13-2018-2019-2020 - Enabling next-generation of smart energy services valorising energy efficiency and flexibility at demand-side as energy resource
Call	H2020-LC-SC3-EE-2019
Funding scheme	IA – Innovation Action
Project duration	42 months (1 June 2020 – 30 November 2023)
Project adviser	Ulrike Nuscheler - EASME
Coordinator	CIRCE – Fundacion Circe Centro de Investigacion de Recursos y Consumos Energeticos
Consortium partners	CIRCE, S5, EI-JKU, CARTIF, UBITECH, UBE, KONCAR KET, KRK, COMSA, LCTE, VOLT, VERD, IOSA, RINA-C
Website	http://fresco-project.eu
Cordis	https://cordis.europa.eu/project/id/893857

DISCLAIMER OF WARRANTIES

This document has been prepared by frESCO project partners as an account of work carried out within the framework of the EC-GA contract no. 893857.

Neither Project Coordinator, nor any signatory party of frESCO Project Consortium Agreement, nor any person acting on behalf of any of them:

(a) makes any warranty or representation whatsoever, expressed or implied,

  (i). with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or

  (ii). that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or

  (iii). that this document is suitable to any particular user's circumstance; or

(b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if the Project Coordinator or any representative of a signatory party of the frESCO Project Consortium Agreement has been informed of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n. 893857

Disclaimer: *The European Commission is not responsible for any use made of the information contained herein. The content does not necessarily reflect the opinion of the European Commission.*

Deliverable D3.4
Definition of the frESCO PMV methodology

Deliverable number	D3.4
Deliverable name	Definition of the frESCO PMV methodology
Lead beneficiary	CIRCE
Description	Definition of the frESCO PMV methodology
WP	WP3
Related task(s)	T3.4
Type	Report
Dissemination level	Public
Delivery date	30.11.2021
Main author	José Manuel Longares (CIRCE), Juan Aranda (CIRCE), Victor Ballestín (CIRCE)
Contributors	Tasos Tsitsanis (S5), Javier Antolín (CARTIF), Katerina Drivakou (UBE),

Document history

Version	Date	Changes	Author
V0 – first draft	02.11.2021	First draft for internal review	CIRCE
V3 – review draft	09.11.2021	Draft ready for partner review	CIRCE
V4 – Reviewed draft	16.11.2021	Draft reviewed	EI-JKU, CARTIF, UBE
V5 – Final draft	23.11.2021	Final draft	CIRCE
Final	30.11.2021	Final Version for submission	CIRCE

ABBREVIATIONS

Abbreviation	Meaning
AI	Artificial Intelligence
ASHRAE	The American Society of Heating, Refrigerating and Air-Conditioning Engineers
BRP	Balance Responsible Party
CA	Consortium Agreement
CDD	Cooling degree days
CEN	European Committee for Standardization
D	Deliverable
DB	Database
DER	Distributed Energy Resource
DHW	Domestic Hot Water
DoA	Description of Action
DR	Demand Response
DSM	Demand Side Management
DSO	Distribution System Operator
DOE	Department of Energy
EC	European Commission
EE	Energy Efficiency
EEM	Energy Efficiency Measure
EU	European Union
EPC	Energy Performance Contract
ESC	Energy Sales Contract
ESCO	Energy Service Company
ESI	Energy Saving Intervention
ESPC	Energy Savings Performance Contract
EV	Electric Vehicle
EVO	Efficiency Valuation Organization
FEMP	Department of Energy’s Federal Energy Management Program
FP	Framework Programme
GDPR	General Data Protection Regulation
HDD	Heating degree day

H2020	Horizon 2020 EU Framework Programme for Research and Innovation
HVAC	Heating, Ventilation and Air Conditioning
IPMVP	International Performance Measurement and Verification Protocol
IPR	Intellectual Property Right
ISO	International Organization for Standardization
MA/CP	Mitigation Action / Contingency Plan
MBL	Maximum Base Load
P4P	Pay for Performance
PMV	Performance Measurement and Verification
PV	Photovoltaic
RES	Renewable Energy Sources
SCE	Southern California Edison
SME	Small and Medium Enterprise
T	Temperature
THI	Temperature - Humidity Index
TSO	Transport System Operator
UMP	Uniform Methods Project
VPP	Virtual Power Plant
WP	Work Package

Table Of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

This document describes the frESCO Performance Measurement and Verification (PMV) methodology that must be used to clearly measure and verify the performance of the novel frESCO energy services (D3.1) in residential buildings. These services are bundled and exploited in the new business models for ESCOs and aggregators of demand flexibility (D3.3) under a Pay-for-Performance (P4P) approach. Hence, the frESCO PMV is key to deploy the new P4P energy and non-energy services.

The novel frESCO energy services utilize large amounts of data to monitor, forecast and design the best implicit and explicit strategies to obtain energy savings.

The methodology definition starts with an analysis of PMV and baseline methodologies currently in use to understand and depict the state of the art and be able to apply the knowledge to our methodology. The review covers methodologies from EVO, FEMP, ASHRAE for energy efficiency measurement and verification, as well as most recent methodological proposals for demand response assessments, carried out in several EU-funded projects. The frESCO PMV method developed in this deliverable is based on these previous methods, however, adapts them to the requirements and characteristics of the frESCO project.

The frESCO PMV methodology is data driven. In other words, it relies on the continuous collection and ingestion of near real-time dependent (energy demand) and independent variables to produce dynamic short term forecast models for event-based energy shift assessments such as Demand Response services where Flexibility is aggregated in short events triggered by market signals.

In the same way, historical data enables long term baseline adjustments for holistic energy efficiency saving measurements. This is possible by means of the on-site data collection and storage system of the frESCO data platform. Machine Learning techniques are used for baseline creation from pre-defined models. The use of the data platform entails a simplification of the process and optimizes the model to the building and user consumption profiles at all times, ensuring flexibility to adapt to the different service conditions.

The methodology differs for the two main service bundles comprising the frESCO business models: energy efficiency services and demand flexibility services.

- Energy Efficiency PMV. The EE PMV aims at measuring energy savings of the smart retrofitting and energy efficiency frESCO services. Savings are measured holistically at the dwelling or building level in long periods (billing periods). Savings are directly enjoyed by the end users and paid to the ESCO for the service delivery, proportionally to the savings obtained in the period. Savings are achieved from various implicit (behavioural changes) and explicit (automation) strategies, as well as from building and equipment retrofitting. Baselines are constructed upon pre-selected models and then trained automatically for a given period using the data flow of energy consumption and the parameters that most influence it. There could be more than one baseline to model the energy performance of the building such as seasonal baselines (three or four baselines in a year) or monthly baselines (12 baselines in a year. Baselines are fixed in the reporting period but they are subject to continuous accuracy checks that may reveal the necessity of non-routine baseline adjustments calculated periodically at every billing or reporting period.
- Demand Flexibility PMV. Demand response services offer aggregated domestic consumer demand flexibility to grid operators for congestion management and grid balancing services. These services are demanded on a short event basis and the performance is measured on the energy shift achieved by the automated operation of available Distributed Energy Resources (DER) by the aggregator. The aggregator delivers this demand flexibility to the grid and shares the market remuneration with the flexibility providers or the building users. Demand flexibility-related events need a short-term forecast so the shortest possible reference period for baseline training is selected to ensure the best baseline fit to the most recent boundary conditions and energy profiles, thus guaranteeing an accurate prediction. The baseline obtained this way is load-based and dynamic, as it is recalculated on a continuous basis as new data comes into the moving reference or training period. This way the baseline is updated to any change in external weather conditions or internal user behaviour changes, avoiding the need of continuous manual adjustments, fitting to the latest user energy profile. Finally, the baseline is calculated using the values of the independent variables just prior to the event, thus reflecting the actual conditions closest to the event.

- Performance assessment of non-energy services. Non-energy services are a set of additional value data-driven services that use the available user and building data to deliver optional benefits to the end users such as comfort, noise control, air quality or others, all under P4P contracts. The performance of these optional services is not based on energy measurements but on compliance to the contractual service levels. Measurements of the involved service parameters are compared to the target values to derive service payments from the degree of compliance with the set targets.

The frESCO PMV meets the necessary requirements for a P4P contract, such as accuracy of the measurements, fairness of the contracts under a P4P approach and the involved economic settlements, simplicity of the methodology with no or few baseline adjustments needed, replicability to any type of building and users and flexibility to adapt to a wide range of user profiles, buildings and equipment. In addition, the methodology is dynamic and uses data flows to continuously generate the baseline which fits best to the new boundary conditions and user demand profiles.

The frESCO PMV will be implemented technically in the data platform in WP4 and tested in several buildings and dwellings in the four demo-sites in WP6. Demo sites are selected to include a broad sample of climatic areas, residential building types, and usage profiles. Baseline models and training procedures will be subsequently refined with the results of these tests.

1 OBJECTIVES AND SCOPE

The main objective of this task is to establish the frESCO performance measurement and verification (PMV) methodology in order to evaluate the impact of frESCO services, both in terms of energy efficiency and flexibility, and to assess the performance of other non-energy services. This PMV methodology is the basis of the economic arrangements between service providers and service contracting parties using a Pay-for-Performance (P4P) approach.

The methodology is based on precise algorithms both in the short term and in the long term in order to create baseline models that take into account the weather, changes in the building, seasonal conditions, and user consumption profiles, always maintaining the comfort preferences of each user.

frESCO's energy services are intended for users in the residential sector, generating large data flows. This information is integrated in the frESCO data platform and accessible for the baseline calculation that is necessary to apply the PMV protocols. The established methodology has to be very versatile and able to adapt to different services and scenarios in order to achieve an accurate baseline. Additionally, the protocol meets the requirements of transparency, accuracy, fairness and cost-effectiveness.

This task of PMV methodology development follows after frESCO energy services, service value chain, revenue streams, and service provider business models are defined, and occurs prior to business model simulation and contractual and legal arrangements. The methodology ensures the full deployment of Pay for Performance type of energy contracts that enable a fair and transparent payment and remuneration of energy services along the value chain. The results contained in D3.1, D3.2 and D3.3 are used in this document to design optimal PMV methodology that will be tested in the demo sites and in the simulation tests of T3.5. The implementation details of the PMV should be reflected in the P4P contracts to be developed in T3.6.

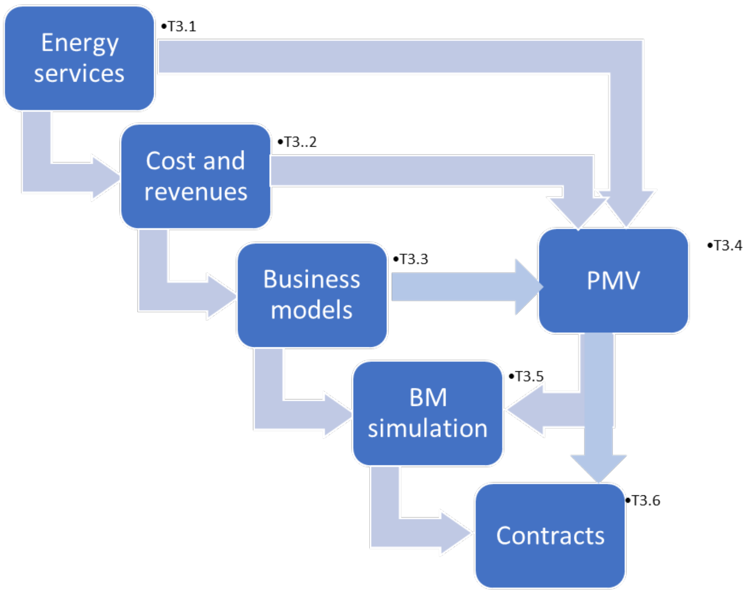

Figure 1. WP3 task dependencies and relationship.

This deliverable describes the methodology to verify performance in the three main types of frESCO energy services: energy efficiency, energy flexibility and other non-energy services. However, the proposal of mathematical baseline and forecasting models that may be used for the measurement and verification of energy savings and energy shifts is out of the scope of this document. The approach and calculation methodology are common and generic but forecast models are generated based on building energy data on a case-by-case basis. The independent variables suggested are defined by the energy service and type of energy demand. Forecasting and baselining analytics are in the scope of tasks T4.4 event-based and flexibility short term forecasting and T5.1 holistic building demand and generation forecast for long term efficiency. Baseline calculation details are not in the scope of the methodological PMV framework.

2 INTRODUCTION: THE ROLE OF THE PMV METHODOLOGY IN THE NEW FRESCO P4P ENERGY SERVICES

This document provides an overview of the frESCO Measurement and Verification (PMV) procedure that is on the basis of the Pay-for-Performance innovative energy services for domestic consumers. A thorough review of existing PMV methodologies in application for energy savings and under development for energy flexibility is carried out in section 3. Several demand flexibility assessment methodologies have been proposed in previous EC-funded projects and partially replicated by frESCO. In the same way, several Demand Response measurement experiences in the US are scrutinised for their relevance in frESCO.

One of the main issues when applying PMV methodologies is how baselines are estimated. Baselines are mathematical models that describe the energy performance of a building or system as a function of one or more contextual dependent variables. In annex 1, a review of the most important baseline estimation methods is carried out, using different exploratory data analysis and baseline adjustments.

