Ref. Ares(2025)69822 - 07/01/2025

SmartSPIN

Smart energy services to solve the **Sp**lit **IN**centive problem in the commercial rented sector

Call topic: H2020-LC-SC3-2018-2019-2020
Start date of the project: 01/09/2021
Duration: 40 months

D5.3 – EVALUATION OF THE M&V FRAMEWORK TO VERIFY THE EFFECTIVENESS OF THE SMARTSPIN PREDICTIVE ALGORITHMS

Due date	**31/08/2024**	**Delivery date**	**07/01/2025**
Work package	WP5 – Measure and verify impact		
Responsible Author(s)	SOTIRIS PAPADELIS (HEBES INTELLIGENCE)		
Contributor(s)	HARA TZOUVARA (HEBES INTELLIGENCE)		
Dissemination level	PUBLIC		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

Version and amendments history

Version	Date (MM/DD/YYYY)	Created/Amended by	Changes
0.1	31/08/2024	Sotiris Papadelis	First version of methodology and results
0.2	15/10/2024	Sotiris Papadelis	Event detection method finalized
0.3	31/12/2024	Sotiris Papadelis	Complete draft text of deliverable
0.4	06/01/2025	Luciano De Tommasi	Internal quality review
0.5	07/01/2025	Sotiris Papadelis	Improvement after internal quality review
1.0	07/01/2025	Luciano De Tommasi	Final Review and submission

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

Table of Contents

EXECUTIVE SUMMARY 6
1 INTRODUCTION 7
1.1 The Workflow of the SmartSPIN M&V Methodology 7
1.2 The Goal of the Deliverable 8
1.3 The Structure of the Deliverable 9
2 OUTLIER DETECTION 10
2.1 The Outlier Detection Method 10
2.2 Outlier Detection for Data Quality Monitoring 12
3 ACTIVITY ESTIMATION 14
3.1 Why Activity Estimation is Needed 14
3.2 How Activity is Estimated 18
3.3 The Possible Physical Meanings for the Activity Feature 21
3.4 The Impact of Sample Size 23
4 EVENT DETECTION AND CHARACTERIZATION 26
4.1 Introduction 26
4.2 Application of Event Detection 26
4.3 Event Characterization 28
5 METHODOLOGY DEMONSTRATION 30
6 CONCLUSIONS 34
APPENDIX 35
Predictive Model Evaluation 35

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

List of Figures

Figure 1.1: The workflow of the SmartSPIN M&V methodology 8
Figure 2.1: Evaluation metrics for the outlier detection method 11
Figure 2.2: Evaluation metrics for a prediction-based outlier detection method........ 11
Figure 2.3: Average observation weights for weekly data batches of the Irish pilot........ 12
Figure 2.4: Outlier scores of the Irish pilot dataset 12
Figure 2.5: Average observation weights for weekly data batches of the La Gavia pilot 13
Figure 2.6: Outlier scores of the La Gavia pilot dataset 13
Figure 3.1: Prediction-based M&V without events........ 15
Figure 3.2: Prediction-based M&V with events........ 15
Figure 3.3: The interrelation between mapping and impact variables 16
Figure 3.4: Relationship between energy consumption and outdoor temperature for La Gavia 17
Figure 3.5: The structure of the SmartSPIN M&V method 18
Figure 3.6: The activity envelope of the electricity consumption for HVAC of La Gavia in 2023 18
Figure 3.7: The raw activity for the HVAC electricity consumption of La Gavia in 2023 19
Figure 3.8: The binned activity for the HVAC electricity consumption of La Gavia in 2023 20
Figure 3.9: A 2-stage modeling process for energy consumption prediction 20
Figure 3.10: The interplay between the granularity of the activity feature and the prediction results 21
Figure 3.11: The smoothed activity for the HVAC electricity consumption of La Gavia in 2023 21
Figure 3.12: Decomposition of the electricity consumption for HVAC of La Gavia in 2023 22
Figure 3.13: Decomposition of the electricity consumption for HVAC of La Gavia when activity is used 22
Figure 3.14: Contribution of the number of occupants to the estimated activity........ 23
Figure 3.15: The training/testing scheme for evaluating the in-sample accuracy of the activity estimation... 24
Figure 3.16: Impact of sample size on the in-sample accuracy of the activity estimation 24
Figure 4.1: Average daily energy consumption profiles per year for the La Gavia building 27
Figure 4.2: Differences between expected and observed activity levels for the La Gavia building........ 27
Figure 4.3: Detected events for the La Gavia pilot 27
Figure 4.4: Average activity profiles for reference and test samples........ 28
Figure 4.5: Observed and expected energy consumption for the test sample 29
Figure 4.6: Supporting plots for characterizing the detected event 29
Figure 5.1: Activity estimation after an event........ 30
Figure 5.2: Activity estimation for the 2020 lockdown period 31
Figure 5.3: Counterfactual energy consumption estimation for the 2020 lockdown period........ 31
Figure 5.4: Activity estimation after a retrofit 32
Figure 5.5: Observed and counterfactual energy consumption for June 2022 of La Gavia pilot 32
Figure 5.6: Comparison of estimated energy savings and metered generation from onsite solar PV 33
Figure 5.7: Estimated energy savings using a prediction-based model 33

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

List of Abbreviations

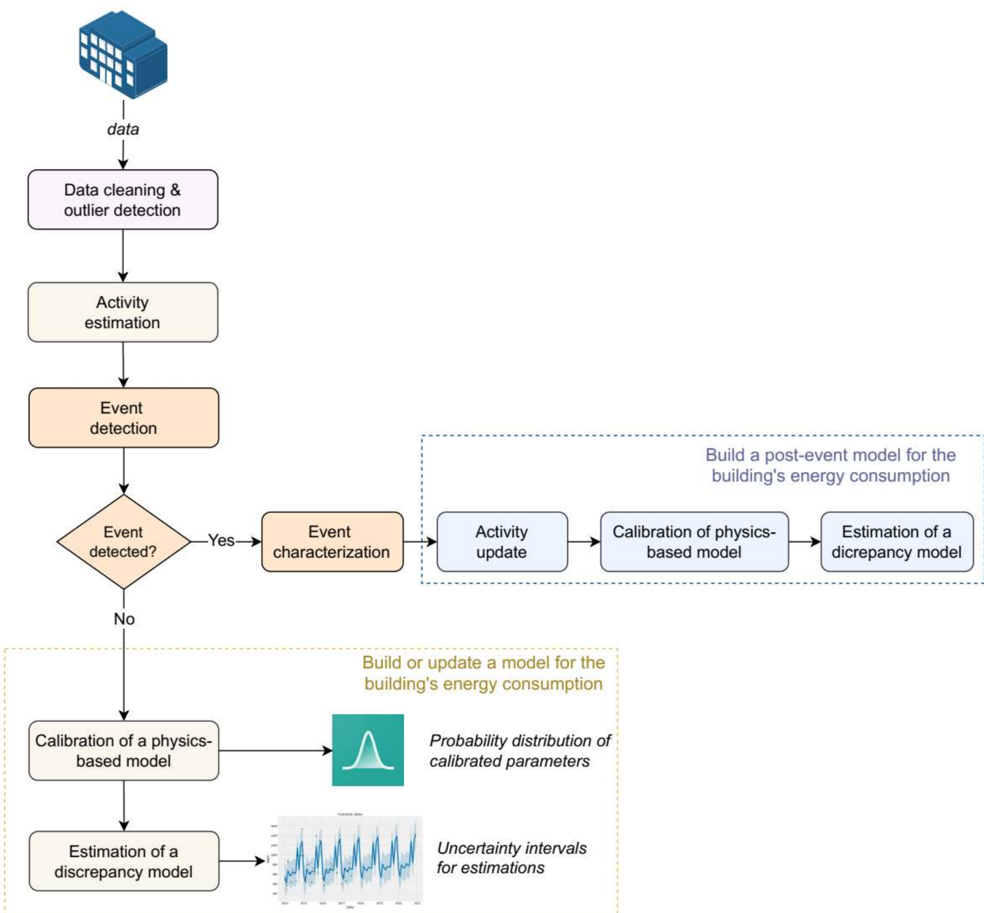
BDG2	The Building Data Genome 2 dataset
CV(RMSE)	Coefficient of Variation of Root Mean Squared Error
IPMVP	International Performance Measurement and Verification Protocol
M&V	Measurement and Verification
RMSE	Root Mean Squared Error

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

EXECUTIVE SUMMARY

This deliverable presents an evaluation of the M&V methodology that was developed by the SmartSPIN project. The methodology represents an innovative approach for the M&V of energy savings produced by energy retrofits of buildings, and it constitutes a significant step forward for the state-of-play in M&V compared to both the current literature and the professional practice.

The utility of the methodology stems from the fact that it provides solutions for every step of the M&V workflow: evaluation of data adequacy, baseline model development, non-routine event detection and characterization, counterfactual energy consumption estimation, and adaptation/correction due to events. The workflow of the methodology is outlined below:



The innovation of the methodology stems from the way it categorizes the different events that may take place inside a building and change the dynamics of its energy consumption, as well as from its ability to adapt to these events so that to accurately estimate energy savings from energy efficiency upgrades in a building's envelope and HVAC systems.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

1 INTRODUCTION

1.1 THE WORKFLOW OF THE SMARTSPIN M&V METHODOLOGY

The SmartSPIN Measurement and Verification (M&V) methodology is based on three (3) main components:

1. **Detecting events** that signify that the dynamics of the building's energy consumption have changed. The methodology treats energy efficiency upgrades, changes in the building's operating schedule and changes in the density of the consumption (such as more or fewer people being present in the building during peak hours) in the same way: events that split the available data into pre-event and post-event periods.

2. **Characterizing events** according to the source of the changes in energy consumption. There are two (2) types of events. The first type concerns changes in the energy efficiency of the building. If these changes are deliberate, such as when an energy retrofit has taken place, the goal of the M&V method is to estimate the impact of the event. If the changes are not deliberate, such as when the HVAC equipment is malfunctioning, their impact will affect the estimated energy savings. The second type concerns changes in the activity levels of the building, such as when the HVAC system remains the same, but it is used more frequently or more intensively due to increased internal heat gains. In this case, the goal of the M&V method is to adapt/correct the model for estimating the building's energy consumption.

3. **Building a model for the building's energy consumption** using a combination of mechanistic modeling and statistical modeling. A mechanistic or first-principles model uses physics-based formulations to describe the relationships that govern an underlying system or process. As a result, its parameters have physical meaning. For example, a mechanistic model for a building envelope would include parameters for its U-value, its geometry, its thermal mass, and so on. A mechanistic model for the HVAC system would include parameters for its efficiency, the temperature setpoints, as well as the settings of the temperature control. In contrast, statistical models – such as regression or more flexible machine learning models – approximate the relationships that govern an underlying system directly from the observed data.

