White Paper

The Meval Tool for Estimating and Verifying Energy Savings from HVAC System Upgrades

Sotiris Papadelis, HEBES Intelligence

This White Paper is available at no cost from www.hebes.io/white-papers

March 16, 2026

Disclaimer

While reasonable efforts have been made to ensure the accuracy and reliability of the information presented, HEBES Intelligence makes no representations or warranties, express or implied, regarding the completeness, accuracy, or suitability of the content for any particular purpose. The information is provided "as is," and HEBES Intelligence shall not be liable for any losses or damages arising from its use.

By accessing or using this white paper, you agree to respect all intellectual property rights of HEBES Intelligence and to use the information solely for lawful and non-commercial purposes unless explicitly authorized otherwise.

List of Acronyms

BMS	Building Management System
DiD	Difference-in-Differences
IPMVP	International Performance Measurement and Verification Protocol
M&V	Measurement and Verification
TOWT	Time of week and Temperature

Executive Summary

Energy efficiency has long been recognized as a critical pillar of the energy transition, yet its potential remains underutilized as a tradable and financeable resource. A central barrier lies in the inherent fragility of energy savings estimates, which are highly sensitive to changes in building usage patterns, occupancy behavior, and HVAC system operation. These uncertainties complicate Measurement and Verification (M&V), limiting confidence in reported savings and hindering the development of scalable business models and financing mechanisms for energy efficiency upgrades.

Traditional M&V approaches predominantly rely on predictive models that estimate baseline energy consumption and compare it with post-retrofit performance. While effective under stable conditions, these methods struggle to maintain accuracy when non-routine events or operational changes occur. As buildings are dynamic systems, such changes are the norm rather than the exception, causing M&V to deviate from a pure prediction problem and exposing fundamental limitations in conventional modeling techniques.

This white paper introduces a novel methodology developed by HEBES Intelligence and implemented in the Meval tool. The approach reframes M&V as a problem of mapping comparable operational states between pre- and post-retrofit periods, rather than solely predicting a counterfactual baseline. By identifying and categorizing the evolving states and conditions within a building, the methodology enables continuous adaptation to changes such as activity shifts, operational adjustments, and other non-routine events.

At the core of the Meval method is a dynamic modeling framework that detects relevant system states across time and establishes correspondences between them. Energy savings are then quantified as differences in consumption between matched states, ensuring that comparisons remain valid even as building behavior evolves. This state-based approach enhances robustness, reduces bias, and improves the reliability of savings estimates in real-world conditions.

The paper further examines the limitations of prediction-based M&V, explores causal-informed redesign options, and positions Meval within the broader landscape of M&V methodologies. By addressing key sources of uncertainty and enabling more trustworthy quantification of energy savings, the proposed approach represents a significant step toward making energy efficiency a credible, transactable resource.

Ultimately, Meval provides both a practical tool and a conceptual advancement for M&V, supporting more resilient business models and unlocking new opportunities for financing energy efficiency at scale.

Table of Contents

Introduction

Energy savings estimation is fragile with respect to changes in the way a building and its HVAC systems are used. This is the main reason why energy efficiency is yet to be treated as a transactable resource. This affects business model development — many business models for energy generation assets would struggle to apply to efficiency upgrades — and the financing of the upgrades.

In HEBES Intelligence, we have been working on innovative approaches for the Measurement and Verification (M&V) of energy savings produced by energy retrofits of buildings. The result of our work is a methodology that constitutes a significant step forward for the state-of-play in M&V.

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.

Unlike other approaches, the methodology does not rely exclusively on pure predictive models to estimate energy savings. Instead, it treats M&V as a task of devising and applying a mapping from states and conditions after an energy efficiency intervention to states and conditions before it. In this way, the impact of the intervention is the difference in energy consumption between **matching states and conditions**. The identification of the relevant states is carried out in both pre- and post-retrofit periods, so that the underlying model continuously adapts to changes inside a building.

The methodology is implemented as a web-based tool called **Meval**. **Meval** is freely accessible at https://mevalapp.com/ and it will remain free for the foreseeable future as a platform for collaboration and testing of M&V methods that are robust to non-routine events.

This white paper explains the methodology that **Meval** implements and the rationale behind it.

M&V as a Prediction Task

Predictive models for M&V

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.
- **Occupancy-independent information** that is predictive of the building's energy consumption prior to the intervention. The most often utilized information is the outdoor air temperature.

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, 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.

The Time-of-Week and Temperature (TOWT) predictive model is a regression-based modeling approach widely used to estimate counterfactual energy consumption. The TOWT model decomposes energy consumption into two components:

- One that captures the time-dependent operational effects, represented by discrete time-of-week indicators, and
- One for the weather-dependent effects, represented by temperature response functions.

Time-of-week effects are modeled using a set of categorical indicator variables that partition the week into fixed intervals (usually hours). These indicators capture recurring operational schedules that may reflect time-varying occupancy, equipment operation, and control strategies, all of which may not be directly observable.