Section 4 is devoted to describing the frESCO platform data type needs and availability, prior to the design of the frESCO PMV methodology in section 5 for energy saving and for demand flexibility verification. The performance assessment for non-Energy services is also assessed briefly in this section.

Finally, some conclusions regarding the frESCO PMV methodology are presented in section 6. The following introductory subsections link the frESCO PMV with the energy services and the Pay-for-Performance (P4P) nature of the novel service contracts.

2.1 P4P energy service approach and complementarity with current EPC.

The P4P contracts require a direct relation between service payment and energy performance. Hence, this performance must be accurately measured and verified. Indeed, P4P energy services are based on a specific Measurement and Verification Methodology (PMV) that uses real-time data streams to ensure (i) objective validation and assessment of the feasibility and effectiveness of the new business models and (ii) transparent remuneration of the involved actors for the achievement of energy savings and the provision of flexibility to the grid. The PMV methodology focuses on the establishment of a robust and transparent method, based on data streams from local resources and blockchain technology. The new PMV method offers fairness, simplicity, accuracy and replicability in order to foster end users' trust in the remuneration mechanism.

The frESCO PMV approach consists of a continuous and dynamic baselining of generation and consumption of the controllable DERs included in the contracts. Baselines created use real-time data according to modelling forecast algorithms previously trained to adapt to the building thermal performance and users' comfort preferences. This baseline enables a comparison with actual energy metering to assess how much energy has been saved, how much has been self-consumed and how much has been shifted, upward or downward, after a flexibility event. With this available data, service charge and remuneration settlements can be made depending on the service outcome in accordance with the Pay for Performance approach.

The frESCO new services and business models bring under common Pay for Performance Contracts (extended form of current EPCs) two currently differentiated service offerings to enable the realization of next-generation smart energy service packages. The new services combine remuneration from energy efficiency management (savings) and demand response trading in markets (revenues). Both type of services, plus additional non-energy services related to comfort, air quality and security, can be delivered with the same big-data platform, thus increasing the total revenue stream stemming from different services and reducing the payback time.

2.2 frESCO Energy services and revenue streams

The frESCO business models for ESCOs and Aggregators offer advanced data-driven energy services to residential consumers and prosumers under the Pay-for-Performance concept. There are three different energy service business models considered. Firstly, the ESCO business model is the combination of implicit and explicit energy efficiency services for consumers as well as PV optimisation strategies for prosumers, all leading to recurrent energy savings in the dwellings. Energy performance in this case is measured in terms of energy savings obtained in the period of analysis. Secondly, the Aggregator business model deals with delivering demand flexibility for grid management, balancing and congestion issues. The service provider may receive market triggering signals from a grid or market operator and responds by configuring enough DERs operated automatically for a scheduled event to dispatch the requested demand flexibility. In this case, energy performance is measured in terms of the amount of energy that is shifted and dispatched as upwards or downwards demand response flexibility. Finally, a third business model considers the delivery of both type of services simultaneously with the use and exploitation of the same data platform, thus benefiting from both revenue streams.

The Pay-for-Performance concept relies on precise dynamic baselining that facilitates the calculation of the value added as a direct result of energy services through a fair measurement of energy performance. The value added is then distributed between building users and service providers.

The main stakeholders participating in the provision of the energy services are the energy service providing companies (Energy Service Companies ESCOs and Aggregators) and the end users that receive the energy service and provide the demand flexibility. Regarding the energy performance, facility managers and building occupants are the beneficiaries of the efficiency services, whereas grid operators and managers are the beneficiaries of the flexibility services. These benefits are measured, verified, administered, and settled by the energy service provider in their role of ESCO or Aggregator, or both.

The possibility of having near real time data enables ad-hoc dynamic energy profile baselines to measure energy efficiency and to accurately assess the energy performance through which the service provider will obtain the revenues.

The first estimated figures show that frESCO services are economically viable, as long as the dwellings and the service providers meet some minimum prequalification requirements to guarantee payback periods of less than 11 years. Regarding preliminary results obtained from an economic assessment there are several different costs and revenues that depend on numerous factors such as the user profile, the energy demand curves and the scope of services. The following main factors affect the economic viability of the frESCO services:

- Upfront costs, which should be as low as possible, taking advantage of existing infrastructure and limiting the number of onsite visits.

- The maximum number of services that may be contracted simultaneously, thus increasing the sources of income (efficiency, optimization, flexibility and non-energy services).
- Dwellings should have a minimum electricity demand of 500 kWh/year.
- The service provider's customer base must be large enough to share operating costs. A reference number indicates more than 3,500 users.
- The distribution of benefits must not be less than 35% in favour of the financing service provider.

However, to validate or correct the assumptions made in the preliminary calculation, it is necessary to evaluate on a case-by-case basis when actual test data from the testing phase of the frESCO project is received.

3 PMV METHODOLOGY REVIEW

In order to verify and report the energy savings from an Energy Efficiency Measure, Measurement and Verification (PMV) procedures take a significant role. These protocols require planning, collecting and analysing information to reduce the uncertainty and estimate the efficiency savings. Related to the success of PMV protocols, the Energy Services Companies (ESCO) or utility providers experienced an appreciable development in their business models. During the 1990s, different organizations got involved in the elaboration of guidelines and protocols and, consequently, energy savings performance contracts (ESPC) experienced growth.

In 1994, the US Department of Energy (DoE) began working on a strategy to unify the measurement and verification of energy efficiency, and after 2 years of hard work, the North American Energy Measurement and Verification Protocol (NEMVP) was published. This document determined the development of PMV procedures and was consider as the first edition of a PMV protocol. As a result of the extensive international interest, one year later, in December 1997 twenty national organizations from twelve countries worked together and published a new version. The document was renamed the International Performance Measurement and Verification Protocol (IPMVP). This second version, unlike the first, included efficiency opportunities in new construction projects and in the use of water. The IPMVP has been broadly accepted internationally and has become the standard PMV document in several countries.

In 2001, the Protocol Committee of the IPMVP formed a non-profit organization IPMVP Inc. to support and update the existing content including the development of new content [1] . Afterwards, in 2004, this organization was renamed with its current name, the Efficiency Valuation Organization (EVO) [1]. Due to the high levels of adoption of the protocol, during the following years, the IPMVP was updated several times and EVO made translations into different languages.

Considering the collaborative effort between industry, government, financial, and other associations, this protocol acts as the framework for many PMV procedures, being the most used method. As well as the IPMVP, there are other protocols with the same purpose. The U.S. Department of Energy’s Federal Energy Management Program (FEMP) arose in 1996 to reduce energy costs by identifying and procuring energy-saving projects. Hence, in 1996, the FEMP PMV Guidelines were published based on the recent North American Energy Measurement and Verification Protocol (NEMVP), that later became the FEMP. This methodology was specialized towards federal facilities.

At the same time there was another organization that was working on the development ofPMV methodology, namely the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). In 2002, a more technical document called the ASHRAE Guidelines 14-2002 was approved.

Even though the IPMVP protocol was recognised worldwide, the European Committee for Standardization (ECS) published the standard EN 16212:2012: “Energy Efficiency and Savings Calculation, Top-down and Bottom-up Methods”. Taking into consideration the uncertainties of energy supply and the need to limit the greenhouse effect, the aim of this regulation was to harmonize the methodology for monitoring and evaluating energy saving at the European level. The document presents a general approach for energy efficiency and energy savings calculations 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.

Within the framework of the European Directive 2006/32/EC on energy end-use efficiency and energy services, currently replaced by the European Directive 2012/27/EU on energy efficiency, the standard proposes two separated methods, top-down and bottom-up (The European Parliament and the Council of the European Union, 2006).

The top-down method proposes an estimation of savings from indicators established with statistical data, while the bottom-up method is based on actions of end users to improve energy efficiency. Subsequent EED directives such as EU 2018/2002 emphasises the methods to calculate energy savings from individual actions and the reporting requirements.

In 2014, 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" [2], which complements the previous ISO 50001:2011 "Energy Management System" [3], in the context of PMV, a key point for energy management systems based on this standard.

Recently, they also published the ISO 17741:2016 "General technical rules for measurement, calculation and verification of energy savings of projects". The goal of this Standard is to quantify the energy savings during a specific period in new projects or retrofit projects. The energy savings are determined by comparing measured, calculated, or simulated energy consumption before and after and/or without implementation of a project and by adjusting parameters in case of changes in relevant variables (routine adjustment) or in static factors (non-routine adjustment) [4]. This demonstrates the IMPVP's influence in the realization of international regulation.

For this reason, the European Commission DG JRC recommends that performance-based projects base the energy efficiency assessment on PMV protocols.

Aiming at the worldwide scope and recognition of the IPMVP protocols, and the consequent applicability in the frESCO project, a detailed definition of a PMV methodology is required to verify the impact of the project's services. Earlier, other European Commission co-funded projects, such as HOLISDER, FLEXCoop, PARITY or BEYOND have developed PMV methodologies for the verification and assessment of buildings energy performances focused on IPMVP [5] and FEMP [6].

The most recognised protocols and their basis are briefly described below with a summary of their methodologies and important aspects. This includes guidelines such as the American Society of Heating Refrigeration and Air Conditioning Engineers (ASHRAE) Guideline 14 and the US DOE Uniform Methods Project.

3.1 Energy efficiency measurement and verification

PMV methodologies for energy efficiency in buildings are broadly used in the building sector to establish an accurate and representative reference of the historic building energy performance to compare the actual energy measurement against. frESCO PMV for energy savings is related to them, especially to the IPMVP that is described in detail hereafter.

3.1.1 International Performance Measurement and Verification Protocol

The International Performance Measurement and Verification Protocol (IPMVP) offers a framework for implementation of PMV procedures and defines the concepts needed to determine savings. In addition, this protocol functions as a guarantee for the client and the supplier of the Energy Efficiency Measures (EEMs), because it is essential to provide a PMV method accepted by all parties.

Until 2012, the IPMVP was distributed in three volumes:

- IMPVP Volume I, called "Concepts and Options for determining Energy and Water Savings", describes the fundamental concepts and the methodology used in the protocol. This volume includes most of the information needed to apply the IPMVP, hence, it is the most significant volume.
- IMPVP Volume II Concepts and practices for improved indoor environmental quality (2002): this document covers the indoor-environmental-quality issues regarding the design, implementation and maintenance of EEMs [7].
- Volume III known as "Concepts and Options for determining Energy Savings in New Construction", addresses a brief of the PMV methods related to new building construction and incorporated renewable energy systems. This volume is divided in the following two parts:

  Part 1 – Concepts and practices for determining savings in new construction (2006) [7].

  Part 2 – Concepts and practices for determining energy savings in renewable energy technologies applications (2003) [7].

EVO decided to reorganize the IMPVP documents in 2014, and then, published the IPMVP Core Concepts. This document specifies the terminology and principles to apply PMV. Also, it explains the project framework, the contents and requirements, and saving reports:

- Principles
- IPMVP Framework
- IPMVP Options
- IPMVP Adherent PMV Plan and Report
- Adherence with IPMVP

The following review only focuses on the most important ideas of the methodology's principles, framework and options. In other words, it refers to what is necessary to apply the protocol. Thus, the principles of PMV need to be defined first, and they must be considered by any PMV plan based on this protocol:

- Accurate: the report should have the highest level of precision.
- Broad: the project must be defined by a demonstration report, including all the characteristics of the project.
- Conservative: in case of estimation, the potential savings should be underestimated.
- Coherent: the reports must be consistent, in accordance with the different energy efficiency projects, the professionals responsible for energy management and the time periods of a project.
- Relevant: the parameters of interest should be measured to determine the savings and estimate what is of lesser importance or not predictable.
- Transparent: all the PMV activities must be documented in detail.

The IPMVP is a guidance document that promotes efficiency investments by measuring, computing and reporting savings achieved by energy or water efficiency projects in commercial and industrial facilities [6]. Savings are calculated by comparing before and after energy consumption and/or demand, also considering the possible adjustments that could be necessary. Hence a saving is defined by an absence of consumption, the energy, water or demand savings cannot be directly measured. The general PMV equation to compare the two periods is shown below:

$$Savings = (Baseline\ Period\ Energy - Reporting\ Period\ Energy) \pm Adjustments$$

Therefore, the application of this protocol should integrate the identification, development, procurement, installation and operation of energy conservation measures. The IPMVP's work with key points in the process that determine the different periods that must be included as part of good PMV practice [1]. In fact, the first period is titled the *reference period* and includes the consumptions before the application of the EEM. In this period the *reference baseline* is defined, which delineates the consumption curve. Here the independent variables have a significant impact, such as the hours of operation, the outside temperature, or the occupancy. Whereas the period after the implementation of the EEM is called the *reporting period*. In this case the energy function is called an adjusted baseline, after the installation of the EEM and the adjustment corresponding to the variation of some independent variables called static factors.