The workflow of the SmartSPIN M&V methodology is outlined in Fig. 1.1 below:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

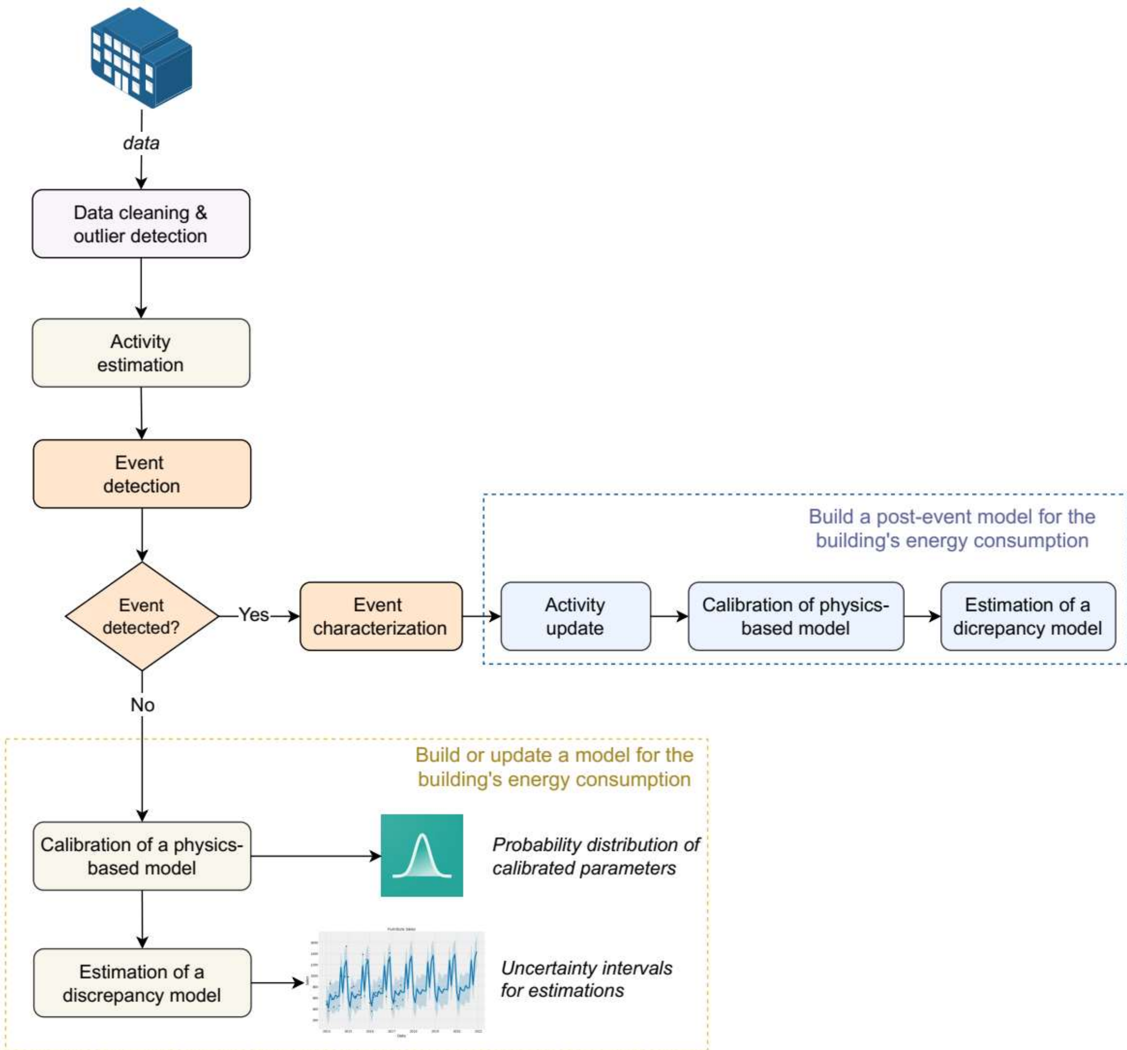


Figure 1.1: The workflow of the SmartSPIN M&V methodology

1.2 THE GOAL OF THE DELIVERABLE

The goal of this deliverable is to evaluate selected parts of the SmartSPIN M&V workflow using data from the project's pilot buildings. When data from a large number of diverse buildings was required, the following datasets were also utilized:

- The Building Data Genome 2 (BDG2) dataset. The BDG2 dataset is an open dataset made up of 3,053 energy meters from 1,636 buildings. The time range of the time-series data spans over two full years (2016 and 2017) and the frequency is hourly measurements of electricity, heating and cooling water, steam, and irrigation meters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

- The open dataset from the Kaggle competition "Large-scale Energy Anomaly Detection – Identify energy usage anomalies in hourly smart electricity meter readings"[1]. This dataset was used in the evaluation of the outlier detection method.

1.3 THE STRUCTURE OF THE DELIVERABLE

The deliverable is composed of the following chapters:

Chapter 2 "Outlier Detection" presents the approach and evaluation results of the outlier detection method that is included in the SmartSPIN M&V workflow.

Chapter 3 "Activity Estimation" explains the way the SmartSPIN M&V methodology estimates the counterfactual energy consumption (i.e. what would the energy consumption of a building have been had an energy efficiency intervention not occurred), as well as the way it adapts to changes in the activity levels of the building.

Chapter 4 "Event Detection and Characterization" presents the approach that the SmartSPIN M&V methodology uses for detecting changes in the dynamics of a building's energy consumption, as well as the tools that it offers for users to decide whether the counterfactual energy consumption model should adapt to these changes.

Chapter 5 "Methodology Demonstration" presents step-by-step the application of the SmartSPIN M&V methodology on the data from the project's building pilots.

Chapter 6 "Conclusions" summarizes the deliverable.

"Appendix" provides information on the commonly used metrics for evaluating the accuracy of a predictive model for energy consumption.

[1] https://www.kaggle.com/competitions/energy-anomaly-detection

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

2 OUTLIER DETECTION

2.1 THE OUTLIER DETECTION METHOD

The M&V workflow starts by screening any new batch of data for outliers. The SmartSPIN outlier detection method is based on Sugiyama and Borgwardt (2013)[2], and Muhr and Affenzeller (2022)[3].

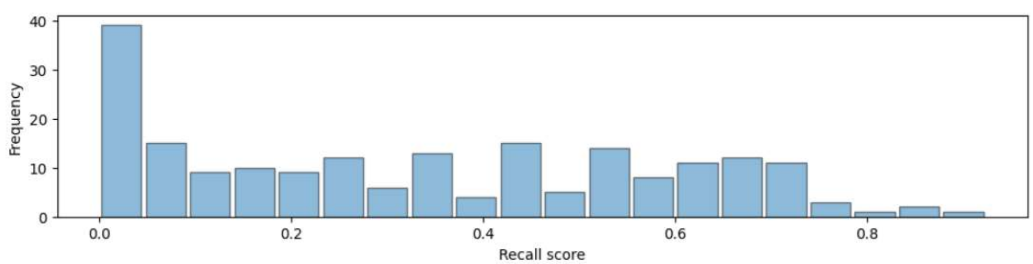
The method is composed of the following steps:

- Randomly select a number of observations in the previously collected data and compute the Euclidean distance of each and every observation in the batch to each and every member of the selected sample.
- For each observation in the batch, use the smallest distance as an indicator of outlierness.
- Fit an exponential distribution on the outlierness scores and use the distribution's cumulative distribution function to scale the scores in the [0, 1] interval: the larger the score, the more probable is that the observation is an outlier.

In order to evaluate the outlier detection method, it was tested on the **200 buildings** of the dataset from the Kaggle competition "Large-scale Energy Anomaly Detection – Identify energy usage anomalies in hourly smart electricity meter readings". This dataset already includes information about the actual outliers, so it can be used for evaluating the *recall* and *accuracy* of the method:

- Recall measures the ability of the algorithm to find all the outlier samples. The best value is 1.0 and the worst value is 0.0.
- Accuracy is the fraction of correct predictions. The best value is 1.0.

Fig. 2.1 shows histograms of the recall and accuracy metrics for all the 200 buildings.



---

[2] Mahito Sugiyama and Karsten Borgwardt (2013) "Rapid Distance-Based Outlier Detection via Sampling", Advances in Neural Information Processing Systems 26 (NIPS 2013)
[3] David Muhr and Michael Affenzeller (2022) "Little data is often enough for distance-based outlier detection," Procedia Computer Science, Vol. 200, pp 984-992

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

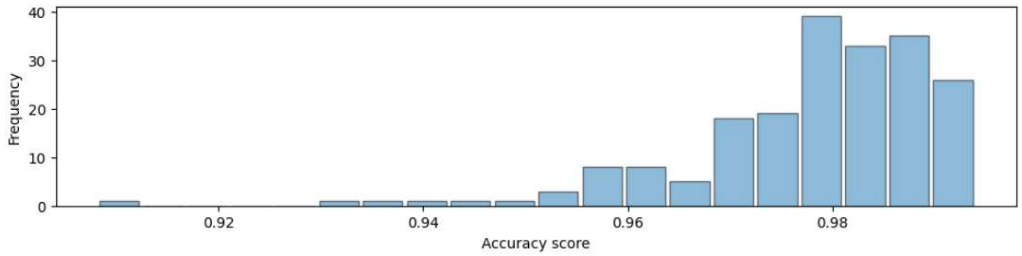


Figure 2.1: Evaluation metrics for the outlier detection method

The results indicate that the algorithm is conservative when marking observations as outliers (average recall is 0.33), but it is very accurate when doing so (average accuracy is 0.98). This is a desirable property for M&V analysis because identifying some outliers with a low false positive rate is better than identifying higher numbers with more false positives. In any case, the outlier detection step should be regarded as a filter to improve the quality of the data that will pass on to the next workflow steps.

In addition, the same dataset was evaluated with an alternative event detection approach that relied on fitting a predictive model on the available data and calculating the prediction errors. The errors are then normalized to reside in the [0, 1] interval. The results were similar: average recall was 0.28 and average accuracy 0.98:

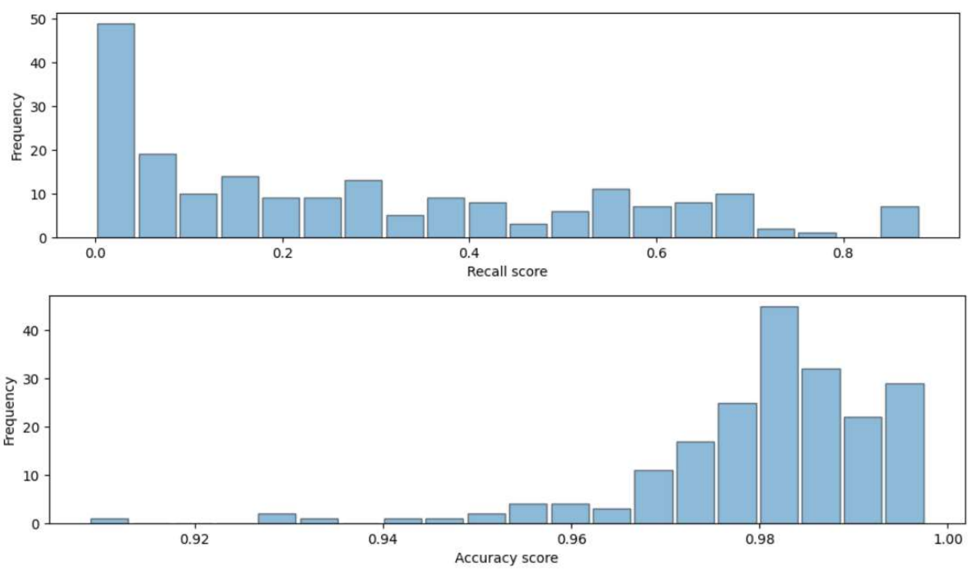


Figure 2.2: Evaluation metrics for a prediction-based outlier detection method

However, the distance-based outlier detection method was adopted as part of the SmartSPIN M&V approach because it is faster and does not require feature engineering for improved accuracy (while predictive models do need feature engineering).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

2.2 Outlier Detection for Data Quality Monitoring

The outlier detection method of the SmartSPIN M&V approach is meant to be applied on each data batch as data becomes available to the user. In this way, a run chart for the percentage of the observations marked as outliers can be maintained. In addition, the SmartSPIN outlier detection method produces outlier scores: the higher the score, the higher the probability that the corresponding observation is an outlier. Consequently, the higher the outlier score, the lower the weight[4] that should be assigned to the observation when used for training predictive models. Thinking in terms of observation weights is generally more helpful than thinking in terms of outliers.