Weather sensitivity is incorporated through temperature-based regressors that are non-linear or piecewise linear functions of the outdoor air temperature.

In its most general form, the TOWT model can be written as:

$$Y_t = a_{h(t)} + f_s(T_t) + \varepsilon_t$$

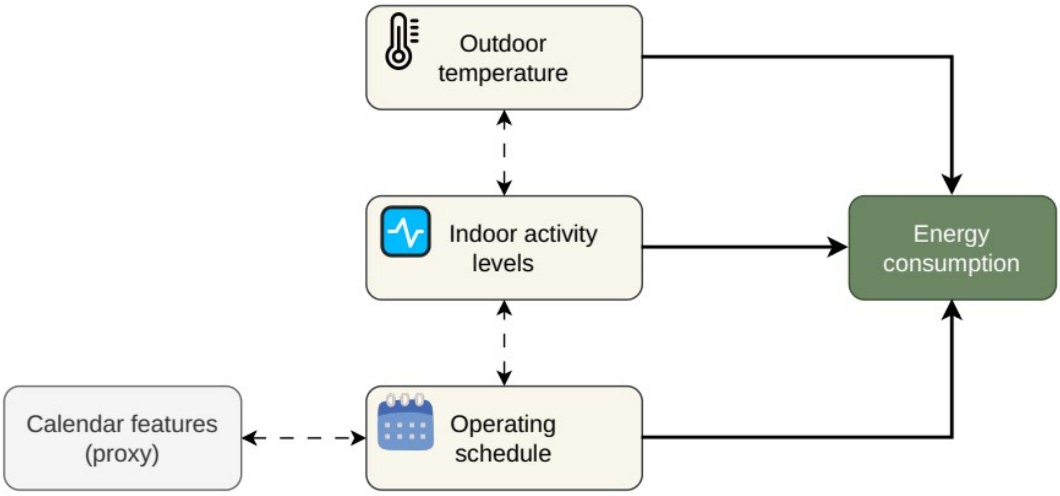
where:

- Y_t is the metered energy consumption of the building at time t
- $a_{h(t)}$ is the coefficient associated with the time-of-week hour h
- $f_s(T_t)$ is a temperature response function of outdoor air temperature that may vary based on the building's schedule ($s = 1$ for setpoint operation and $s = 0$ for setback)
- ε_t is a stochastic error term.

Since M&V is an impact assessment task, it is useful to provide a causal interpretation to any approach to savings quantification. For the TOWT model, and any other predictive model for that matter, we have:

True drivers of energy	- Outdoor temperature - Operating schedule (hidden) - Indoor activity / occupancy (hidden)
Observed variables	- Outdoor temperature - Hour-of-week (proxy for schedule & occupancy patterns)

The causal relationships (solid arrow lines) and the correlations (dashed arrow lines) between these variables are summarized below:



Simplified causal graph for building energy consumption

Predictive modeling works because calendar features are good proxy variables that capture regular, repeating causal patterns, such as weekdays and weekends, work and non-work hours, and typical occupancy cycles. In a more formal way, there is (often) a strong correlation between calendar features and, during the training (baseline) period, the joint distribution of schedule, occupancy and calendar features **is stable**.

When M&V stops being a prediction task

While prediction tasks have favorable properties when it comes to testing and evaluating their accuracy, they typically include the following steps:

- Use historical data to train a predictive model
- Apply the model to get predictions for one of more time steps ahead
- Evaluate the accuracy of the model when receiving new data (by comparing the observed data to the predictions)
- Re-train or fine-tune the model using the newly received data.

This workflow is different from the workflow for M&V:

- Use historical data (baseline period) to train a predictive model
- Apply the model to get predictions over the reporting period
- Assume that the difference between observed energy consumption during the reporting period and predicted consumption is attributed to the retrofit.

In the M&V case, the true values for the counterfactual prediction are never observed. As a result, it is much more difficult to update the predictive model so that it remains relevant for its task.

Let's consider two types of changes that may take place in a building after the retrofit.

Case A: Change in operating schedule

If a building changes its operating schedule (let's say from 9–5 to 24/7), the estimated energy savings **should increase** since a more efficient system is now used more often. However, from a predictive point of view, energy consumption is now higher than expected, so energy savings have eroded. This dichotomy stems from the fact that:

The predictive model assumes

> The learned mapping between calendar features and the operating schedule remains valid, or, equivalently,
>
> $$P(schedule \mid hour) = invariant$$

But in reality

The post-retrofit mapping is different than the learned one,

$$P(schedule \mid hour) \neq baseline$$

This is a **distribution shift** of one of the drivers of energy consumption that must be accounted for as a non-routine adjustment.