Therefore, the savings achieved are determined by comparing the adjusted baseline energy with the reporting period actual energy, as represented in the following figure by the green area. [8]

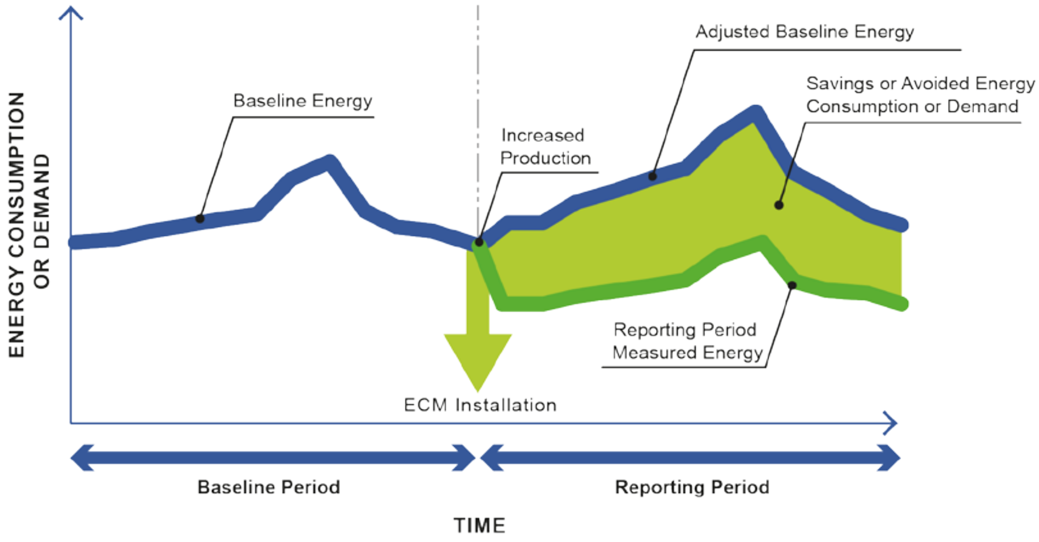

Figure 2 IPMVP framework

Then the application of the Measurement and Verification Plan entails the selection of the most appropriate verification option to define the necessary measurement equipment, establishes the work periods and preparation of demonstrator savings reports. Depending on the scope of the Energy Efficiency Measure (EEM), the precision required and the budget available for measurement and verification, the IPMVP offers four options to calculate the savings:

- Option A – isolated Retrofitting with just one key parameter measurement: this method is based on the measurement of a key parameter that determines the efficiency solution. This option measures only essential parameters and estimates the rest. In this case, the measurement can be continuous or punctual.
- Option B –Isolated retrofitting with all parameters measured: all the variables involved in the EEM are measured. The savings are determined by measuring the energy consumption of the system in which the EEM has been implemented . Also in this case, the measurement can be continuous or punctual, depending on the expected variation in savings and the duration of the savings demonstration period.

- Option C – Whole-facility measurement: this option is associated with the analysis of a factory or building as a whole, considering all the energy demands. So, savings are determined by measuring the energy consumption of the entire installation or part of it. In contrast with the previous optionshere the measurement of the energy consumption of the entire installation is carried out continuously during the savings demonstration period.
- Option D – Calibrated computer simulation: In this case savings are determined by simulating the energy consumption of the entire installation or part of it, after the calibration of a model. The simulation must be able to model the current energy performance of the installation. The implementation of simulation modelling requires expertise in building energy performance simulation.

In conclusion, there is not a strict rule that leads to the most appropriate option, but there are certain recommendations depending on certain characteristics of the project that can help to choose the correct option or exclude others. All in all, the application of the IPMVP enables the systematization of the measurement and verification procedure, adjusting it to recognized and contrasted standards. In case of some disagreement between the parties, the protocol provides a reliable and accurate support for the ESCO client and the ESCO. Finally, after applying the EEM, the savings reports will emerge, describing both the energy and economic savings achieved.

3.1.2 FEMP

The Federal Energy Management Program (FEMP) was developed by the U.S. Department of Energy (DOE) and its aim is to allow federal agencies to achieve objectives related to energy, like identifying affordable solutions, providing energy leadership to the country and facilitating public-private partnerships. Therefore, this program concentrates on key services that help agencies reach energy and water requirement goals [6].

In order to reach their goals, the FEMP seeks to implement replicable and well-designed projects, taking advantage of the resources and assistance offered by the federal agencies. As a result, federal governments reached a 49% reduction in energy intensity since the implementation of the FEMP [6].

The FEMP focuses its effort on the following areas:

- Strategic programming and integration planning: the agencies are provided by FEMP with information and resources to develop strategic programs and plans to successfully reduce federal energy and water use.
- Facility and fleet optimization: FEMP seeks to ensure the optimal development of operations, looking for profitability.
- Energy and water resilience and security: FEMP integrates advanced and secure technologies, in order to plan the portfolio resilience. It also addresses security threats that the energy management system could face.
- Energy and project procurement development services: taking advantage of the private sector financing with no upfront capital costs, FEMP supports federal projects with acquisitions and technical expertise. Energy savings performance contracts, electric power services contracts and renewable energy procurement are some of the areas of expertise that are considered.
- Federal leadership and engagement: FEMP forms working groups between the agencies and engages them in workforce development opportunities, recognising their efforts.

Intending to work on these areas, the FEMP published the Measurement and Verification (PMV) Guidelines for Federal Energy Projects. This document provides PMV options and methods, considering the different levels of accuracy and cost, for verifying the energy and cost savings associated with federal agency performance contracts. This document focuses on developing site-specific PMV plans for federal Energy Savings Performance Contract (ESPC) projects, intending to be fully compatible and coherent with the IMPVP. The FEMP Guidelines assist users in choosing the most appropriate PMV option and method for specific projects.

In conclusion, the IMPVP and FEMP PMV Guidelines are complementary and provide the instructions for quantifying the results from energy savings projects. The FEMP PMV Guideline contains specific procedures for applying concepts originating in the IPMVP. In this sense, the Guideline represents a specific application of the IPMVP. It outlines procedures for determining PMV approaches, evaluating PMV plans and reports, and establishing the basis of payment for energy savings during the contract [6].

The differences between these two documents are identified in the following key areas [1]:

- IPMVP is a framework of definitions and broad approaches, while FEMP PMV Guideline is an application document based on IMPVP, exclusively for the federal sector in the US. In addition, FEMP PMV Guideline offers more details of the application of the different PMV options.
- IPMVP requires metering data whereas FEMP PMV Guide allows a PMV approach that does not require any metering for some EEMs.
- If necessary, the FEMP guideline contains additional recommendations for annual inspection and measurements, in order to comply with the objectives of PMV in performance-based projects, such as ensuring long-term equipment performance [6].

3.1.3 ASHRAE Guideline 14

ASHRAE Guideline 14, Measurement of Energy, Demand and Water Savings, is a document intended for calculating energy and demand savings associated with performance contracts using measurements. It provides instrumentation, data management guidelines and defines uncertainty assessments associated with models and measurements. Guideline 14 does not discuss other issues related to performance contracting. [6]

3.1.4 The DOE Uniform Methods Project

In order to determine the savings from energy efficiency measures and programs, DOE is developing a series of protocols under the Uniform Methods Project (UMP). There are four categories: commercial, residential, combined commercial and residential, and cross-cutting measures.

The protocols provide an elementary method for evaluating gross energy savings for residential, commercial, and industrial measures usually offered in ratepayer-funded programmes in the United States. The International Performance Verification and Measurement Protocol (IPMVP) incorporates additional procedures necessary to aggregate savings from individual EEMs in order to evaluate program-wide impacts. For commercial measures, the FEMP guideline and the UMP are complementary.

3.2 Flexibility measurement and verification methodologies

PMV is the performance measurement process to calculate and validate the provision of a service in harmony with the specifications of a product. The objective of PMV for demand response (DR) is to establish the amount of energy or power that is "delivered" by a DR resource under the conditions imposed by a DR program. The use of this process is the basis for implementing a fair and transparent remuneration structure, in a way that allows end users to gain confidence in the market. It is very important to correctly determine the amount of flexibility offered to provide a payment adjusted to this 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), permits the improvement of operational efficiency and the achievement of a more sustainable and efficient electricity system.

Likewise, positive economic rent from historical DR as a distributed energy resource (DER) will be the main input for preparing and designing a retail programme.

In brief, PMV for DR is used for:

- Establishment of resource eligibility or capacity: For some 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 retribution achieved, and the financial transaction between the programme or product operator and the participant, based on the flexibility. For DR assessment the estimated load consumption without event participation and the actual load consumption determines the calculated flexibility amount that is the basis for the settlement and remuneration to each of the event participants. More usually, different PMV 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 PMV via impact estimation is valuable for calculating program effectiveness and for ongoing planning.

There are many agreements between the customer and the retail operator that can be established; many of these programmes do not require the demand reduction measurement as the basis for settlement with the retail customer or DR aggregator. However, when the reduction at the programme or segment level is offered as a wholesale resource, the amount of demand reduction measured for the program or segment is usually required for the wholesale settlement. Consequently, the application of a performance evaluation methodology to DR events consists of evaluating, against a baseline, the volume of variation in demand that is sold in the market. This flexibility is calculated by the difference between the actual metered consumption during the event and what users normally consume.

It is not possible to measure the baseline directly. This leads us to the fact that it must be estimated and calculated based on the measured data combined with a robust methodology to evaluate the reference energy consumption. Therefore, measuring any DR resource generally requires comparing the observed load during the time of reduction with the estimated load that would otherwise have occurred without the reduction. The difference is the load reduction (Figure above) [9].

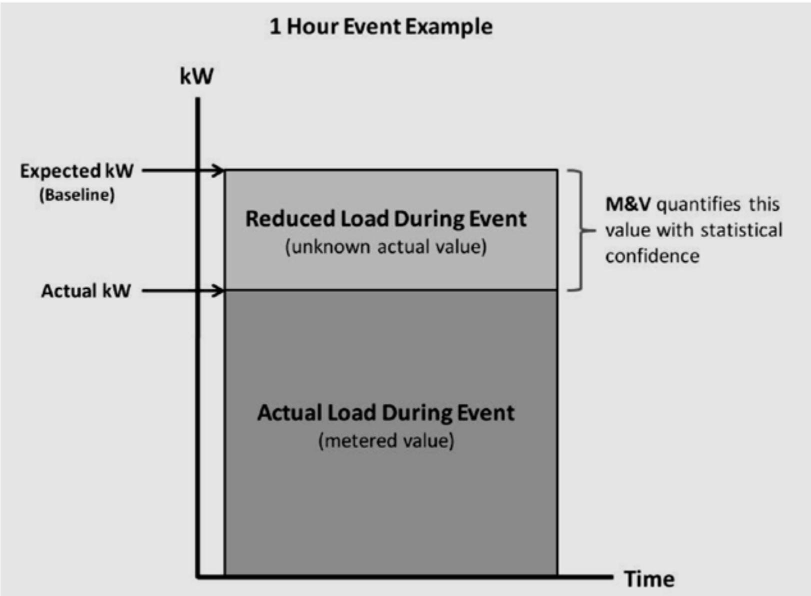

Figure 3 PMV Quantifies Load Reduction Value

The performance evaluation methodology applied for settlement of the DR programme is critical to the success of any DR programme. This could be helpful 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 dissimilar types of baselines. On the other hand, in cases where pay-for-performance is measured against an absolute value, accurate measurement is essential and subsequent verification is simple. However, in some cases where performance is measured relative to a baseline, both defining the baseline and measuring energy are crucial. The challenge is to obtain a simple but accurate estimation 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 for all parties. Being estimations, the baselines are imperfect. Nevertheless, according to NAESB (North American Energy Standards Board) recommendations, good baselines balance four main aspects:

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 intermittent 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 the DR programme when choosing the baseline methodology.

Balancing these aspects is not an easy task. Sometimes, baselines that are resistant to manipulation are complex and difficult to calculate. In simpler cases, they may allow participants to exploit the baseline to their advantage. In addition, it is crucial to consider that baseline estimation should not compensate or penalize the variation in natural load caused by system operations and is generally associated to variance in occupancy or local weather conditions.

Recently, many different PMV methodologies for DR have been implemented in the US and in research projects in EU. The characteristics and specifications of these methods are presented in the following sections.

3.2.1 The eeMeasure methodology

This project studied two different PMV methodologies, based on IPMVP and developed from the experience of current and completed ICT PSP projects which included roughly 10,000 social residences and 30 public buildings (e.g., hospitals, schools). This was the first European project that created a methodology to measure and verify DR in the European Union. The methodologies have been applied in three recognized H2020 projects and one FP7 project, such as NOBEL GRID, MOEEBIUS, ORBEET and Inertia, respectively.

The Residential Methodology [10] applies only to the domestic building sector 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. These methodologies do not focus on changing demand to achieve demand reduction. Only Option C of the IPMVP is applicable in this case. This method verifies energy savings in short or long periods, measuring energy use in the entire installation or sub-installation. This option does not assume constant energy demand or any modelled variation of energy demand but is a before-after evaluation instead.

The Non-Residential Methodology [11] defined in eeMeasure, can characterize any property type (residential, business, etc.) and can utilize data on different temporal resolutions (seconds, minutes, hours, etc.). In this methodology, a process flow is specified, to monitor appropriate variables and to create a precise energy consumption model. A description of the underlying mathematical statistics is also included.