The following plot shows a run chart for the weekly average weight of the available data for the Irish pilot:

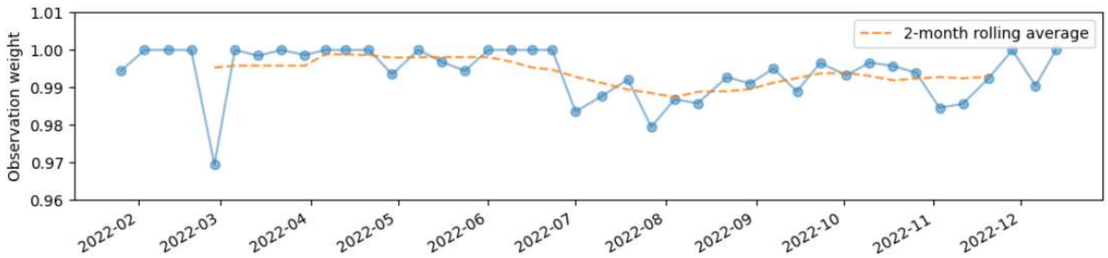


Figure 2.3: Average observation weights for weekly data batches of the Irish pilot

When the average weight of the observations becomes (relatively) low, something has changed in the data collection process. In this way, outlier detection is also a way to monitor the quality of the input data over time. The plot of Fig. 2.4 shows the outlier scores of the Irish pilot data for the same period:

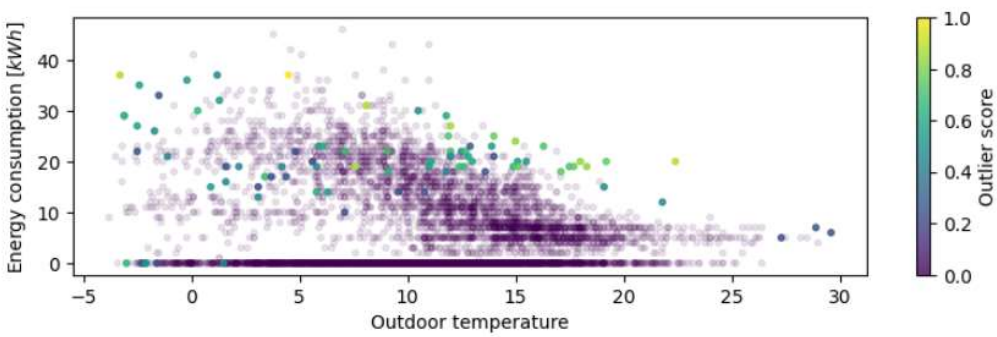


Figure 2.4: Outlier scores of the Irish pilot dataset

[4] A simple formula is *weight = 1 - score*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

Similarly, the next plot shows a run chart for the weekly average weight of the available data for the La Gavia pilot:

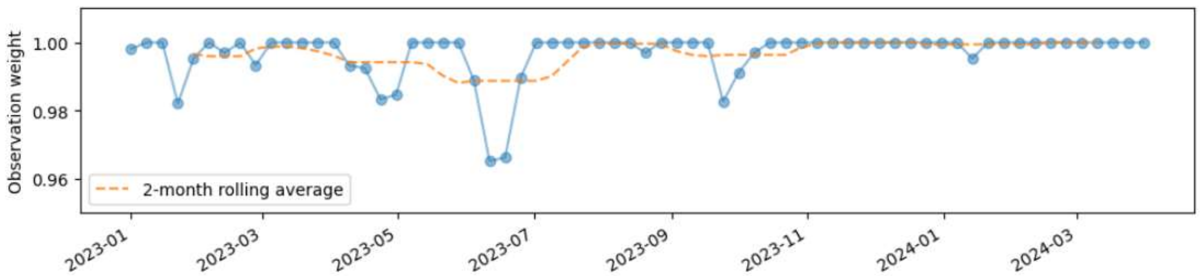


Figure 2.5: Average observation weights for weekly data batches of the La Gavia pilot

Finally, the plot in Fig. 2.6 shows the outlier scores of the La Gavia pilot data for the same period.

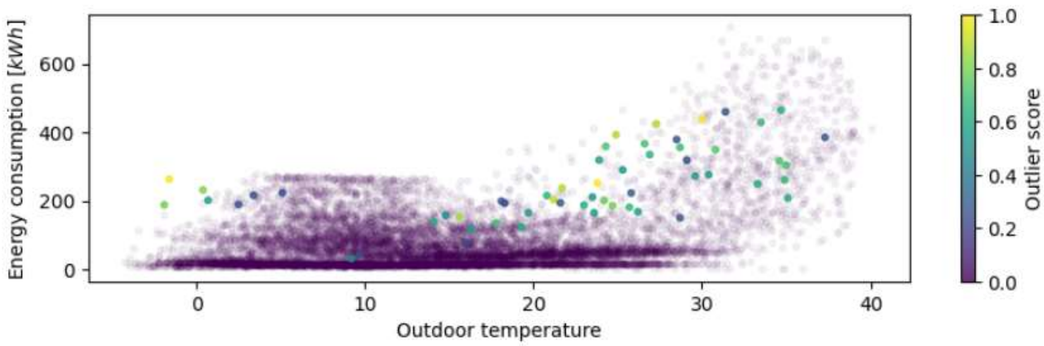


Figure 2.6: Outlier scores of the La Gavia pilot dataset

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

3 ACTIVITY ESTIMATION

3.1 Why Activity Estimation is Needed

Measurement and verification of energy savings is fundamentally an impact assessment problem, where the goal is to estimate the counterfactual energy consumption – i.e. what would the energy consumption of a building have been had an energy efficiency intervention not occurred – using two sources of information:

- Occupancy-dependent information. Energy consumption reflects events and operations that take place inside the building and, as a result, recurrent events and routine operations lead to daily, weekly and yearly seasonality in the consumption that can be exploited (to the extent that the seasonality remains unaffected by the efficiency intervention).
- Occupancy-independent information that was predictive of the building's energy consumption prior to the intervention. The most often utilized information is the outdoor air temperature.

The relationship for the quantification of the energy savings from a retrofit project is commonly defined as:

Energy savings = Counterfactual consumption based on pre-retrofit (baseline) period data
− Actual consumption during the post-retrofit (reporting) period
+/− Non-routine adjustments

Some definitions:

- **Baseline period**: The period of time prior to the intervention during which data is gathered so as to determine the relationship between energy consumption and the different independent variables that can predict it.
- **Reporting period**: The period of time following the intervention during which data is gathered so as to calculate energy savings (avoided energy use).
- **Non-routine adjustments**: Non-routine adjustments account for unexpected changes in energy use. The fact that the changes are "unexpected" means that the driving factors of these changes were not included as independent variables in the baseline predictive model. By definition, they render the predictive model less relevant and require adjustments either to the model or to the baseline period energy data so as to reflect the same set of conditions as the ones observed during the post-intervention period.

The most common approach to the M&V of energy savings from a retrofit is to treat it as a prediction task. In this case, a predictive model is developed using pre-retrofit data to predict the building's energy consumption given the values of a set of observable variables. In most cases,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

these variables correspond to calendar features (as a proxy for the operation schedule), such as the week of the year and the hour of the week, and outdoor temperature information. After the energy retrofit, this model is used to predict the counterfactual consumption. The difference between the counterfactual and the actual consumption is regarded as the avoided energy usage that can be attributed to the intervention. This approach works well when no other event has taken place and, as a result, the counterfactual prediction can be constructed using only pre-retrofit data (Fig. 3.1).

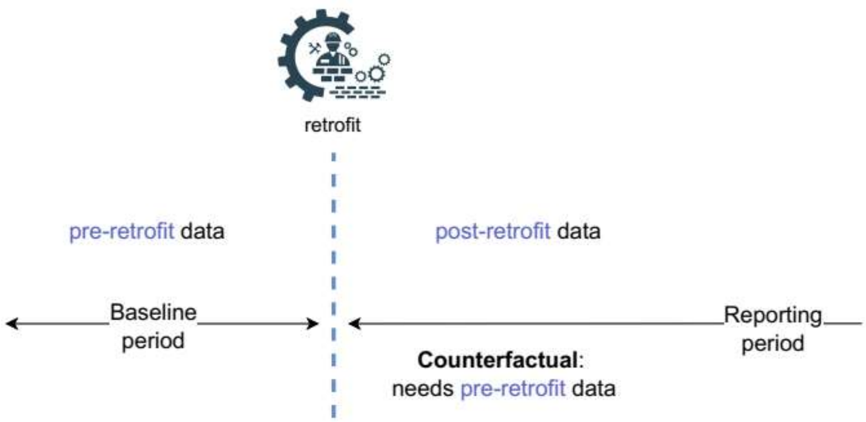


Figure 3.1: Prediction-based M&V without events

However, this approach makes it difficult to adapt to events that affect energy consumption independently of the energy retrofit (as an example, a change in the maximum number of people in the building). The reason for this is that the prediction-based approach requires us to train the predictive model on data that is not available (Fig. 3.2).

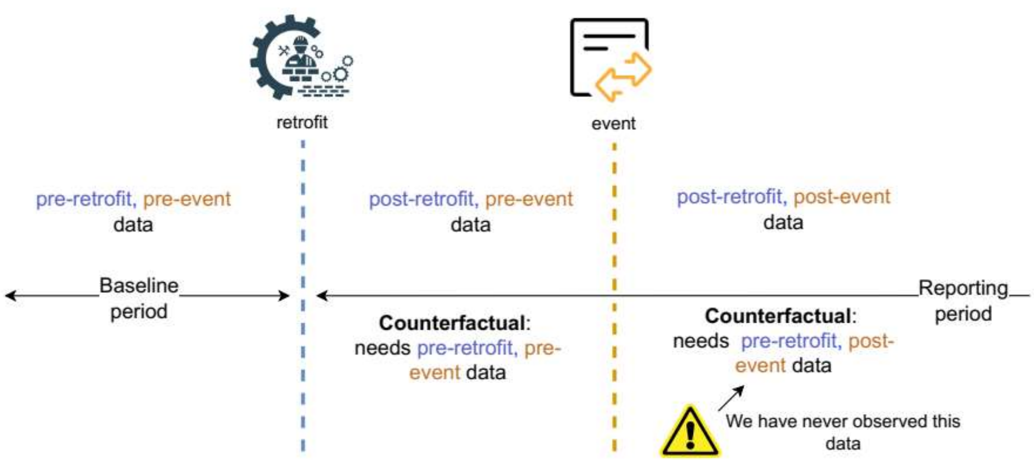


Figure 3.2: Prediction-based M&V with events

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

Contrary to the prediction-based approach, the SmartSPIN M&V method defines the M&V goal as one of devising and applying a mapping from states and conditions after an energy efficiency intervention to states and conditions before it. The impact of the intervention is the difference in energy consumption between matching states and conditions. To this end, the methodology makes a distinction between mapping variables and impact variables:

- **Mapping variables** are defined as the observed and/or unobserved variables that must be similar between the pre- and the post-retrofit periods so that a counterfactual estimation for energy consumption would make sense. In other words, mapping variables help in mapping from states and conditions after the intervention to states and conditions before it. The weather, the building's occupancy schedule and the intensity of the occupancy (e.g. number of people or level of plug loads that correspond to full occupancy) are examples of mapping variables. As a general rule, an M&V model must be able to adapt to changes in the mapping variables.