<u>Option C for residential</u>

The discrepancy between the consumption after the Energy Efficiency Measure (EEM) has been applied and the consumption under the equal demand conditions without this intervention, will result in a before-after rough calculation of the energy savings [10]:

Reference data is the estimate of consumption without the EEM. The baseline expansion is the projection of consumption before the intervention into the period after the intervention. The time after the event is called the reporting period. Once the EEM has occurred, the energy consumption decreases.

To calculate the consumption avoided, a model that adjusts depending on the independent variables, such as occupation, size, or outside temperature, is necessary. In case of not being able to measure these variables, a reference period is necessary to have precision. It is advisable to generate regression models that simulate a base energy consumption for these independent variables. Within the residential sector, the climate is the variable that has the highest level of variability. If this variable is not measured, the average temperature or the heating and cooling degrees days are usually used. For regression models, an acceptable accuracy of modelling of the dependent variable is crucial to estimate the adjusted baseline in the reporting period. One metric to calculate this accuracy is the squared multiple correlation coefficient R2, which shows the proportion of variance explained in the model. If R2 is low (less than 0.7), the independent variables chosen must be revised to include or reject some and improve predictions. In addition to R2 there are other statistic to be checked to validate a regression model, such as t-statistic, CV, etc.

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

1. Choose a period for the design of the baseline which captures all variations of the non-measurable independent variables and calculate an average which can fairly 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. Check with the T-student statistics that all chosen independent variables are meaningful in the model and do report signals or variations in the dependent variable rather than just noise. Check the R2 statistics to validate the model fitting.
4. Choose a period for the reporting period which is again long enough to capture all variations of the non-measurable 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.

Steps 1, 2 and 3. Baseline period estimation

Energy consumption must be accurately and constantly measured. This is measured by smart meters. Other variables such as room temperature, occupation, and statistic factors such as surface area or energy source, directly influence energy savings. They are independent variables and should be measured before the intervention.

The variables that affect consumption, such as hot or cold periods or vacation periods, are those that determine the minimum duration of the baseline, but the reference or baseline period should cover a complete operation cycle to comprise all possible variations. Since it is not possible to directly calculate 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.

Other variables that may be of interest can be collected through surveys and are subject to the data protection legislation.

Step 4 and 5. Reporting period estimation

After the EEM and an adjustment time, the energy savings should stay stable for a certain period in the case where the 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 assessed weekly to verify their continuity over time after the ESI, mainly if the savings depend on social behaviour.
- In the long-term, it is very significant to confirm equipment renovations as the baseline estimations may vary.

3.2.1.1 eeMeasure baseline methodology

This methodology considers four specific baseline methodologies to estimate the degree of peak shaving achieved in a DR scenario [10].

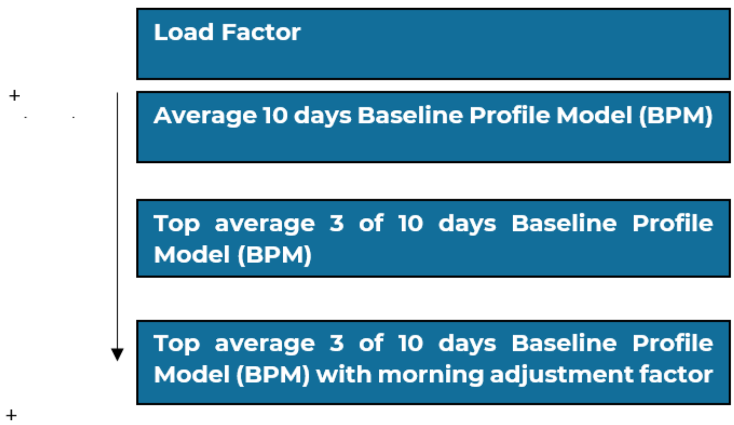

Figure 4. eeMeasure baseline methodology flowchart

Load factor

The value obtained by dividing the minimum power demand by the maximum power demand of a building:

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

As the difference between minimum and maximum power demand decreases (redistribution of peak load hours), the value of the load factor approaches one. In case the peaks correspond to the electrical network, changing this factor to one can represent useful peak shaving for the utility.

10 days Baseline Profile Model

The models are used to estimate peak cut-offs, or peak "events", which occur on random days. Generally, a period of ten business days is taken before the event, since it is taken as a reasonable and representative period, is used to establish the baseline. The reporting period is typically the 24 hours of the event day.

The non-intervention reporting period (event day) estimate is compared with the actual consumption on the event day to quantify the peak shaving. The calculation of average consumption follows the following formula:

$$baseline\ (10\ days) = \frac{d1+d2+d3+d4+d5+d6+d7+d8+d9+d10}{10} for\ the\ event\ number\ of\ hours$$

$$Or$$

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

Top 3 of 10 days Baseline Profile Model

The model averages the 3 highest consumption figures from the previous ten days, excluding other event days, holidays etc. The estimator for the non-intervention event day consumption is:

$$baseline\colon \max(1,10)(\ \Sigma dn(t,h))/3$$

Or

$$DR\ consumption = Baseline\ (average\ high\ 3\ of\ 10\ days) - \ Demand\ event\ day\ (day11)$$

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

This 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:

$$Baseline' = \max\ (1,3)(\Sigma dn(t,h))/3$$

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

$$DR\ consumption = Baseline(average\ high\ 3\ of\ 10\ days) Demand\ event\ day(day11) + morning\ adjustment\ factor$$ [10]

3.2.2 Moeebius project - Modelling Optimization of Energy Efficiency in Buildings for Urban Sustainability [12]

The Moeebius project presents a holistic energy efficiency optimization framework, enhancing the modelling approaches that have been used to date and offers innovative tools.

The project offers simulation predictions according to the complexities of building operations, thus significantly reducing the "performance gap" and improving performance optimization.

This methodology can be found published on the project website [11] and is based on the IPMVP and FEMP methods [6]. This PMV consists of three phases: ex-ante analysis, implementation and measurement and verification. The simulation model and the baseline are compared in the ex-ante analysis. The baseline is characterized by:

- The evaluation of the energy consumption over a sufficient period (about one year) and with sufficient resolution (hourly if possible) to identify variations in consumption.
- Expected breakdown in energy consumption according to use (e.g., lighting, heating office equipment, servers, etc.).
- Independent variables and static factors that affect the energy consumption and the important values (i.e., degree days for heating or cooling, floor area for lighting, building opening hours, metering period length, etc.).

Independent variables must be evaluated at the same time as the energy consumption data. In addition, the simulation model must be defined. This model will be used to evaluate the difference between expected and actual consumption.

This process consists of first identifying the energy sources, marking the measurement points, and then tracking the energy consumption.

Finally, the last phase of PMV calculates, analyses, and evaluates the final performance of the system.

3.2.3 OrbEEt project - Organizational Behaviour improvement for Energy Efficient administrative public offices [13]

The OrbEEt project presents an innovative solution, which seeks to facilitate citizen participation in energy efficiency actions, thus providing real-time evaluations of the energy impact. Using options C and D of the IPMVP, it creates a methodology that manages to combine annual bills and building sub-metering data [13]. This PMV establishes a continuous validation approach (different measurement periods) but in parallel for different loads (different load types). Certain periodic savings adjustments based on independent variables are needed to re-establish the baseline demand for reporting periods under a common set of conditions. If no historical data energy uses are simulated based on utility hourly or monthly billing data.

Depending on the EEM, different time horizons for reference periods may be proposed. This definition of the reference period is shown below, depending on the type of EEM:

Fuel/Gas: HVAC systems

- Baseline period: a period of a year is required for baseline definition.
- Information to register: monthly consumption.
- Independent variables (for routine adjustments): Heating degree day(HDD) or Cooling degree day(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 long period is required for baseline definition.
- Information to register: weekly 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.

In the first tests, information about occupancy levels and environmental conditions were available, and routine adjustments applied. Non-routine adjustments are fixes for changes that are not predictable. Such as:

- 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.

3.2.4 HOLISDER project - Integrating Real-Intelligence in Energy Management Systems enabling Holistic Demand Response Optimization in Buildings and Districts [14]

The HOLISDER project integrates different mature technologies in an open and interoperable framework, comprising the full range of tools necessary to develop the entire DR value chain. This places great importance on the consumer and their conversion into active market players, through the deployment of a variety of implicit and hybrid DR schemes, backed by a variety of end-user applications.

This methodology combines IPMVP option B and C, takes the methodological steps of Option B and uses option C to protect against unexpected events, such as the loss of sub-metering information, etc. Sub-measurement is applied at the beginning of the project reference period for the entire duration of the project.

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

The FLEXCoop project intends to generate an end-to-end Automated Demand Response Optimization Framework, that facilitates the realization of new business models, allowing energy cooperatives to enter energy markets as aggregators.

It equips cooperatives with effective and innovative tools to achieve solid business practices to exploit their microgrids and dynamic VPPs as balancing and ancillary assets toward grid stability and alleviation of network constraints.

FLEXCoop has a great diversity of base technologies to build an open and interoperable DR optimization framework, incorporating a fully-fledged tool suite for energy cooperatives (aggregators) and prosumers implicated in the DR value chain.

This European project takes relevant aspects of options B and D of the IPMVP, since it continuously measures loads and individual parameters to establish the baseline, as occurs in option B. However, the information is used to generate forecast models and calibrate them thereafter, which is similar to Option D. In this situation the difference is that the models are not created at building level, but for every electrical use participating in DR events.

The obstacles in the selection of the reference and reporting interval, in the case of FLEXCoop PMV structure can be overtaken thanks to the methodology itself and to the different duration of EEM implementation that, in the case of DR events, is restricted to a short period.

With the minimum required data possible, the FLEXCoop models are created and calibrated, and this period is defined as the reference period. This minimal amount of data required gives this method an important advantage.

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

The Parity Project works to achieve the integration of blockchain and IoT technologies in a local flexibility platform. This also includes network management tools to address the structural inertia of the distribution network. The main objective of the project is to increase the efficiency and durability of the network by favouring the insertion of energy from green sources in the electricity mix above 50%.

Like other projects, the PARITY PMV is based on a continuous measurement of individual loads and parameters responsible for preparing the baseline, considering a minimum reference period. This baseline is defined by specific algorithms for each DR System.

Depending on the characteristics of each building integrated in the Project, the minimum comfort conditions for each event vary. To develop the PMV, various user-acceptance scenarios were deemed in each of its steps.

3.2.7 BEYOND project - A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond

BEYOND services evaluate the impact on both energy savings and demand flexibility. The methodology of this project is based on precise short-term algorithms, which allow a dynamic and continuous baseline in real time. In addition, different changes that may occur are considered, such as changes in the building, climate, seasonality, etc. In this method, a minimum user comfort level is set, which is always respected. The comparison of demand vs calibrated model, either short-term or seasonal depending on the nature of the event, it permits to assess the impact of the BEYOND services for the customers.

This model is designed to be implemented in buildings, where there is an important data flow, which is integrated into a service contracting platform at the service contracting. Due to this, this methodology is more data-based and therefore simpler than others. However, there must be versatility and adaptability to different services and scenarios.

****

The frESCO PMV methodology for Demand Response measurement and verification is based on the eeMeasure methodology that has been enriched and improved with learnings from the EU-funded Projects FLEXCoop, PARITY and BEYOND, as described in section 5.2

4 PRE-ANALYSIS OF FRESCO PLATFORM DATA

In the case of the frESCO project, the information required for the PMV implementation, that is historical consumption data, system information, comfort, and other independent variable data, will be gathered by the frESCO data collection system, transferred by a communication gateway, and stored in the frESCO big data platform. The platform collects the following three data sets:

Historical energy consumption data

The platform records both energy consumption and the necessary independent variables that may affect the energy demand that is subject to reduction or variation in the future. Energy consumption is measured by smart meters connected to the load supply circuit. Dwelling level readings can be obtained by means of the utility meters if accessible or from newly installed meters or clamps. These variables are used for the creation and self-calibration of the frESCO PMV models and focus on HVAC, hot water, lighting, and other controllable demands as well as on user comfort conditions.

System information

The data platform shall contain information about a list of the energy generation and consumption systems included in the service contract, such as domestic hot water, photovoltaic generation, HVAC, energy storage, etc. In addition, useful information about these systems should be collected to assess flexibility. This information includes power rating, efficiency, type of technology, etc.

Major refurbishments, technology and equipment changes applied in the building as a result of the energy service package must be registered as well. These changes are important to establish a stable reference period and clearly separate the reference period and the evaluation period in the PMV.

Comfort and sensoring data

These variables include those parameters that significantly drive energy consumption per load or at dwelling level and are used for the creation and self-calibration of the frESCO PMV forecast models. The loads involved may be HVAC, hot water, lighting, and other controllable demands aand importantly the user comfort conditions related to these loads.

The minimum comfort conditions are important parameters within the development of the new energy services. Some minimum conditions of comfort between the contractor and the client must always be maintained to avoid comfort dissatisfaction. When customers cannot specify accurately comfort limits, the limits are set by service level agreements.

These end user comfort-related data will feed the platform and the modules to achieve the optimization of consumption and the demand flexibility targets while respecting the pre-defined comfort requirements. In addition, comfort boundaries may be variable throughout the year, so they must be continuously adapted to the seasonal climate variations and customer's demand profile. The comfort boundaries may be defined manually by the users or, alternatively, they can be inferred automatically by checking the users' interaction with the HVAC system. For non-energy services the target parameter values of the contracted services should also be stated, like air quality, noise level, temperature, and humidity, etc.

Other service data

Data on the services provided and events in which the user participates in both energy efficiency and flexibility should be communicated to users for them to be aware of the actions executed and the results obtained in both domains:

- Energy efficiency: information about energy efficiency recommendations that can be applied by the end users and the savings expected. In addition, the smart automation interventions should be notified.

- Flexibility: Information about the types of events in which the user participates, including the duration, the event schedule, the DERs involved in the event and the flexibility target and direction (upwards / downwards) and the expected remuneration.

5 DESIGN OF THE FRESCO PMV

The frESCO PMV methodology is the base of a fair P4P contract type for the new energy services and sets the basis of the service settlement and remuneration between service providers and users. Two clear scenarios should be depicted where different methodological approaches are followed: energy efficiency measurement and verification and demand flexibility measurement and verification.

5.1 Energy efficiency PMV

Energy efficiency measurement and verification include the measurement and verification of any refurbishment, improvement, or energy efficiency action (whether implicit or explicit) whose aim is to provide energy and/or economic savings. These savings benefit the building user and are monetized in the form of a lower energy bill. The aim of the PMV is to separate those savings obtained by the service provider intervention from other seasonal and occasional energy profile variations.

The end user pays an amount of their savings that is proportional to the savings obtained by the ESCO saving strategies implemented in the verification period, retaining the full energy savings of the period. In this case, the frESCO energy efficiency PMV should be able to include all type of energy efficiency measures, ranging from informative billing, efficiency awareness and metric calculation for residents' behavioural change, and human-of users' preferences.

Figure 5 Energy efficiency PMVrepresents the baseline period and the validation period with periodic reporting of savings for billing. The dotted red line represents the period in which the Energy Efficiency services are operational in the building. The blue line in the reporting period is the result of applying the seasonal forecast model using the seasonal values of the independent variables used to characterize the building energy performance.

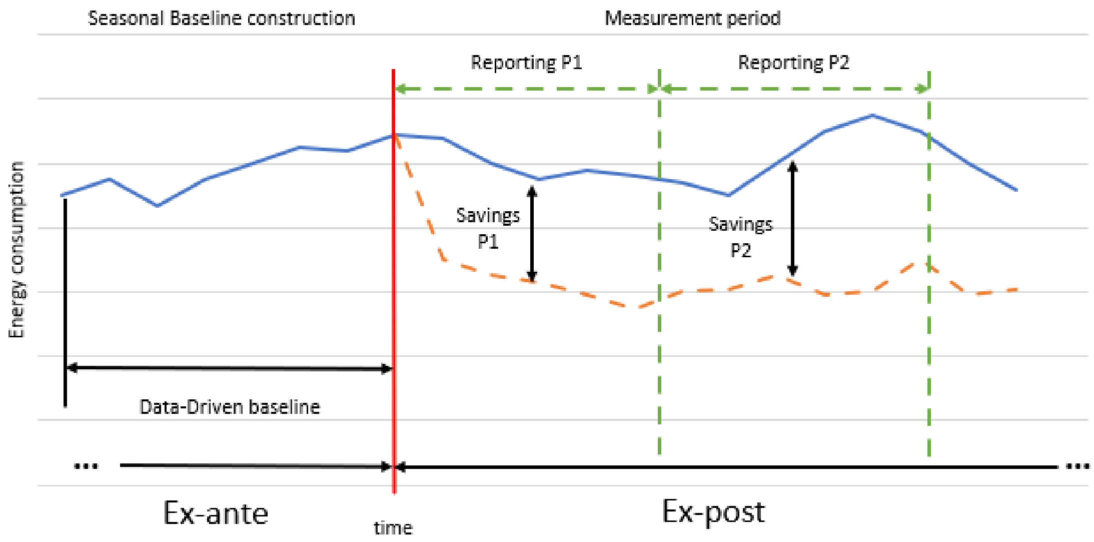

Figure 5 Energy efficiency PMV

5.1.1 Characterisation of the verification scenarios

The reference period is the time horizon in which data is captured to build a solid and reproduceable baseline to calculate ex-post energy savings. The chosen reference time-period establishes a series of important points in the frESCO PMV methodology:

- The reference period should cover all the possibilities of operation of an installation or building. In this choice there are two options, either a seasonal or an annual horizon. However, the seasonal forecast is more adjusted precisely because it delimits the modes of use of the systems and the type of seasonal climate. In other words, it seems easier to specify the consumption of a season like summer than that of the whole year since the demands and user profiles are seasonal. It could be monthly as well.
- Gather data for the entire time cycle. If data is missing from any of the months of the chosen period, it must be taken from averages of similar sampled seasons if available. All the parameters that significantly affect the energy consumption within the building must be available for the chosen period.
- The efficiency baseline is set up by means of pre-trained models that are selected as per the best fit to the building and user profile types. These models are data-driven and should be calibrated for a minimum time horizon with newly collected data. Baseline models should be validated based on statistical values and should cover a complete operational cycle (one full year weather it is split into monthly or seasonal

baselining periods). R2 and T-Student statistics will be used for the validation of the energy efficiency baseline

- The calculation of the baseline is done in homogenous periods (usually seasonally), creating as many baseline functions as needed to cover a full operational cycle (1 year).

The PMV baseline takes overall dwelling electric consumption over the period as the dependent variable and the parameters affecting consumption as independent variables, such as indoor temperature and humidity, occupancy level, setpoint temperature, domestic hot water temperature and consumption, etc. Overall dwelling consumption may be read from real-time digital smart meters whereas independent variables are obtained from the sensors installed in the premises. Data frequency between consumption metering and sensor signals should be balanced so as to obtain synchronous readings.

The minimum comfort conditions defined will feed the model to optimize consumption and demand reductions, but always without exceeding these comfort boundaries. In addition, these values may change throughout the year and should be adjusted, by monitoring the comfort parameters selected by the user.

It is very important to identify and define the main variables that govern the system consumption. These variables must be monitored and used in the consumption forecast models. The creation and self-calibration of the model depends on the availability of continuous data for these variables. For HVAC loads, indoor and outdoor climatic conditions must be taken into account:

- Indoor, outdoor and setpoint temperature.
- Indoor, outdoor and setpoint humidity.
- Domestic hot water consumption and temperature.
- Occupancy.
- Schedule.

The baseline adjustments to be made on this data are based on:

- Non-Routine adjustments: these are parameters that influence energy consumption, but they tend to keep stagnant over time. In this case we find parameters such as the design, the type and number of existing equipment. The changes that these variables may experience have to be monitored on a non-routine basis.

Other additional aspects when applying the protocol for energy efficiency would be the following:

- Price of energy. Adjustments are necessary for the price of energy, as it is variable. These adjustments must be made to all price components for them to be correct. Several adjustment possibilities may be taken for energy prices (annual, seasonal, ….) but using the most updated energy prices for the reporting period is recommended.
- Changes in the boundary system conditions with respect to the reference period where historical data was captured: Once the protocol is applied, there may be changes that may affect measurements directly and indirectly. These changes are usually considered non-routine adjustments, implemented by baseline permanent or temporal adjustments and are monitored on a continuous basis.
- Accuracy: An accuracy check calculating the least square error of the baseline prediction to the actual measurement must be continuously performed to check the suitability of the baseline model to the reality and force necessary adjustments otherwise.
- Cost: Obviously, the cost of PMV implementation must be fully consistent with the savings obtained by the PMV. In the case of frESCO, the usage of the data collection system and data platform entails no or little additional costs to run the PMV assessments and it reports in an automatic way.

5.1.2 Definition of the efficiency assessment baseline

The frESCO energy efficiency baseline is holistic. In other words, it simulates the total electricity consumption of the dwelling.

In the case of PMV applied to energy efficiency, the reference period must comprise a whole year, but it can be made up of seasonal or monthly modelling periods. This split enables the creation of best fit functions adapted to the demand profiles and equipment in use at every season or month of the reference period. For example, the envelope of a building makes the energy performance different in winter than in summer, so a seasonal approach is highly recommended for a better fit to climatic changes throughout the year.

The modifications and retrofitting of the equipment and facilities (building) as a result of the implementation of the efficiency energy services (PV, HVAC equipment, lighting…). have a direct impact in the household energy consumption. These changes must be classified in the following groups:

- Behavioural changes: changes based on the efficiency recommendations provided by the frESCO analytic tools.
- Automated efficiency interventions performed by the frESCO efficiency modules
- Technical improvements: energy efficiency measures based on equipment, such as the acquisition of smart devices, control systems, retrofitting, etc.

As mentioned, energy efficiency evaluation will follow a more traditional method. Option C of the IPMVP, is closest to our case as it implies the overall energy performance assessment for a long period (reporting period). In case no historical data is available, which is the most likely case, pre-selected models will be made available, to select the one that fits best to the building and user profile. The pre-trained model is calibrated automatically for a given period (minimum one full reporting period) using the data flow of energy consumption and the parameters that can influence it. This is similar to IPMVP's option D. Models are fixed along the reporting period, but regular accuracy checks and dynamic adjustment mechanisms based on the continuous data collection will be applied to ensure a correct model fit.

5.1.3 Energy efficiency assessment

Savings in the period are calculated as the difference of the adjusted seasonal baseline evaluated in the reporting period and the actual metering consumption during the reporting period. Non-routine adjustments are made to dynamically adapt the baseline to the new metering data that is read on a continuous basis.

5.1.4 Definition of the Efficiency PMV report

A PMV report shall be issued individually per user / dwelling and based on the actions carried out in energy efficiency during the reporting period. This information must contain at least:

- The type of pre-trained model in use to build the baseline.
- The training period of data collected with the beginning and end dates of that period,
- The systems involved and the type of energy efficiency actions deployed in the period by the service provider.

- Description of issues and deviations and the adjustments performed to the baseline if any.
- The total energy savings calculated in energy and monetary units.
- The average comfort parameters

This report is issued to the end-user through the frESCO platform visualization toolkit, depending on the user's preferences.

5.2 Demand flexibility PMV

The elicitation and extraction of demand flexibility from building residents is carried out by a business actor (aggregator) to offer a service to a grid owner or an operator with different final objectives:

- Optimal infrastructure dimensioning and planning by peak shaving and demand curve smoothening.
- Congestion management
- Grid balancing and network operator ancillary services

Hence, most of the demand flexibility takes place during events requested by the final beneficiary of the service, i.e., the distribution and grid operators. They can run their own local markets or purchase the flexibility in official open markets at country or European level. The retribution generated by this service is then shared by the aggregation company according to the flexibility delivered by building users under a P4P scheme. The way to determine the amount of flexibility dispatched by every aggregated load implies the use of Demand Flexibility PMV protocols.

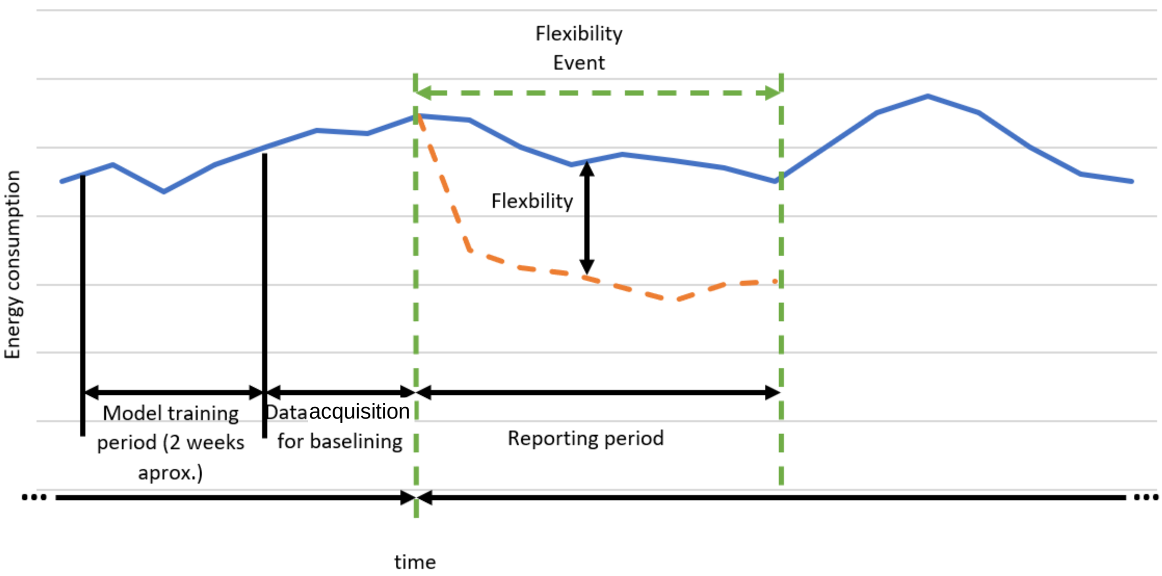

Figure 6 Demand flexibility PMV

The protocol implementation is event-based. Thus, the time horizon is the closest to the flexibility event, and the consumption forecast is calculated for the whole event duration based on the near-real-time data collected before the event and using a trained data-driven consumption baseline for every load involved in the flexibility event. The longer the training horizon the more data that can be gathered to adjust the forecast models into a reliable baseline. However, a higher weight should be given to the data acquired just prior to the event since this data reflects more accurately the system status just before the event takes place. The flexibility baseline is dynamic, renewing itself as new data comes in. This is why the frESCO flexibility baseline does not need additional adjustments and the training periods are not excessively long. From 10 to 15 representative days should be enough for a good algorithm training as proven in past project feedback such as FLEXCoop and Parity projects.