- **Impact variables** are defined as the observed and/or unobserved variables that directly affect the impact of the energy efficiency intervention. The U-value of the building's envelope, as well as the efficiency and/or the control strategy of the HVAC system are examples of impact variables.

The interrelation between mapping and impact variables can be summarized using the conceptual model of the following diagram:

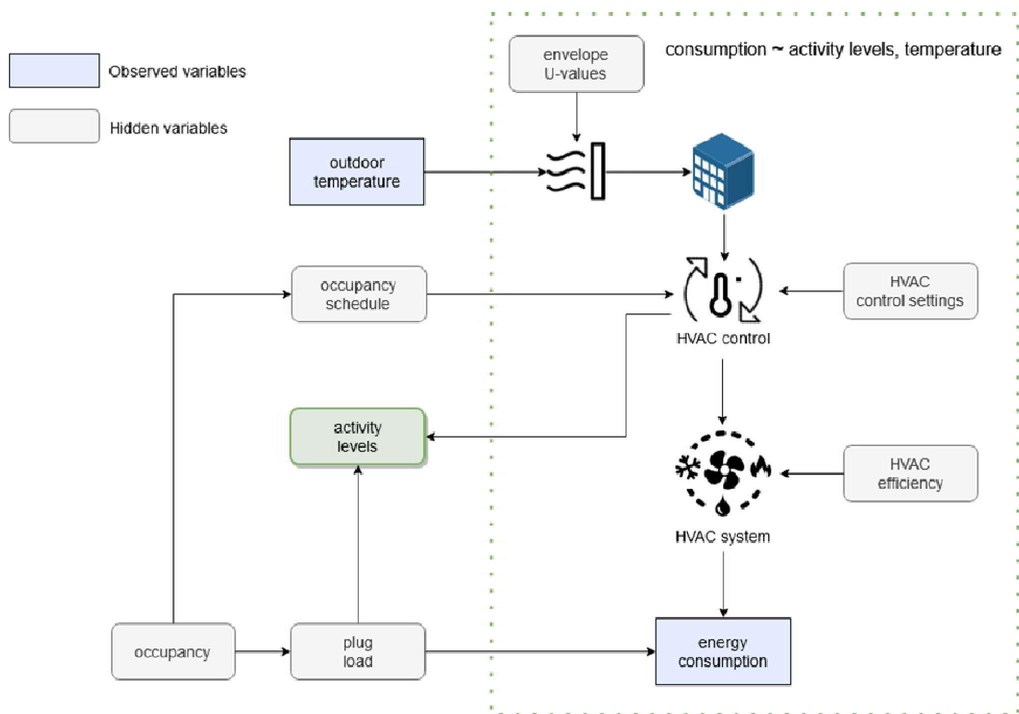


Figure 3.3: The interrelation between mapping and impact variables

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

A central aspect of the developed M&V methodology is the estimation and utilization of a mapping variable called *activity*. The activity feature explains the variation in the energy consumption of a building that cannot be explained by external conditions (such as outdoor temperature).

To better understand the activity feature, let's consider the relationship between the energy consumption for HVAC and the outdoor temperature for the La Gavia pilot during 2023 (Fig. 3.4). For any given outdoor temperature, energy consumption varies significantly. A large part of the variation reflects differences in occupancy. However, even for the same occupancy status, energy consumption varies. This variation reflects differences in activity levels, as well as unstructured stochasticity (noise).

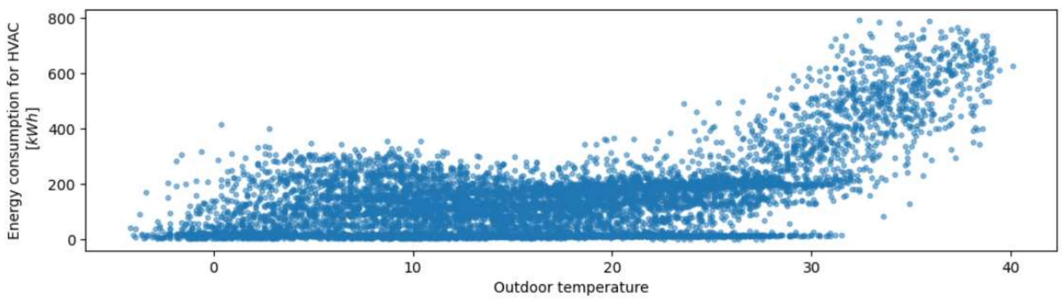


Figure 3.4: Relationship between energy consumption for HVAC and outdoor temperature for La Gavia in 2023

The SmartSPIN M&V methodology calculates the counterfactual energy consumption of a building through the following steps:

Baseline period

1. Use pre-retrofit data about the external conditions and the observed energy consumption to estimate activity levels.
2. De-noise activity levels to remove stochasticity.
3. Use pre-retrofit data about the external conditions, data about the internal conditions and the de-noised activity to train a predictive model for the pre-retrofit energy consumption.

Reporting period

4. Use post-retrofit data about the external conditions and the observed energy consumption to estimate activity levels.
5. De-noise activity levels to remove stochasticity.
6. Apply the predictive model for the pre-retrofit energy consumption on post-retrofit data about the external conditions, data about the internal conditions and the de-noised activity to estimate the counterfactual energy consumption.

The main idea behind the SmartSPIN M&V methodology is that we can have a valid, counterfactual prediction model by comparing the energy consumption before and after an

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

intervention for similar activity levels and similar values of the variables reflecting external conditions (such as the outdoor temperature) or internal operations (such as number of occupants). The M&V model is derived by combining the activity estimation of the post-retrofit data with an energy consumption model that is trained using the pre-retrofit data. This concept is summarized in the next diagram:

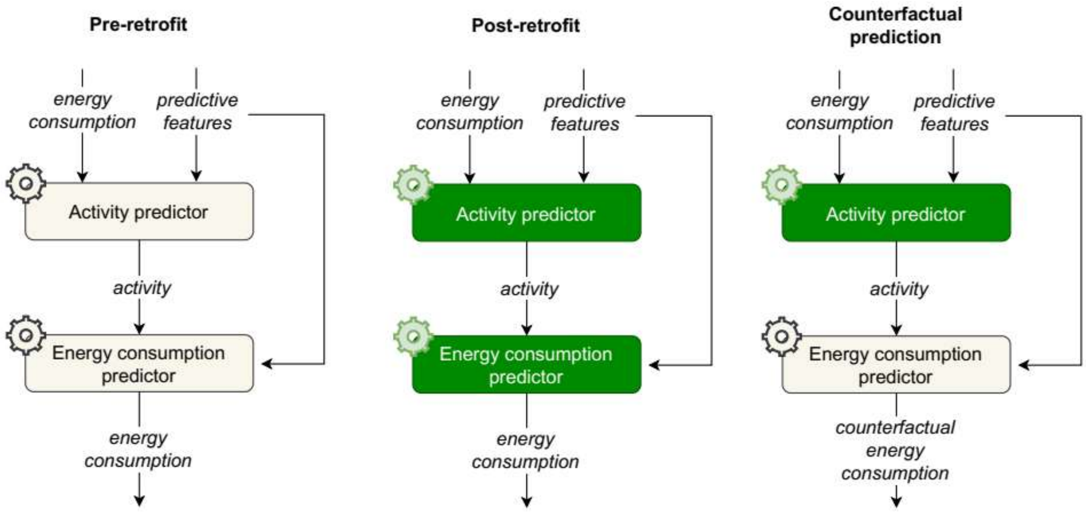


Figure 3.5: The structure of the SmartSPIN M&V method

3.2 How Activity is Estimated

The first step in estimating activity is to perform two (2) quantile regressions between the energy consumption (dependent variable) and the external conditions such as outdoor temperature (covariates). One quantile regression model targets the 0.01 quantile and the other regression model targets the 0.95 quantile. This creates an envelope for the energy consumption data. The plot of Fig. 3.6 shows the activity envelope for the HVAC electricity consumption of the La Gavia pilot.

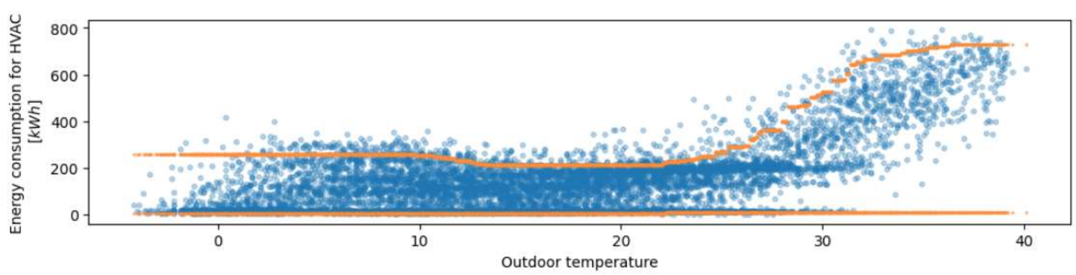


Figure 3.6: The activity envelope of the electricity consumption for HVAC of La Gavia in 2023

The electricity consumption is normalized from the interval that is defined by the envelope to an interval for 0 to 1. The output of the normalization is the **raw activity** estimation:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

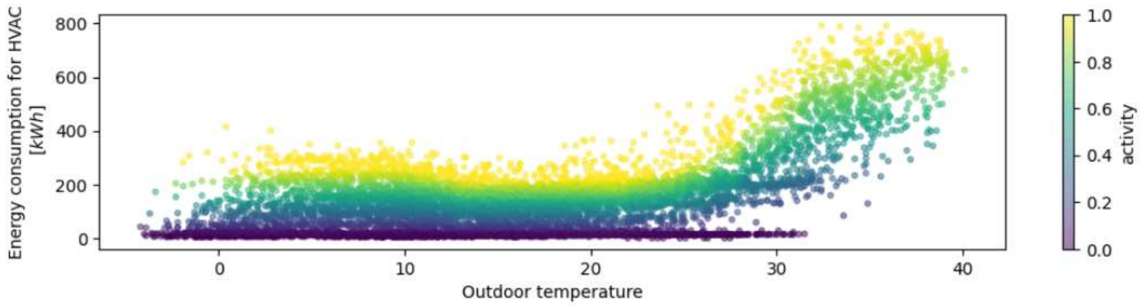


Figure 3.7: The raw activity for the HVAC electricity consumption of La Gavia in 2023

The raw activity estimation includes the **noise/stochasticity of the data**; the noise must be removed before the activity feature can be of any use.

In order to evaluate the effects of any smoothing operation, we first need an in-sample prediction of the energy consumption that we know it does not suffer from overfitting. For this, a Gradient Boosting regression model was used. The model used as covariates the outdoor air temperature, as well as the month of year (1-12), day of week (1-7) and hour of day (1-24) of the timestamps of the observed energy consumption's time-series. This is the model described in Touzani et al. (2018)[5]. To avoid overfitting, the dataset was split into a training and evaluation set, and the training process stopped when the CV(RMSE) of the evaluation set stopped improving. The trained model produced predictions with a **CV(RMSE) of 24.2%**.