5.2.1 Characterisation of the demand flexibility events

The first step is to know and classify the type of event in advance. In this way, we can evaluate the data needs to be able to define a proper baseline. In the case of demand flexibility, events are of a short period of time, for example, modifying the dwelling setpoint temperature to vary consumption upwards or downwards during the event duration and using the building thermal inertia to keep the comfort conditions. This example needs a short-term approach, modelling consumption with occupancy and indoor temperature just before the event takes place.

Depending on the type of event, the measurement needs vary. The two main flexibility sources in domestic dwellings are defined below:

Table 1 Main flexibility demands for PMV

Event type	Use	Baselining period	Minimal metering scope/ Data source
Flexibility	Electric Hot water	Short-term forecast (15 min to 1 hour ahead)	Load level
	Electric HVAC		

Faced with the lack of historical data or erroneous data, a pool with different pre-defined models will be generated for each demand. With these models, it is possible to create a baseline model with just a few weeks of data for training. The set of pre-trained models will cover different climatic zones, building envelopes and user consumption profiles.

Flexibility-related events need a short-term forecast so they will select the shortest possible reference period for baseline training, while guaranteeing an accurate prediction. The baseline obtained this way is dynamic as it is recalculated on a continuous basis as new data comes into the moving reference or training period. This way the baseline is updated to any change in external weather conditions or internal user behaviour changes, avoiding the need of continuous manual adjustments and adjusting to the latest user energy profile. Finally, the baseline is calculated using the values of the independent variables just prior to the event, thus reflecting the actual conditions closest to the event. This point is important when choosing the granularity of the data, which ranges from 15 minutes to an hour before the forecast.

The event may affect one or more of the users' DERs. A different baseline should be calculated for every involved load in the event. The involved loads are chosen by the aggregator based on previous experiences and the VPP configuration parameters used, such as energy power, availability, reliability, and other contractual boundary conditions enabling the aggregator to stack up enough flexibility resources to meet the committed flexibility delivery.

5.2.2 Definition of the Demand Flexibility baseline

This baseline mathematical model is created by machine learning algorithms based on long term valid historical data if available. Often, such precise data may only be available when the frESCO data platform is installed on premises and hence, there is no valid historical data to create the model. In this sense, the data platform will provide a catalogue of pre-trained building models to choose the one that fits better to the target building. In any case, models must be trained using actual data (consumption and performance data) for every selectable DER in the pool. In the case of flexibility, the baseline consumption model is trained using short time horizons as opposite to the baseline for energy efficiency where training periods may last for the full baseline applicable periods like full reporting months or seasons. In DR assessment, a minimum two-weeks training period is estimated necessary for an accurate model training. The baseline consumption is continuously updated as new data is ingested in the data platform prior to the event using short time calculation periods to ensure a best fit to the actual conditions of the flexibility event.

This baseline estimation method follows the individual regression analysis described in Annex 1. The frESCO data platform collects the data to perform the baseline training and estimation, excluding data that may be erroneous while applying rules to complete data gaps.

5.2.3 Demand flexibility assessment

Demand flexibility is calculated by computing the difference between actual energy consumption and the expected baseline consumption during the event. This difference could be positive, corresponding to an upward flexibility request, or negative, following a downward flexibility request. The actual consumption corresponds to the reporting period and the expected consumption corresponds to that of the actions given by the energy baseline.

It is important to remark that, in order to provide an accurate demand flexibility assessment, no other simultaneous energy efficiency action, whether implicit or explicit, should be programmed at the time of the event, so as to avoid mixing results of strategies with different aims. Downward flexibility events do not correspond to long-term energy efficiency strategies since they often imply a consumption overshooting after the event.

5.2.4 Definition of the Flexibility PMV report

At the end of every successful DR event, a PMV report will be issued individually by the aggregator to the end users based on the automated actions carried out in the flexibility event, with the results of demand response delivered during the event. An estimation of economic settlement should be reported as well. This information must contain:

- Type of flexibility event and market requester (ordinary, extraordinary, upwards, downwards, …).
- DERs involved in the flexibility event
- Schedule and duration of the flexibility event.
- Amount of shifted demand (kWh) per load.
- Unitary price of the energy shifted (€/kWh) and estimation of remuneration.
- Comfort conditions during the event (temperature, humidity, etc.).

This report is issued to the end user through the frESCO platform application, depending on their preferences.

A monthly or regular report at the billing period should be issued showing the period total demand response participation in aggregated terms, the economic settlement, and the distribution of the benefits between aggregator and the users according to the contractual agreements.

5.3 Performance assessment of non-energy services

Non-energy services satisfy additional needs of the residential sector, but they are not directly related to energy flexibility or efficiency. This type of service does not generate savings or revenue streams related to an amount of energy saved or shifted since the purpose is to provide other subjective or well-being value added outputs such as comfort or noise reduction. This group of services are complementary and take advantage of the data available in the frESCO Platform to offer value-added services for which some users may be willing to pay.

This service bundle is also offered under a P4P basis. In this case, performance is the ultimate objective of the service (comfort, noise, air quality and surveillance) measured through the frESCO data platform and compared to the contractual service levels. In this way, we can objectivise the measurement and establish a P4P payment mode for every service.

From frESCO D3.1 we identify four non-energy services:

- NE1: Thermal Comfort services. Comfort preservation and automation at minimum energy costs. Requires comfort parameter sensors, smart controls and switches. Although thermal comfort is a rather subjective parameter, the objective parameters measured are indoor temperature and humidity in representative spots of the user's dwelling. The performance is measured as the time that the system is able to keep the dwelling within comfort limits or the number of deviations from the contractual comfort boundaries.
- NE2: Indoor air quality preservation. Preservation of Indoor air quality by means of air quality sensors. Smart ventilation. The CO2 concentration is the objective parameter measured to assess the indoor air quality, requiring ventilation when the air does not meet the right air quality levels. The performance is measured as the time that the system is able to keep the dwelling within air quality limits or the number of deviations from the contractual air quality boundaries.
- NE3: Noise reduction services. Noise sensors. Scheduling of noise devices and appliances at certain periods of time, smart ventilation, others. The noise level is the objective parameter measured to assess the indoor noise level, requiring shifting noisy loads when the level of noise does not meet the right levels. The performance is measured as the time that the system is able to keep the noise within contractual limits or the number of deviations from the contractual noise level boundaries.
- NE4: Security and surveillance services. Presence sensors, scheduling of lighting at night / absences to create a dissuasive security system. The level of security is the subjective output of this service and the number of false positives or negatives of failures in automated systems could be the objective parameters measured to assess the performance of this service. The performance is measured as the number of incidents from the contractual target, that should be zero incidents.

The P4P type of contract is based on a direct relationship between service payment and service performance, therefore it is very important that the service parameters associated with performance are measured with great accuracy. Service payments are directly related to the performance obtained in the billing period as stated in the service contracts. The P4P contracts are handled in frESCO's D3.6

5.4 FrESCO PMV suitability and compliance

frESCO PMV has been created following approaches from existing methodologies such as IPMVP and eeMeasure described in section 3. The methodology must align with the basic principles of all PMV protocols such as fairness, accuracy, dynamicity, simplicity, replicability and flexibility. frESCO PMV achieves these premises in the following way:

Fairness. During the project, the frESCO PMV model will be defined and tested with and for different types of users, buildings and climatic conditions. This will provide a standard method for a just distribution of savings and demand flexibility between users and service providers according to the energy performance measured in terms of energy savings for EE services, and in terms of energy shifted for DR flexibility. Estimates of demand reduction are used to carry out the economic settlements under the P4P frESCO service contracts for service billing purposes.

Accuracy. The approach of our PMV allows the measurement of the variables that most affect the energy demand of the building. The accuracy is measured as the least square deviation of the demand forecast with respect to the actual energy consumption in a period with no events. This accuracy should stay within controlled error limits both in terms of average error and standard deviation. In addition, the flexibility PMV is applied individually per load, thus increasing the accuracy per load or demand. The baseline is then calculated using actual data just prior to the event for higher accuracy. The accuracy contributes to the fairness of the measurement and verification process.

Dynamicity. The baseline and forecast models are continuously learning from data before the event and adjust the baseline on a continuous basis in a dynamic way. As an advantage, we have greater accuracy than with traditional models, since the loads are measured independently, continuously and in almost-real-time. This avoids errors produced by measurements affecting the whole building. Dynamic baselining also ensures that baseline models adapt to changes in climate and user behaviour.

Simplicity. By measuring independent variables or loads, more information is provided to both clients and service providers. This information provided to the users supports the billing process with transparency and simplicity when it comes to understanding the billing concepts. DR events are reported individually and per participating load to help users understand better the final service settlements.

Replicability/Flexibility. Using different forecasting algorithms for generation and demand, and for holistic dwelling consumption and per load, the methodology is modular and simple to replicate in other buildings and DERs. Each forecast is individually calibrated to adjust the baseline at every event with real-time data. That is why it is possible to obtain great flexibility.

6 CONCLUSIONS

This document describes the frESCO Performance Measurement and Verification (PMV). A study and review of the current PMV and baseline methodologies has been carried out in order to apply them to design a specifically tailored methodology to assess energy savings and demand response flexibility for domestic consumers. The frESCO PMV is supported by the use of large amounts of data arranged and curated in the data platform to create a forecast of energy consumption based on historical data, that will be used as a reference to compare with the actual energy metering in the evaluation period. The collection of dependent variables (energy demand) and independent variables that drive energy consumption in near real time is used to produce reference baselines for long term holistic measurements of energy efficiency savings as well as for short term event-based energy shift measurements associated to Demand Response events.

Since the nature of the measurements is different, the designed methodology is divided into Energy Efficiency and Demand Flexibility measurement and verification. A thorough review of existing and new PMV methodologies and baseline estimation methods has been made. frESCO PMV protocols are partially based on the previous practices and developments made in this matter. For Energy Efficiency assessments the frESCO methodology relies on IPMVP's option C, using pre-calibrated mathematical models if no historical data is available in a similar way as option D. For Energy Flexibility assessments the frESCO methodology follows a similar approach to the eeMeasure methodology with individual regression analysis based on short-term data, expanded with the learnings and improvements coming from previous EU-funded projects such as FLEXCoop, PARITY and BEYOND.

The specific frESCO methodology to measure performance in the three types of energy and non-energy services is summarised hereafter:

- Energy Efficiency PMV. The EE PMV aims at measuring energy savings of the smart retrofitting and energy efficiency frESCO services. Savings are measured holistically at dwelling or building level in specific periods of medium and long-time horizons (e.g., billing periods). Savings are directly enjoyed by the end users and used to pay the ESCO for the service delivery, proportionally to the savings obtained in the period. Savings are achieved from various implicit (behavioural changes) and explicit (automation) strategies, as well as from building and equipment retrofitting. Baselines are built based on historical data comprising a full year. If no historical data is available, baselines are constructed upon pre-selected models and then calibrated for a given period using the data flow of energy consumption and the energy consumption driving parameters in the reference period. There can be more than one baseline to model the energy performance of the building along the year such as seasonal baselines (three or four baselines in a year) or monthly baselines (12 baselines in a year). Baselines are fixed in the reporting period, but they are subject to continuous accuracy checks that may reveal the necessity of non-routine baseline adjustments calculated periodically at every billing or reporting period.

- Demand Flexibility PMV. Demand response services offer aggregated domestic consumer demand flexibility to grid operators for congestion management and grid balancing services. These services are demanded on short event basis and the performance is measured via the energy shift achieved by the automated operation of available Distributed Energy Resources (DER) by the aggregator. The aggregator delivers this demand flexibility to the grid and shares the market remuneration with the flexibility providers or the building users. Flexibility-related events need a short-term forecast so the shortest possible reference period for baseline training is selected, while guaranteeing an accurate prediction. Thus, the baseline is load-based and dynamic, as it is recalculated on a continuous basis as new data come into the moving reference or training period. This way, the baseline is updated to any change in external weather conditions or behavioural changes, avoiding the need for continuous manual adjustments, fitting to the latest user energy profile. Finally, the baseline is calculated using the values of the independent variables just prior to the event, thus reflecting the actual conditions that are the closest to the event.
- Performance assessment of non-energy services. Non-energy services are additional value-added data-driven services that use the available user and building data to deliver optional benefits to the end users such as comfort, noise control, air quality or others, all under P4P contracts. The performance of these optional services is not

based on energy measurements but on compliance with the contractual service levels. Measurements of the involved service parameters are compared to the target values to derive service payments from the degree of compliance with the set targets.

The frESCO PMV meets the necessary requirements for a P4P contract, such as accuracy of the measurements, fairness of the contracts under a P4P approach and the involved economic settlements, simplicity of the methodology with no or few baseline adjustments needed, replicability to any type of building and users and flexibility to adapt to a wide range of user profiles, buildings and equipment. The Efficiency baselines are fixed for monthly or seasonal evaluations along the year with continuous accuracy checks to make the necessary non-routine baseline adjustments. On the other hand, the methodology for flexibility baselining is dynamic and uses the data flows to continuously adapt the baseline function to new boundary conditions, user profiles and climatic changes, and take the parameters prior to a flexibility event to evaluate the energy baseline.

The fresco PMV will be implemented technically in the data platform in WP4 and tested in several buildings and dwellings in the four demo-sites in WP6. Demo sites are selected to include a broad sample of climatic areas, residential typologies, and usage profiles. Baseline models and training procedures will be subsequently refined with the results of these tests.

7 REFERENCES

[1] Efficiency Valuation Organization, «International Performance Measurement & Verification Protocol. Concepts and Options for Determining Energy and Water Savings,» 2021.

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

[3] International Organization for Standardization (ISO), *ISO 50001:2011 Energy management systems – Requirements with guidance for use,* 2011.

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

[5] Efficiency Valuation Organization, «International Performance Measurement and Verification Protocol. Concepts and Options for Determining,» January 2012. [En línea]. Available: http://www.eeperformance.org/uploads/8/6/5/0/8650231/ipmvp_volume_i__2012.pdf.

[6] FEMP of the US Department of Energy, «PMV Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0, Federal Energy Management Program,» November 2015. [En línea]. Available: https://www.energy.gov/sites/prod/files/2016/01/f28/mv_guide_4_0.pdf.

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

[8] [En línea]. Available: https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp.

[9] AEIC Load Research Committee, «Demand Response Measurement & Verification,» 2009. [En línea]. Available: https://www.smartgrid.gov/files/demand_response.pdf.

[10] European Commission, «The ICT PSP methodology for energy saving measurement - a common deliverable from projects of ICT for sustainable growth in the residential sector,» April 2012. [En línea]. Available: http://eemeasure.smartspaces.eu/static/files/eemeasure_residential_methodology.pdf.

[11] European Commission, «Moeebius project,» [En línea]. Available: http://www.moeebius.eu/.

[12] KEMA, *PJM Empirical Analysis of Demand Response Baseline Methods.,* 2011.

[13] Enernoc, *The Demand Response Baseline,* 2009.

[14] Association of Edison Illuminating Companies (AEIC), *Demand Response Measurements & Verification,* 2009.

[15] Efficiency Valuation Organization (EVO), *Uncertainty Assessment for IPMVP - Release of a New IPMVP Application Guide,* 2018.

[16] California Energy Commission , *Protocol Development for Demand Response Calculation. Draft and Findings.,* 2002.

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

[18] Southern California Edison , *Southern California Edison Company Demand Response Appendices to SCE-1, Volumes 1-4,* 2011.

ANNEX 1

1 BASELINE ESTIMATION IN PMV METHODOLOGIES

1.1 Experiences of PMV baselining methodologies for DR in the US.

This chapter is devoted to a literature analysis of existing baseline methodologies in application to the energy measurement and verification domain for flexibility assessment. The most advanced area in the application of PMV methodologies for Demand Response verification is North America, due to the longer tradition of the ESCO business models. In North America, the precision of the different baseline estimation methodologies has been studied by empirical verification of the baseline estimation in DR. The main methodologies are listed hereby.

1.1.1 California Energy Commission

A large number of possible baselines are compared using real data in the report "Protocol Development for Demand Response Calculation – Findings and Recommendations" [12] produced by the California Commission on Engineering (CEC). Interval load data were provided from several parts of the U.S., for both curtailed and uncurtailed accounts. The methods used were classified according to three characteristics:

- 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 a study of more than 100 combinations of data selection criteria, estimation methods and adjustments, and detailing specific findings, the final conclusion is that no approach offered a comprehensive solution. However, some recommendations were made:

- The best solution for a 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.
- Regressions including weather variables 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.

1.1.2 ERCOT Demand Side Working Group

ERCOT [13] performed an analysis of the options for baselines for weather-sensitive loads with short curtailments. This study compared eleven calculation methods at four levels 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 found:

- 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.

1.1.3 Southern California Edison – Methods for Short-duration events

Southern California Edison (SCE) [14] conducted a study between 2007 and 2011, about 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. SCADA data was used to assess the impact of the load and the event analyses. This study demonstrates the value of short-term direct load control programs. In addition, the existing technological barriers are in evidence.

- Short duration events were found to have a minimal impact on customer comfort and a reduced post-event snapback.
- Because there was no pre-event notification of dispatch to participating customers and snapback was minimal, baseline modelling approaches that utilized both pre and post event load information proved to be effective.
- The ex-ante forecast has improved, but the inherent variability in the measurable load impact of the aggregate resources still remains a challenge for market integration. Telemetry of the aggregate resource through technological developments in AMI deployment is the best option to face this barrier

1.1.4 PJM

In 2011, PJM[1] sponsored an analysis of baseline options for PJM DR programmes [15].

In this analysis, their performance was classified by measuring variability and relative error. In addition, the expected administrative costs were taken into account for this analysis.

The cost of administration was an important variable when deciding on a baseline, when these offered similar levels of precision. DR customers represented 39% of the total number of DR customers across PJM territory and 54% of Peak Load Contribution (PLC).

[1] 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

This study consisted of covering all the baselines currently used by ISOs and representing a variety of estimation methods and data selection criteria. Four of them were based on the average load of a subset of a rolling window, two flat baselines, two based on regression and two on split day.

Within optimal conditions, four types of fit were applied to the baselines, including a ratio fit, additive, and a weather-sensitive PJM regression-based additive. The additive and ratio adjustments were the same day load-based adjustments common across the industry. This setting is not based on the load on the day of the event but on the weather on that day. In this way it avoids efforts related to the adjustment based on the load of the day of the event, but uses a regression of the climatic sensitivity, which requires additional data in addition to computational complexity. The following conclusions are drawn from the analysis of these reference methods:

- Methods that use an average load, with the same additive or multiplicative adjustment, perform better than any adjusted baseline with the climatic or unadjusted PJM adjustment.
- Variable loads do not perform as well as expected under all conditions. Clients with this type of load must be segmented to apply a different methodology, suitable for this type of load.
- The best results were provided by the PJM weather-sensitive adjustment method without load adjustment. This entails additional costs and complexity.
- One of the most accurate lines was the PJM's existing high 4 of 5 baseline, and it did not require any additional cost for its implementation.
- Other methods have shown better precision, 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.

1.2 Baseline estimation methods

1.2.1 Maximum Base Load

This method is the least complex and tries to define the performance in DR events. This refers to the ability of a resource to operate both at and below a defined load level. Normally, the data from the previous year are necessary to define the power which the customer cannot exceed. Often, this line is not representative, due to possible demand profile changes that the consumer may make in the facilities. That is why this technique is based on the maximum base load (MBL) from the previous year. According to PJM [12] this method is the most accurate to evaluate the contribution of DR in the capacity market.

1.2.2 Before / after metering

Through the meter readings before implementation and during the response period, a baseline is defined, thus comparing the measured performance with it. This method shows variations in real-time, collecting data before and after to measure the change in demand, which is why it is used in fast-response programs. According to PJM and NAESB it is the most appropriate method to evaluate load reduction in ancillary services such as frequency regulation and reserve events when individual interval meters are available. Nevertheless, it requires demand resources with flatter load profiles. The actual level of load reduction may be affected if the resource has high variability, that is, it may overestimate or underestimate the actual level of load reduction.

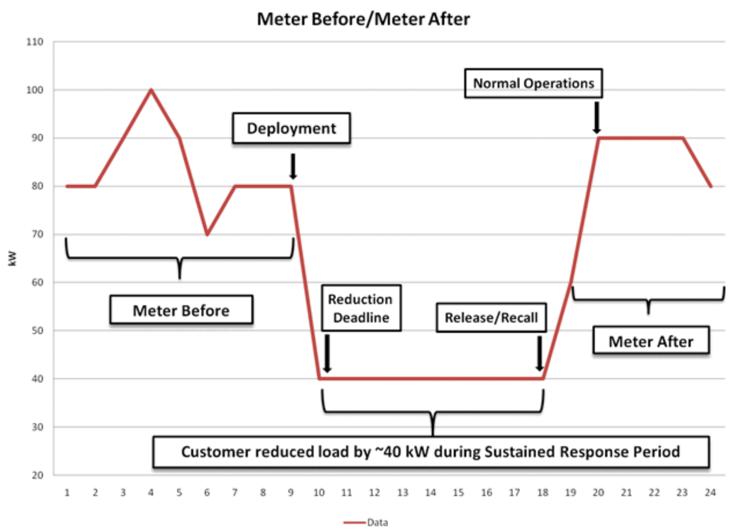

Figure 7 Before / After Metering

1.2.3 Baseline Type I

Using historical interval metering, meteorological and / or calendar data, the baseline is generated. The techniques used for data analysis are typically moving averages, coincident day values, and period averages.

- Moving averages: use historical meter data weighted towards more recent data to reflect representative conditions based on actual data.
- Matching days: use a representative day from the past.
- Average of periods: By means of the average of the historical energy data, they create the baseline. They are also known as X high / medium baselines, where Y marks the most recent number of days with X high load days. As an example, the High 4 of 5 baselines indicate that the four highest values from the last five days are taken.

This method is used to measure and verify the contribution of DR in the daily markets or in real-time, according to PJM. For a DR program that facilitates the aggregation of individually metered end users, the calculation of the aggregate baseline is viable from the aggregation of the individual end users' interval load data and the comparison with the aggregate observed load to determine the demand reduction. Otherwise, the aggregated demand reduction may be calculated as the sum of individual end user reductions, each calculated from its own baseline and own actual load.

1.2.4 Baseline Type II

This method uses statistical sampling, the data are usually from aggregated meters, which is the most common scenario where this method is used. Using these meters, the baseline is created and it is appropriately distributed to individual sites or loads that could not be measured. It is usually more appropriate or used for the residential sector, as commercial and industrial facilities can meter energy usage in a cost-effective way. This method, in addition to not guaranteeing good results, is complicated. Only the North American Energy Standards Board (NAESB) recommends this method as an alternative to Baseline Type I when all individual intervals metered are not available or in case of aggregated loads

1.2.5 Experimental design

This method has been used in recent years as an impact evaluation method, through the random assignment of eligible participants to treatment and control groups. This method could be introduced as an application of the 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 can be useful for cases with a large number of homogeneous clients, such as residential and small businesses, which have a relatively similar consumption pattern. In addition, it is advantageous to use it in groups since individually it can be costly and time-consuming. The estimation of the impact is achieved by comparing the types of load resulting from the set of clients with other similar ones. A well-defined target market is required to generate these forms of charge. Target markets are segments of larger customer classes defined by specific characteristics. There are two groups of clients: those who accept the offer and are considered as participants and those who reject it who are non-participants (control group). Hence, another variable which is the random assignment of clients in the two groups; is the “participants” group and the "control" group. The difference in averages between the two groups is what marks the average reduction in demand per participant.

Another type of calculation consists on a difference-difference method. This method is a baseline calculation or an individual load model, both in both the participant and the control groups. The impact is then calculated as the difference between the moderate and observed load of the participant group minus the corresponding difference for the control group. With this approach, we obtain an estimate of how the real burden of the participant group is compared to its model.

It is not always possible to randomly assign customers to different groups. In these cases, comparison groups of customers identified as similar to the participants after the event are sometimes used for impact estimation. Without randomization, the differences between participants and non-participants are unknown, leading to a possible error in the estimation based on the comparison of groups. The randomized control experimental design is conceptually the best form of evaluation but has been limited until recently. This is because large-scale programs and regulatory contexts do not allow for the random assignment of clients to participate.

A recent exception in the context of energy efficiency are behaviour-based programmes that provide information to a large number of randomly selected residential customers.

In some feasible cases, the experimental design produces very accurate results for estimating load reduction. This method virtually eliminates any difference between participation and control groups, providing an unbiased estimate and a large sample to provide the best precision. On the other hand, it would not be effective with fewer clients or large industrial clients due to the short sample size for statistical estimations in random groups.

When most participants have interval metering data available, this method offers many advantages:

1. Since PMV is done separately for each day of the event, participants do not have to be assigned to either experimental or control groups permanently. Instead, the control group is a different set of participants randomly selected for each event. This ensures that the experimental group and the control group are homogeneous, including the fact that they have equivalent programme experience.
2. Large control samples, that is, a large-scale program achieves very accurate results without drastically reducing the total resource.
3. For ex post estimation a randomized treatment control difference has been shown to provide a very accurate estimate of savings. Using meteorological modelling, 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 participating group from the difference between the modelled and actual load of the control group.
4. For ex ante estimation, observing a large number of participating and non-participating customers during each event provides a much more precise basis for modelling the effects of events as functions of weather or other variables. This type of modelling can be inaccurate if the ex-ante period is very unusual and not representative.
5. Taking a random control group as a basis for evaluation, convening events on unusual days do not create any problem for PMV.