As a next step, the raw activity is divided into three (3) levels: low, medium and high. All activity values inside a level are averaged so that activity is now represented only by one of the following values:

Level	Value
Low	0.08
Medium	0.5
High	0.85

The plot of Fig. 3.8 shows the new activity levels:

[5] Samir Touzani, Jessica Granderson, and Samuel Fernandes (2018) "Gradient boosting machine for modeling the energy consumption of commercial buildings," Energy and Buildings, Vol. 158, pp. 1533-1543

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

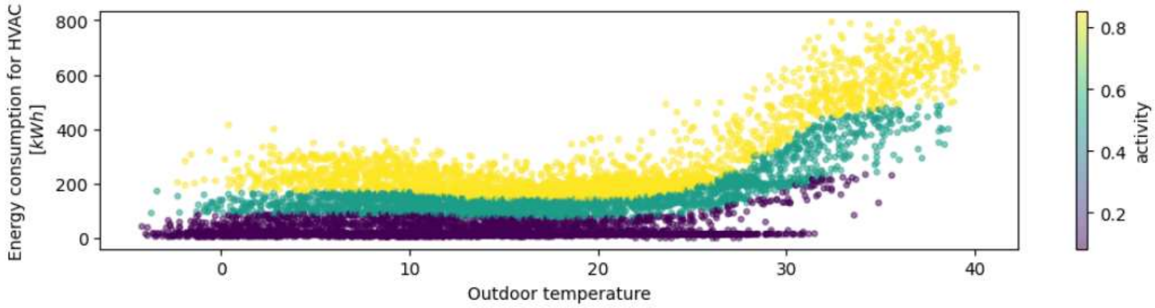


Figure 3.8: **The binned activity for the HVAC electricity consumption of La Gavia in 2023**

Instead of using calendar features directly for predicting energy consumption, we can adopt a 2-stage process: use calendar features to predict activity levels and, then, use predicted activity levels and outdoor temperature to predict energy consumption:

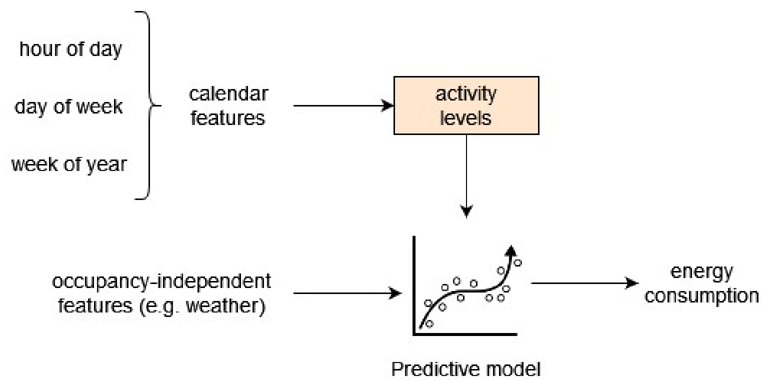


Figure 3.9: **A 2-stage modeling process for energy consumption prediction**

The 2-stage scheme produces predictions with a **CV(RMSE) of 28.5%**. However, we can now evaluate how well the mapping between calendar features and activity levels actually worked. The CV(RMSE) of the activity level prediction was **32.2%**. However, and since M&V is not a purely prediction task, there is no reason to use calendar features as proxy of activity in the first place; we can use the estimated activity levels directly. A model that predicts energy consumption given the actual values of the 3-level activity and the outdoor temperature produces predictions with a **CV(RMSE) of 22.5%**. The improvement in accuracy is driven by the fact that we have removed the error from mapping calendar features to activity levels.

The plot of Fig. 3.10 offers some guidance on the interplay between the granularity of the activity feature and the prediction results. As the number of levels increases, the mapping between calendar features and activity levels improves until about a number of ten (10) discrete levels. Both the 2-stage model (that uses the predicted activity levels) and the direct model that uses the actual activity levels improve until ten (10) levels.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

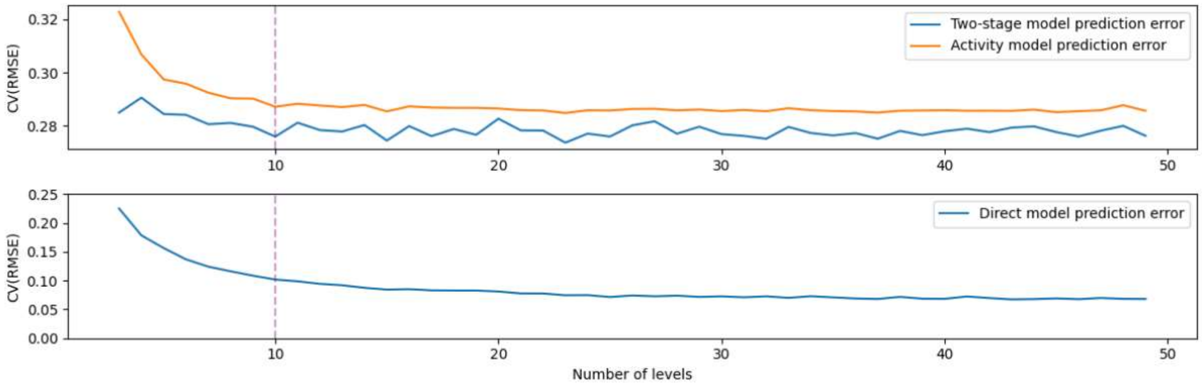


Figure 3.10: **The interplay between the granularity of the activity feature and the prediction results**

At ten (10) levels, we have the best 2-stage model since we have minimized the error due to mapping calendar features to activity levels. If we skip the mapping step and use the activity levels directly as input feature, we get the best direct model with predictions with a CV(RMSE) of around 10%.

The SmartSPIN M&V approach smoothes the raw activity based on the idea that similar energy consumption profiles indicate similar activity levels. The following plot shows the electricity consumption and the **smoothed activity** for the La Gavia pilot. Using a regression model to predict energy consumption given outdoor temperature and the ***smoothed*** activity leads to a **CV(RMSE) of 10%**.

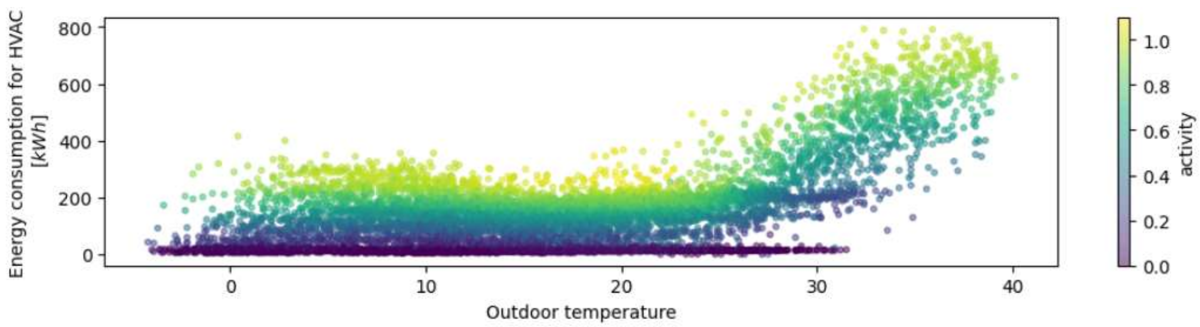


Figure 3.11: **The smoothed activity for the HVAC electricity consumption of La Gavia in 2023**

3.3 The Possible Physical Meanings for the Activity Feature

In order to establish a reference for comparison, the electricity consumption for HVAC of the La Gavia building is decomposed into the contributions of individual components. The plot in Fig. 3.11 shows the contribution – relative to the average prediction – of the daily seasonality (upper panel) and the outdoor temperature (bottom panel) to the building's electricity consumption for HVAC:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

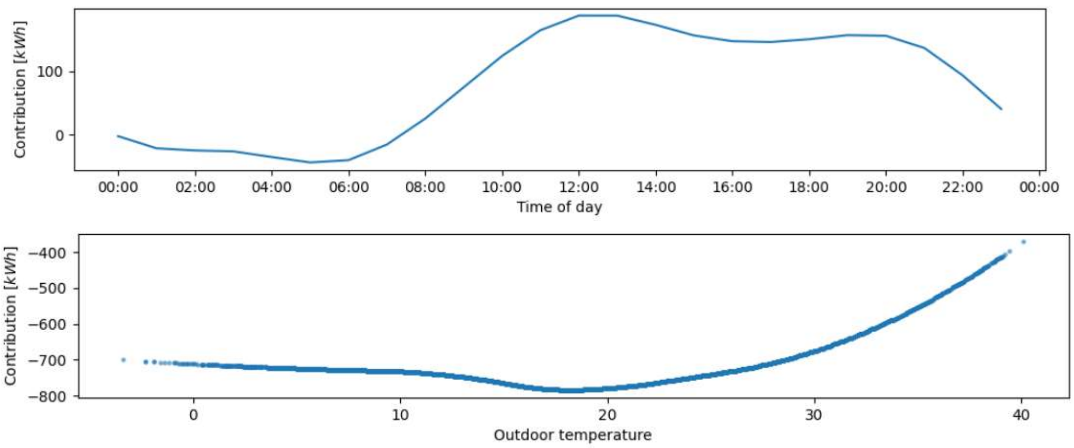


Figure 3.12: Decomposition of the electricity consumption for HVAC of La Gavia in 2023

Next, we explore the impact on our understanding of the building from replacing the calendar features with an estimation of the activity levels. In this case, the HVAC electricity consumption of the building is decomposed into the contributions of the activity and the outdoor temperature. The plot of Fig. 3.13 shows the average daily profile of the activity's contribution (upper panel) and the estimated contribution of the outdoor temperature (upper panel).

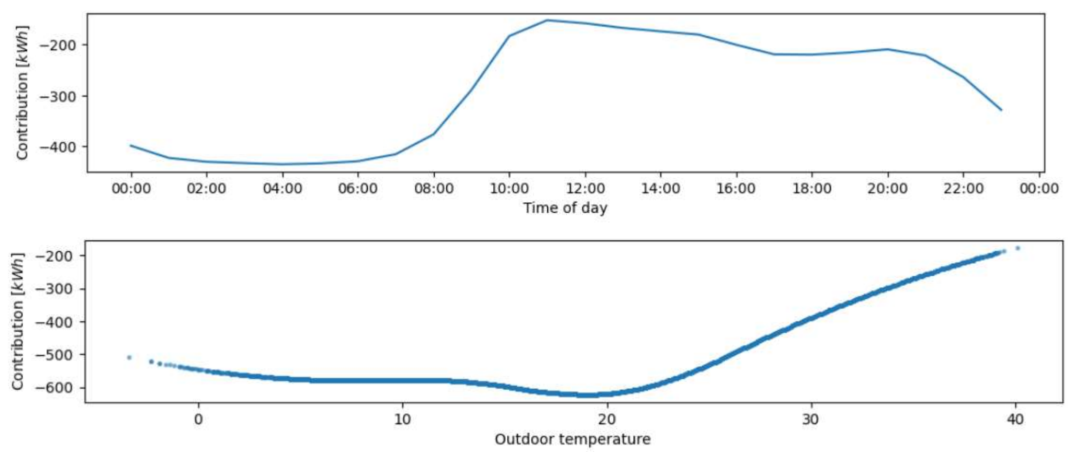


Figure 3.13: Decomposition of the electricity consumption for HVAC of La Gavia in 2023 when activity is used

The plot indicates that the activity feature can be considered as a proxy for the daily operation schedule of the building. Furthermore, the estimated contribution of the outdoor temperature is similar to the one estimated when calendar features were utilized. This implies that the effects of activity and outdoor temperature are unrelated or, in other words, activity does not contain information that is part of the building's external conditions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

The La Gavia pilot building has sensors installed to monitor the number of occupants in the building. This information provides the opportunity to explore the potential physical meaning of the activity feature. The plot of Fig. 3.14 shows the contribution of the number of occupants to the estimated activity for outdoor temperatures that are higher than 22 °C (to capture the cooling operation of the HVAC system). The plot indicates that activity is strongly correlated to the number of occupants: the number of occupants has a large impact on activity levels up to a point, and then the relationship is slightly linear. The trend line from 3,000 occupants and upwards could be representing the effect of internal heat gains of the energy consumption for cooling.