6. The experimental design approach can allow robust load reduction estimates to be developed for a wide range of conditions, while exposing any individual customer to a limited number of control events.

1.2.6 Metering Generator Output

The output data is generated to measure the demand shifting, taking into account that all loads taken by the generator have been in the system in another way. 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 the behind-the-meter generation.

1.3 Exploratory data analysis

Within all the aspects that make up the PMV protocols, the estimation of the baseline is critical, especially for residential users, who have highly variable and climate-sensitive demands. These PMV protocols have been around for a long time, but they have mainly been used for the estimation of savings derived from the implementation of energy efficiency measures (EEM) and not for Demand Response assessment. The implementation of an EEM is permanent while a DR event is temporary, so the main difference is the time of impact. The baseline for DR assessment is estimated based on historical data, but any energy shift in an interval could be due to special and unusual conditions beyond the DR intervention. Hence, the reference techniques usually take recent data, around ten days before the event, to ensure a minimum accuracy level. However, having a long period of measurements is advantageous, because it can hide the issue of temporary lack of data due to error or bad readings

The most common techniques for handling data are the use of this Day-matching technique and regression analysis. Both techniques are shown below.

1.3.1 Day matching

This method involves choosing a subset of days from the historical period that match the event day conditions, and averaging them, often with an adjustment for the current day's conditions. Common bases for identifying match days for a given event day include:

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

All clients have their reference load that corresponds to the load of each user on the matching day. In the case that there are several matching days, the average of those days is taken. The reductions are estimated by the hourly difference between the matching day and the load of the event day. This is the High X of Y method, which has been examined and recommended by the EnerNOC "Demand Response Baseline" White Paper [13] and the KEMA "PJM Empirical Analysis of Demand Response Baseline Methods" [12] as the best for baseline construction. However, defining the number of days to take depends on many factors and there are concepts that have to be clear:

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, and scheduled shutdowns (as these are not representative of "normal" operation).
3. X to Y Ratio: 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 adjustment reflects percentage demand comparisons and additive adjustment 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 estimation 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.

Simplicity and transparency are considered a major advantage in day matching. Furthermore, if variable loads are unknown, matching techniques can achieve more precision than other techniques such as the regression, as long as the matching criteria include characteristics of the individual client's load. In contrast, regression methods are more accurate when loads are well-known in terms of hourly consumption and independent variables such as climatic data. A disadvantage of this method is that it is highly dependent on the availability of historical data for accuracy, if historical data set is not available, regression models are then recommended.

The problem and complexity for assessing the accuracy of a match-day estimate is greater than assessing the accuracy of a regression model. Testing for lack of fit or systematic bias is not as straightforward with a matching procedure as it is with an explicit model and is not commonly included in match-day analysis. It is also not easy to measure the uncertainty of a match-day estimate. 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. The variability between matching days does not show any error if the analysis is carried out for a sample of clients instead of for the entire population and, therefore, it is very complex to determine the true uncertainty.

1.3.1.1 Proxy Day Approach

This method selects a specific day to represent the user's hourly loads during the day of the DR event. The day of the performance must have similar characteristics to the day of the DR event. Some of the key aspects for selecting this proxy day are temperature, day of the week, etc. Most methods currently in use limit the period that may be considered when selecting the proxy day to the prior sixty days.

1.3.1.2 *Previous Days Approach*

By averaging the hourly load data for a subset of days included in a historical period prior to the DR event, we create the baseline in this method. The selected set of days must be of the same type as the day of the DR event. That is, the baseline load curve is the result of the average of the hourly values calculated from the user's previous actual loads. Here is an example [20]:

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 8 Example of hourly baseline construction from average loads

1.3.1.3 *Average Daily Energy Usage Approach*

This method uses the most representative days to calculate the baseline according to the matching criteria. These days are selected depending on the daily demand as compared with the daily demand of the day before the DR event. 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 the ratio between the daily load of the suitable days and the selected day. Additionally, for the selection of comparable days the ratio between the daily load of the suitable days and the selected day is also considered.

When these values have been selected, the daily ratio between them is calculated, as shown in the figure below [20].

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 9 Example of days' selection for baseline construction

Once this process is done, the baseline is calculated as shown in the following figure. It is done in a similar way to the PJM methods, by averaging the hourly load of the days with the five highest daily ratios [20].

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 10 Example of baseline construction from average loads

1.3.2 Regression analysis

This method is widely used to derive the user's demand profile during an event day. From an accuracy point of view, this allows a DR programme to use advanced statistical tools to calculate a baseline, leading to the highest degree of accuracy. If we take complexity into account, it is less accepted by consumers and therefore, understanding the relationship between their actual curtailment efforts and the performance for which they are credited is progressively more difficult. The baseline calculation difficulty also appears, due to the data requirements of this regression approach, which limits the ability of understanding event performance in near real-time. The data collected to create the baseline can be divided into two groups:

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

1.3.2.1 Individual regression analysis

This type of analysis fits a regression model to the data of an individual client for a year or season and is suitable for frESCO efficiency PMV. A simple model defines 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). For a more complex model, the cooling degree-day base can be defined by a better regression fit, which can contain calendar and day-of-week effects, lag terms that inform about the temperature over multiple hours, and humidity. Typically, the individual regression models are fit to loads on non-event days and are used with the conditions of each event day to give an estimate of the customer's load that would have taken place on that day without the DR event. To estimate the impact, the variation between the modelled and measured load is performed for each hour of the event period. If load data is only available for a section of participating customers, the reduction in total load is approximated by expanding the sample that comes from the individual customer impacts.

When load data is available to all users, load reduction is the sum of impacts for each user. This model can incorporate event day terms and adjust to any type of event. Nevertheless, unless there are multiple event days spanning a wide range of the other terms in the model, incorporating 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.

In individual regression models there is the possibility of a more accurate estimation error because the dispersion of the results reflects both the spread of individual responses and the estimation "noise" or random errors. However, the opposite situation can occur; statistically substantial effects are observed for large numbers of control group customers who had no event to respond to. Generally, if an identical model structure is applied for both individual and pooled regressions, the coefficients of the pooled fit will be roughly the average coefficients of the individual fits. Furthermore, if both models use the same independent variables, the models are even more identical. In addition, this method has some other advantages:

- The information is customized for each customer. These results give a database that contributes to richer analysis since it gives information from the observation of results distributions rather than averages only. Specific customer results can also be related to other customer information.
- Significant results can more easily be created for groups of customers whose load patterns are divergent, since each is modelled separately.
- Results can be included into any segments that may be of interest after that initial analysis is completed.
- Customers for which the basic regression shape is not a good description can be detected by model diagnostics and treated separately.
- Weather response terms such as the best degree-day base can be defined separately for each customer, achieving better and more significant overall fits.
- Ex-ante results can be derived by fitting individual regressions to design or severe temperature data and then aggregating the resulting estimates.
- Results serve to understand the relative customer situation in programmes that promote behavioural changes and how they occur.

1.3.2.2 Pooled regression analysis

The structure of this method is similar to the individual regression analysis model, but in this case, there is a single model for a large group of participants. To describe the average load pattern of all clients, a single set of coefficients is used and variables of the day of the event are usually included. With the larger sample size, it is possible to recognize indicators that may not be well determined otherwise.

A pooled model approach is more complex when compared to an individual one, because there will be correlations between series and patterns in the regression errors that, if not appropriately accounted for, can lead to much more unprecise estimates. This is especially true for single models applied to large number of customers. However, there are also some advantages to using this 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 accurate.
- 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 resume of all the data analysis techniques for baseline estimation, as well as advantages and disadvantages of each of them. The individual data availability in frESCO enables the individual regression models rather than the pooled models, thus issuing customised assessments per user and a more human-centric approach.

Table 2 Summary of data analysis techniques for baseline estimation

Exploratory analysis	PROs	CONs
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.

1.4 Baseline adjustments

In order to improve and adjust the baseline depending on the conditions on the day of the DR event, new calculations should be applied to the initial estimate. This is done by determining the difference between the actual customer load and the calculated baseline. Once defined, the calculation that makes the pre-event period estimated load equal to the pre-event period observed baseline is applied to the event period. There are different variables that influence these settings, such as humidity, day of week, temperature, etc. There are two basic types of baseline settings:

- 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%).

Normally, in the period before the event, the adjusted baseline coincides with the actual load and it is possible that there is a relationship with the same day of the event or the day before. However, the NAESB guidance notes that the adjustment window will begin no more than four hours before the event unfolds. Examples of snap 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)

If we have 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. In cases of sensitive loads, such as heating or cooling, it is advisable to have adjustments based on the load observed before the event. These types of situations are common for residential customers. It is advisable to use the climate or climatic characteristics as the basis of the adjustment system.

It is difficult to measure the nature and scope of the adjustments but conceptually they depend on the timing of the notification along with the specification of the adjustment window and method.

Event effects during the setting window can occur in the following ways:

- From the time of event notification up to the event start: once the event is notified up to the starting instant, the building cools down to a higher level than normal. This is a legitimate, reasonable response that makes programme participation more viable for the building. However, if the adjustment window includes hours between the notification and the event, the baseline will be inflated, and the load reduction overestimated.
- Anticipatory increase prior to the notification: from the moment a very hot day is predicted, the building during the morning gets colder than usual, which makes a DR

event likely. 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.

- Manipulative increase: A DR asset deliberately ramps up 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. A major disadvantage is that the opportunities for this method are limited, depending on whether the adjustment window is set to end before notification.

Weather conditions on the day of the event could be another possibility for adjustment, provided that previous responses are not allowed to influence 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 main advantage is the adjustment to the weather conditions of the day of the event, without requiring a load prior to the event. The main disadvantage is that it does not adjust for an asset's natural, non-distorting operations on the event day, only adjusts for weather.

However, there is a limit to both additive and multiplicative adjustments. In contrast, asymmetric adjustment is only applied if the adjustment value increases the baseline. Another limit to any adjustment could be the use of a stop. 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 maximum, that additive adjustment would be restricted: if the cap were 20%, then the additive adjustment would be 20 kW.

This setting suffers a risk of malfunction if there is a very changeable climate or unforeseen changes, this malfunction occurs due to peak demand on a hot day after a few colder days.

In these cases, where the client has a variable demand for the weather, it is very likely that the real demand is much higher than that observed during the time prior to the event.

However, if there is a limit, it is likely that the customer will receive little or no credit despite taking curtailment action and delivering real value to the grid.

In summary, the baselines based on averages of the last days are poor, this is especially the case for residential customers with significant climate sensitivity. In the case of programs with a large number of clients, the experimental design is the alternative with greater precision.

In the particular case of the frESCO project, baselines for a DR use pre-trained models that are trained for 10 – 15 days to adapt it to the weather and user demand profile. The trained model is used to calculate the baseline consumption value with data prior to the event (1 hour maximum), thus gathering the conditions of the load or system just before the event. Model training is dynamic and is performed continually so as to adjust the model to the seasonal specific conditions. Hence, no adjustments are actually required for Flexibility verification and measurement.

This is not the case for longer term baselines of EE measures where adjustments are actually needed to correct models and take into account the changes occurring from the reference period (baseline) compared to the measurement period.

1.5 Uncertainty

There are errors in the measurements because the instruments in charge of this task are not perfect. These errors are the difference between the observed and the true energy use. The management of these errors is important since they prevent the exact determination of savings. The uncertainty of the measurements can be managed by controlling random errors and bias in the data. The former is affected by the measurement equipment, techniques, and design of the sampling procedure. On the other hand, data bias is affected by the quality of the measurement data, the assumptions and the subsequent analysis. Investing in error reduction normally increases the cost of PMV, which is why the PMV plan must include a method for quantifying errors, in such a way that the resulting error is acceptable. According to [12], 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. Normally, the magnitude of such errors is defined by manufacturer's specifications and is managed through periodic re-calibrations.

- Modelling: uncertainty may be due to the inclusion of irrelevant variables or the exclusion of relevant variables resulting in the inability of mathematical forms to explain variations in energy use.
- Sampling: Using a sample from the entire population can provide errors due to variation in values within the population or biased sampling. Sampling can be done in a physical sense or in a temporal sense. One of the most typical requirements for the definition of sampling precision is that the load should be estimated to have a confidence interval that is ± 10% of the estimate at a 90% confidence level.
- Interactive effects (outside the measurement scope) which are not totally included in the savings computation methodology.

Correct reporting of savings requires a level of accuracy and confidence. We speak of confidence when we refer to the probability that these savings are 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 ± 10% of 1,000". If there is no confidence level, this statement of statistical precision is meaningless. That is why the PMV process can generate high precision with little 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. Savings should be greater than twice the standard error of the reference value. If the variation is excessive, there is a random behaviour in the use of energy and the calculated savings are not reliable. To address the uncertainty issues, the following solutions may be applied:

- more precise measurement equipment
- more independent variables in any mathematical model
- more complex mathematical models (like higher degree polynomials)
- larger sample sizes
- a PMV option that is less affected by unknown variables.