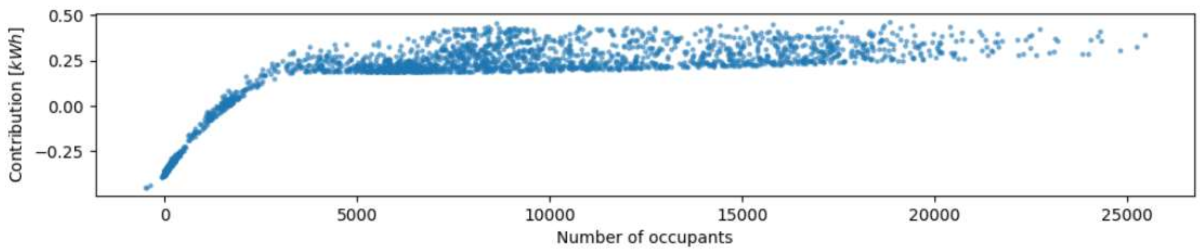


Figure 3.14: Contribution of the number of occupants to the estimated activity

3.4 The Impact of Sample Size

After an event that affects impact variables, the SmartSPIN M&V methodology learns to estimate the new activity levels using the observed, post-event energy consumption data. In this case, we need to know how much data we must have before starting to estimate the activity levels of a building. For this, the following question must be answered: *How much is the in-sample accuracy of the activity estimation affected by the amount of available data*? The following approach was used to answer this question:

- Randomly choose **500 meters** from the BDG2 dataset;
- For each meter, estimate activity levels for a randomly selected week and compare the consistency of the estimated activity with estimates carried out using: (a) one (1) month of data, (b) two (2) months of data, (c) three (3) months of data, (d) six (6) months of data, and (e) a whole year's data after the beginning of the selected week.

This testing scheme is summarized in the following diagram:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

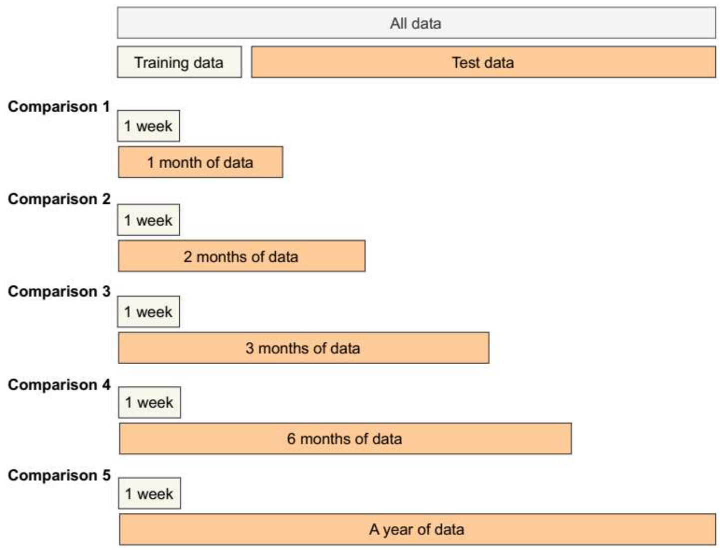


Figure 3.15: The training/testing scheme for evaluating the in-sample accuracy of the activity estimation

In addition, the estimated activity was compared with an estimation given only one (1) month of heating and one (1) month of cooling data.

The results (Fig. 3.16) support a strategy for the activity estimation where:

- Activity estimation can start as soon as one (1) month of data has become available;
- Activity estimations get retroactively updated when one (1) month of heating and one (1) month of cooling data has become available.

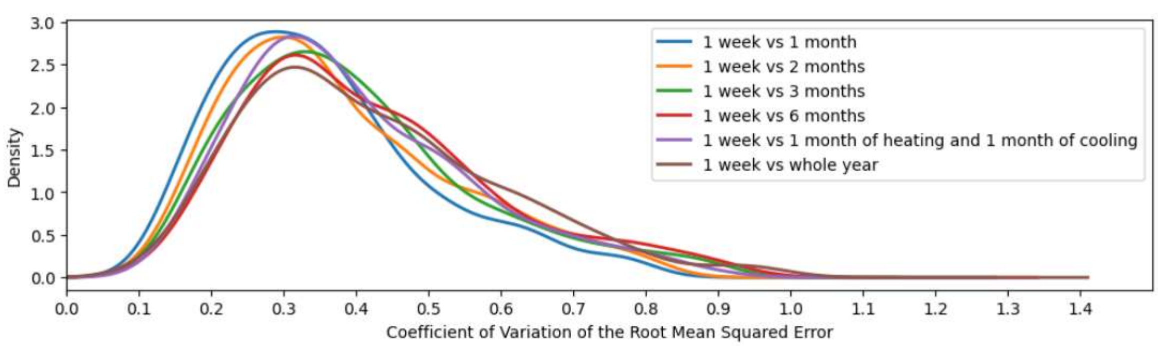


Figure 3.16: Impact of sample size on the in-sample accuracy of the activity estimation

The results provide an answer to an additional question: *how much data do we need to have from the baseline period so that we can trust the counterfactual predictions*? The answer to this question is related to the notion of *data adequacy*. When prediction-based models are utilized for estimating the counterfactual energy consumption, a baseline of a whole year's data is required. However, since mapping-based models estimate activity levels given the post-retrofit energy consumption,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

they **do not need** data for all the months of a year; a month of heating and a month of cooling can provide enough information.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

4 EVENT DETECTION AND CHARACTERIZATION

4.1 INTRODUCTION

Every new batch of data that the M&V workflow receives is scanned for non-routine events. Events are detected by monitoring the difference (RMSE) between the expected and the observed activity levels. It is up to the user to characterize an event, i.e. to decide if the change in the activity levels signifies a mapping event or if activity levels must be recalculated to match the new energy consumption (an impact event).

The event detection step of the SmartSPIN M&V approach uses the tabular CUSUM method for statistical process control. The tabular CUSUM uses two statistics, one (c^+) that accumulates deviations that are above the expected long-term average of the observations μ and one (c^-) that accumulates deviations that are below μ:

$$c_t^+ = \max\left(0,\ x_j - (\mu + SL) + c_{t-1}^+\right)$$

$$c_t^- = \max\left(0, (\mu - SL) - X_j + c_{t-1}^-\right)$$

where x_j are all the observations up to and including time t, and SL is the slack value.

The slack value is commonly set to the middle between μ and an observed value that should signal a shift in the data generating process. If we assume that a shift should be identified for values larger than three (3) standard deviations σ, we get:

$$SL = 1.5 \cdot \sigma$$

Finally, the method requires a decision threshold TH that will signify a shift if the CUSUM metric exceeds it. In statistical process control, TH is typically set to five times the standard deviation of the monitored process.

4.2 APPLICATION OF EVENT DETECTION

The available data for the La Gavia pilot includes one major event: a lockdown due to the COVID-19 pandemic, imposed on 14 March 2020. This is a mapping event, since it affected activity levels (changed the operation schedule and number of occupants) without changing the building's energy efficiency.

In addition, we know that at some point in early 2022, solar PV self-consumption was introduced. This is reflected in the change in the average daily profiles of the energy consumption before and after 2022 (Fig. 4.1). Although M&V of solar PV generation is not needed in practice (generation can be metered directly), if the self-consumption component is ignored, the energy consumption data resembles an (outdoor temperature-independent) efficiency upgrade.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

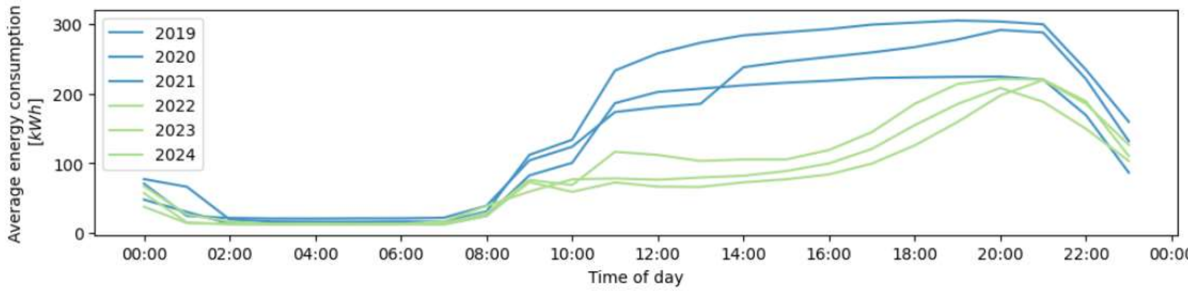


Figure 4.1: Average daily energy consumption profiles per year for the La Gavia building

The plot of Fig. 4.2 shows the difference between observed and expected activity levels as the event detection algorithm processes batches of weekly data for the La Gavia building. In addition, the lockdown event is presented in the plot as a red vertical line.

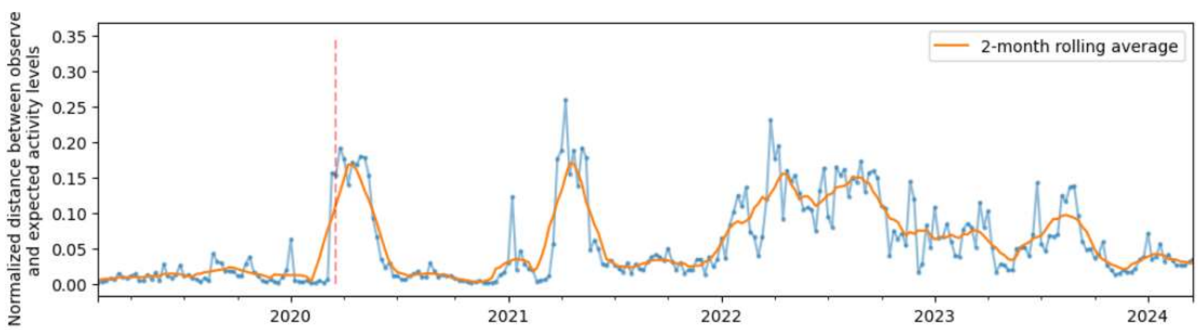


Figure 4.2: Differences between expected and observed activity levels for the La Gavia building

It is easy to spot that the beginning of the lockdown leads to a large upward trend of the monitored difference. There is another event that seems to have taken place in the first half of 2021, but no relevant information was available at the time of the deliverable's writing. Furthermore, the plot shows that the introduction of solar PV generation leads to a sustained upward trend of the monitored difference. Since the algorithm does not look further than one (1) year in the past, the differences in 2023 and 2024 return to their mean value even if solar generation is still available.

The plot of Fig. 4.3 shows the energy consumption of the building and the events identified (as purple vertical lines). In addition, the lockdown event is presented in the plot as a red vertical line.

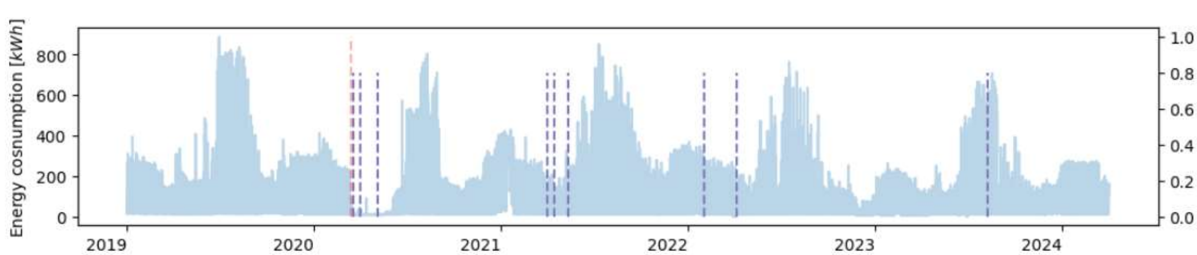


Figure 4.3: Detected events for the La Gavia pilot

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

4.3 EVENT CHARACTERIZATION

There are two (2) types of events that an M&V model must be able to deal with:

1. **Events that affect impact variables**. An energy retrofit is the most characteristic example of such an event, as well as when changes in operating settings take place that affect the energy consumption of the building for the same activity levels and weather conditions. Impact events require re-training both the activity prediction model (to match new energy consumption to new activity levels) and the energy consumption prediction model (to match activity and weather to energy consumption).

2. **Events that affect mapping variables**. Changes in occupancy affect the activity levels but not the relationship between activity and energy consumption. Mapping events do not require re-training any of the models.

Event characterization relies on maintaining a sample of the activity levels for the *most relevant* month (reference sample). The most relevant month is defined as the same month from the previous year of the reporting period or, if not available, the same month from the baseline period (assuming that the energy retrofit did not change the activity profiles) or, if not available, the previous month.

When the event detection procedure reaches the lockdown period, the energy consumption of the new data batch will look significantly different than the expected. This difference will trigger event detection, and the user is responsible for deciding whether the event represents a change in the activity levels or a change in the energy consumption without affecting activity. To support this decision, the user will examine two plots:

- A plot that compares the average observed activity levels with the average activity levels of the reference sample (Fig. 4.4);

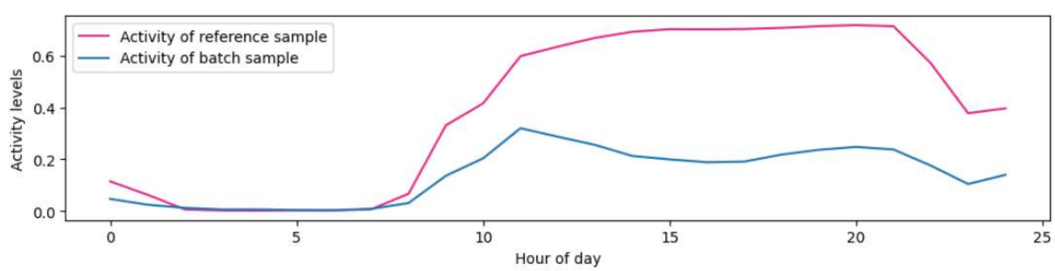


Figure 4.4: Average activity profiles for reference and test samples

- A plot that compares the observed energy consumption with the energy consumption that should be expected according to the average activity levels of the reference sample (Fig. 4.5).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

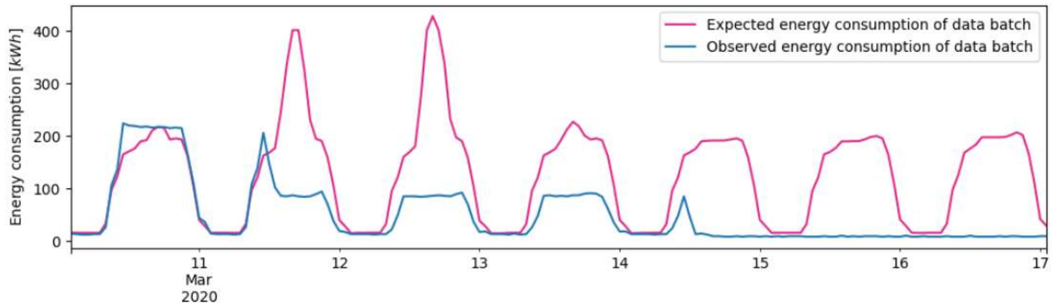


Figure 4.5: **Observed and expected energy consumption for the test sample**

Similarly, when the first event of 2022 will be triggered, the user will be provided with the plots of Fig. 4.6, and tasked to decide[6] if the event should be attributed to a change in the building's energy needs (due to changes in its operation schedule/conditions) or to a change in the building's efficiency to serve its energy needs.

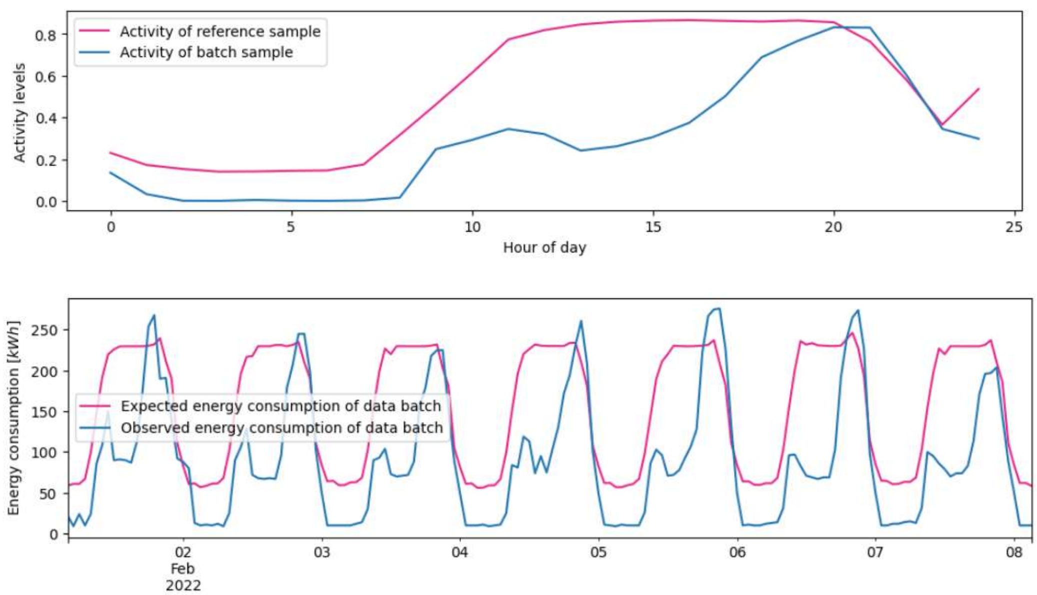


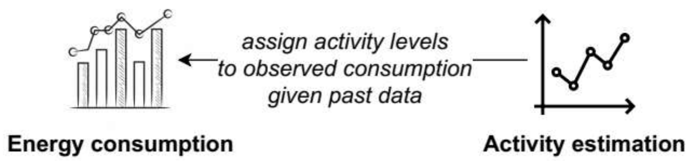
Figure 4.6: **Supporting plots for characterizing the detected event**

[6] Using feedback provided by the building's managers and/or occupants

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

5 METHODOLOGY DEMONSTRATION

Two events have been already identified in the previous chapter for the La Gavia building. The first event is a mapping event that affects only the activity/occupancy levels of the building: lockdown from 2020-03-14 to 2020-05-27. When changes in the building's occupancy occur that affect the activity levels but not the relationship between activity and energy consumption, activity can be estimated from previous data:



The diagram of Fig. 5.1 summarizes the way the M&V workflow adapts to such events:

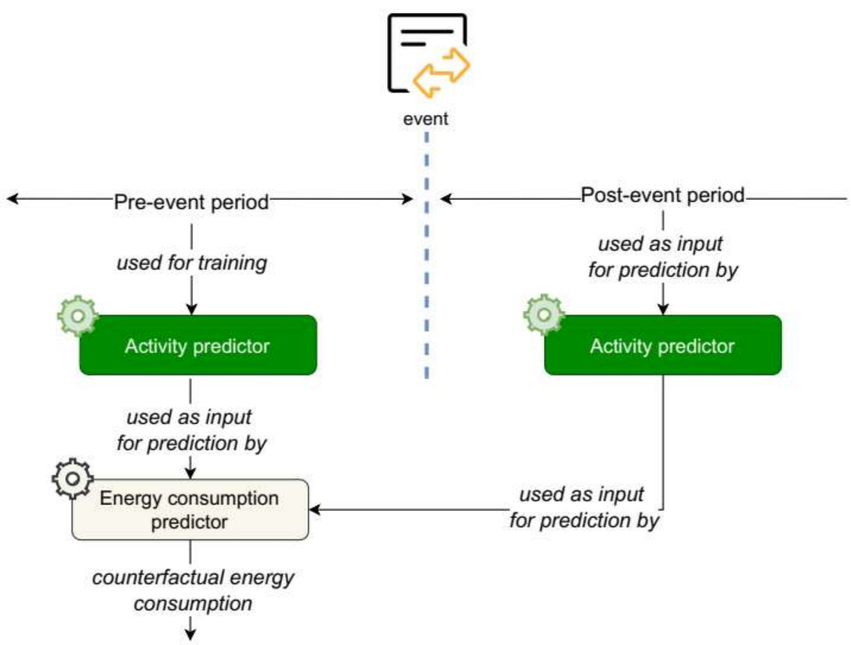


Figure 5.1: Activity estimation after an event

The plot of Fig. 5.2 shows the activity levels as estimated by an activity predictor model that was trained using the data for the energy consumption of 2019, and applied on the observed energy consumption during the lockdown period from 2020-03-14 to 2020-05-27:

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 101033744.

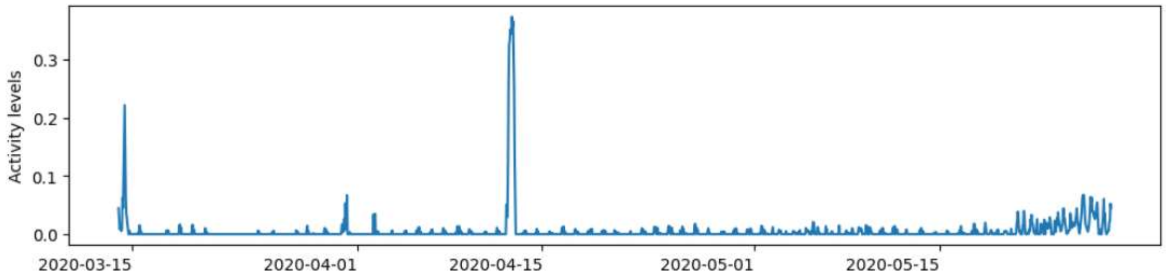


Figure 5.2: Activity estimation for the 2020 lockdown period

Since the activity levels have been estimated using the observed energy consumption data, they reflect the actual usage of the building. The counterfactual energy consumption is estimated by providing these activity levels to the energy consumption predictor that was trained using the data of 2019. The counterfactual energy consumption is presented in the following plot:

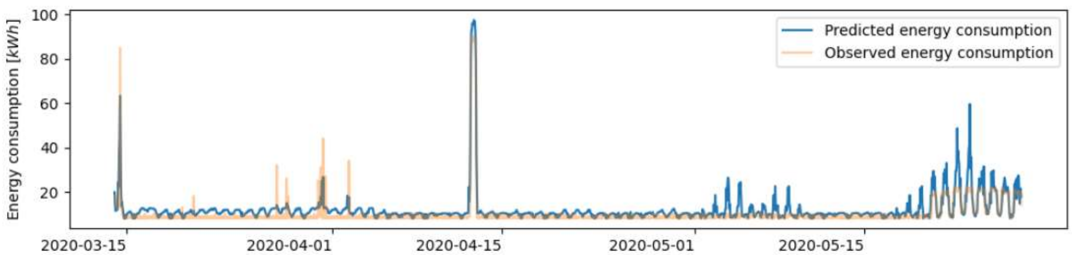
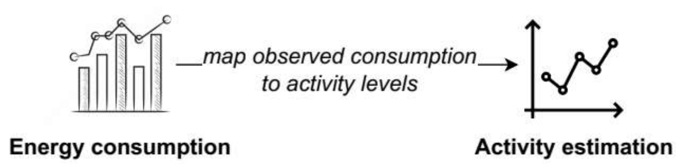


Figure 5.3: Counterfactual energy consumption estimation for the 2020 lockdown period

In addition, solar PV self-consumption was introduced in the beginning of 2022. Although on-site energy generation can be metered directly, this section will treat the generated energy as energy efficiency savings for demonstration purposes.

After an event that affects energy efficiency, the SmartSPIN M&V methodology learns to estimate the new activity levels using the observed, post-event energy consumption data:



The diagram of Fig. 5.4 summarizes the way the M&V workflow adapts to these cases:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

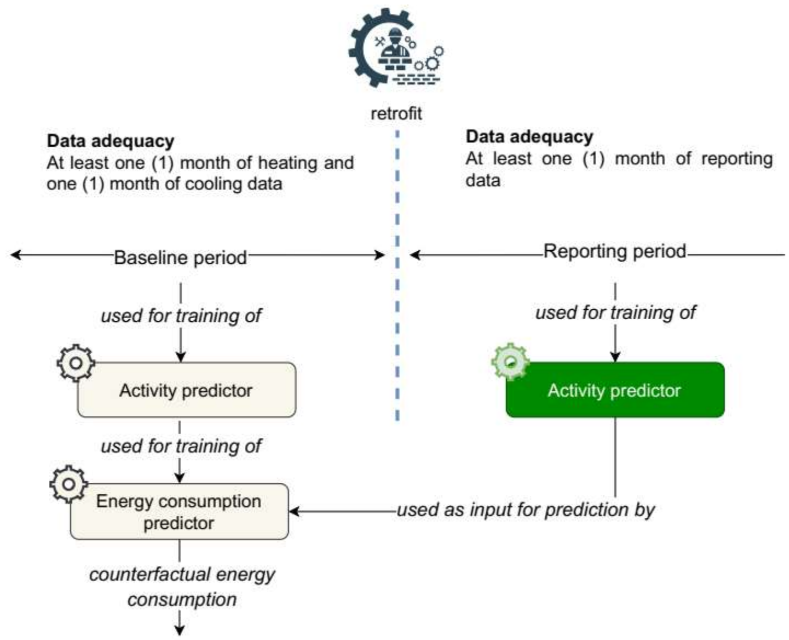


Figure 5.4: Activity estimation after a retrofit

The activity estimation is based on the assumption that outdoor temperature is the main driver for the (explainable part of) the energy consumption variation under the same activity levels. This is true when the energy consumption corresponds to HVAC loads, but not when solar PV generation is the source of energy savings. Consequently, both outdoor temperature and solar irradiance will be used for estimating the activity levels.

Following the flow of Fig. 5.4, activity levels are estimated separately for 2021 and for 2022. Next, an energy prediction model is trained using the pre-retrofit (2021) activity levels and weather conditions. Finally, the counterfactual energy consumption is estimated using the 2022 activity levels and weather as input to the energy prediction model that was trained using the 2021 data. The plot of Fig. 5.5 shows the observed and counterfactual energy consumption for the month of June 2022.

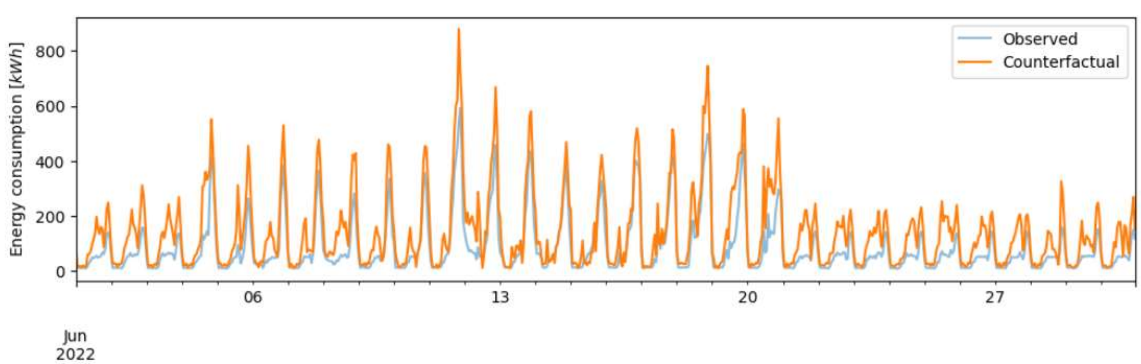


Figure 5.5: Observed and counterfactual energy consumption for the month of June 2022 of the La Gavia pilot

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

Typically, energy savings are presented as cumulative sums over the reporting period. The plot in Fig. 5.6 compares the cumulative sum of the estimated savings (counterfactual minus observed energy consumption) with the available data for the solar PV generation that is consumed by the HVAC systems. Since the data collection started on September 2022, both cumulative sum curves start on September 2022 as well.

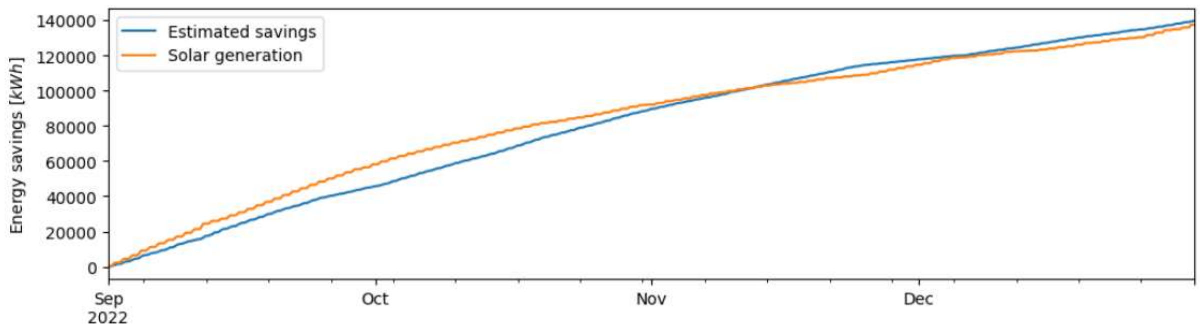


Figure 5.6: Comparison of estimated energy savings and metered generation from onsite solar PV

In order to have a reference for comparison, the savings were also estimated using a prediction-based model. A predictive model was trained using weather data for 2021, as well as calendar features (week of year, day of week, and hour of day). The model was then used to predict energy consumption given the data from 2022. The plot in Fig. 5.7 compares the cumulative sum of the estimated savings with the available data for the solar PV generation.

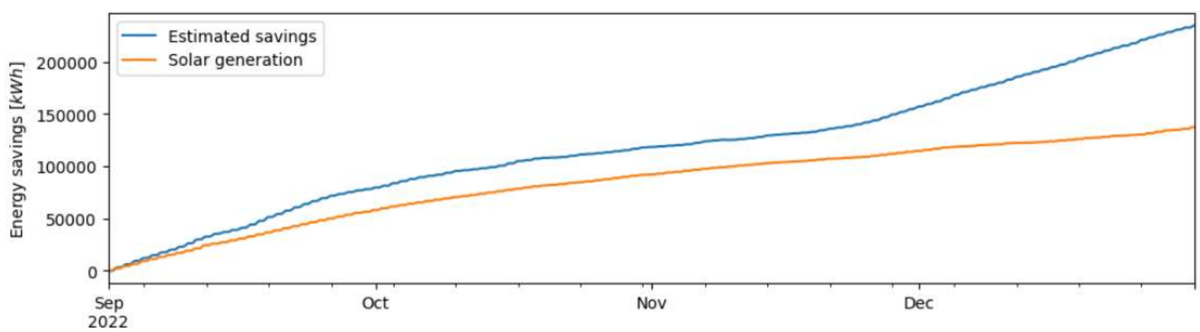


Figure 5.7: Estimated energy savings using a prediction-based model

The results show that the SmartSPIN M&V method is able to estimate the solar generation quite well, even though the methodology is designed to capture the impact of upgrades mainly in HVAC systems (due to the weather-driven approach for activity estimation).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

6 CONCLUSIONS

This deliverable presented an evaluation of the M&V methodology that was developed by the SmartSPIN project. The methodology represents an innovative approach for the M&V of energy savings produced by energy retrofits of buildings, and it constitutes a significant step forward for the state-of-play in M&V compared to both the current literature and the professional practice. The innovation of the methodology is achieved by two (2) main contributions:

- A complete workflow that covers all the steps of the M&V task as it is carried out by M&V professionals;
- A new characterization of the events that may take place inside a building and change the dynamics of its energy consumption, as well as a method to adapt to these events so that to accurately estimate energy savings from energy efficiency upgrades in a building's envelope and HVAC systems. In this way, event detection and adaptation become an integral part of the M&V task.

The immediate next step is the commercialization of the methodology so that to further support the goal of treating energy efficiency as a resource that can be reliably measured and, as a result, a resource that can be financed and exchanged just like energy generation is.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.

APPENDIX

Predictive Model Evaluation

The ASHRAE Guideline 14 proposes the use of the following quantitative metrics to evaluate the fitness of a predictive model:

- The coefficient of variation of the root mean squared error ($CV(RMSE)$);
- The normalized mean bias error ($NMBE$).

The $CV(RMSE)$ metric provides a quantification of the typical size of the error relative to the mean of the observations. The root mean squared error (RMSE) is calculated as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where:

n The number of observations

y_i The consumption of the i^{th} observation ($i = 1,2,\dots,n$)

$\hat{y}_i$ The estimation for the i^{th} observation's consumption.

The $CV(RMSE)$ metric is calculated as:

$$\mathrm{CV(RMSE)} = \frac{1}{\bar{y}} \times \mathrm{RMSE} \times 100(\%)$$

where:

$\bar{y}$ The mean value of the observed consumption data.

The $NMBE$ metric is computed as:

$$NMBE = \frac{1}{\bar{y}} \times \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \times 100(\%)$$

The ASHRAE Guideline 14 requires that $|NMBE| \leq 0.5\%$.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101033744.