Case B: Change in indoor activity levels

In this case, the operating schedule remains the same but the number of occupants has changed significantly (and so have the internal heat gains) or plug loads have changed (for option C of the IPMVP). Again, a main driver of energy consumption has changed and, as a result, masks the actual energy savings:

The predictive model assumes

The joint distribution between the hidden occupancy / activity levels (approximated through calendar features) and the outdoor temperature remains stable, or, equivalently,

$$P(occupancy, temperature) = invariant$$

But in reality

The post-retrofit joint distribution has shifted,

$$P(occupancy, temperature) \neq baseline$$

In a nutshell:

> Prediction works because “Tuesday at 10am” used to mean something. But it fails when “Tuesday at 10am” stops meaning the same thing.

Causal-based redesign options for improving prediction models

The most obvious way to improve a predictive model’s capability to handle distribution shifts in the unobserved variables is to observe them. This requires the presence of a **sensor infrastructure** that can detect occupancy levels (sensors on door frames, WiFi counts, and CO_2 sensors) and/or extract operation schedules from BMS systems. At the same time, such instrumentation requirements would place a lot of buildings out of M&V’s

scope. Any viable M&V approach must be able to make good use of additional data coming from installed sensors, but it **should not require** these data so as to work reliably.

Another approach is to make use of the Difference-in-Differences (DiD) method. In this case, a control group is used so as to absorb the unobserved changes. This is similar to clinical trials: the treated buildings are the building where a retrofit took place, and the control ones are buildings without retrofits. As long as schedule and occupancy changes affect both groups similarly, their effects cancel out when savings are estimated as:

$$savings = \left(y_{treated}^{post} - y_{treated}^{pre}\right) - \left(y_{control}^{post} - y_{control}^{pre}\right)$$

The limitation of this method is the need to find an appropriate control group, as well as the fact that it works best for market-level events rather than events that affect only specific buildings.

The final approach is the one **Meval** utilizes. It is based on the idea that estimating the unobserved variables from the observed ones can make an M&V model more capable of both detecting and adapting to changes that would otherwise break a counterfactual prediction task. To this end, **Meval** operates in the model input domain and focuses on devising the most appropriate inputs for an M&V model.

In particular, **Meval** makes a distinction between **mapping variables** and **impact variables**:

- Mapping variables are defined as the observed and 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 matching 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 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.

Contrary to the prediction-based approach, the **Meval** 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.

The Meval Method for M&V

The modeling approach

It is conceptually straightforward to imagine that calendar features approximate a hidden variable that reflects variations in the activity levels inside a building (operating schedule, number of occupants, recurrent activities that consume energy). This hidden variable – let's call it **activity** – can be estimated based on the requirement that conditional on temperature, it explains the variation in the energy consumption that cannot be explained by a temperature sensitivity model.

This is equivalent to requiring that the difference in energy consumption between two observations (one pre- and one post- retrofit) that have the same activity levels should be explainable by the impact of the retrofit on the building's needs for heating and cooling energy.

Meval uses a physics-based model to capture the relationship between energy consumption and outdoor temperature. This model can be simplified as:

$$Y_t = (\beta_0 + \gamma_0 R_t) + (\beta_1 + \gamma_1 R_t) \cdot \Delta T_t + \varepsilon_t$$

where:

- $\beta_0,\ \beta_1$ are the pre-retrofit intercept and slope of the temperature - consumption relationship,
- $\gamma_0,\ \gamma_1$ are the retrofit effects on the intercept and slope,
- ΔT represents the difference between indoor and outdoor temperature.

Accordingly, activity is estimated through the Bayesian inference of the following model:

$$\begin{gathered} Y_t = (\beta_0 + \gamma_0 R) + (\beta_{hea} + \gamma_{hea} R_t) \cdot H_t + (\beta_{coo} + \gamma_{coo} R_t) \cdot C_t + \beta_A A_t + \varepsilon_t \\ H_t = \max(0, T_h - T_t) \\ C_t = \max(0, T_t - T_c) \\ A_t = \alpha_{cal(t)} + \delta_t \end{gathered}$$

where:

- $\beta_{hea},\ \beta_{coo}$ are the pre-retrofit slopes for heating and cooling
- $\gamma_{hea},\ \gamma_{coo}$ are the retrofit effects on the slopes for heating and cooling

- β_A is the effect of the activity
- T_h is the heating balance temperature (below this point, heating is required)
- T_c is the cooling balance temperature (above this point, cooling is required)
- $A_t \in [0,\ 1]$ the hidden activity
- $\alpha_{cal(t)}$ represents calendar effects as random effects
- δ_t is the residual activity.

An important difference from the prediction-based approach is that **Meval** estimates energy savings by building models for both the pre- and the post-retrofit periods.

Baseline period

During the baseline period, **Meval** jointly fits a physics-based model and a model that uses calendar features to map the deviations around the physics-based model to the hidden variable called activity. In this way, the physics model reflects the average energy consumption at a given outdoor temperature, while activity reflects relative deviation from average. The physics-based model is flexible enough to capture thermal inertia, free-cooling operation and capacity saturation effects.

Activity is regularized according to how strong the link between activity and calendar features seems to be. In the case where calendar features hold no predictive capability (a building with an erratic operating schedule), activity collapses to a binary feature (on/off state).

Reporting period

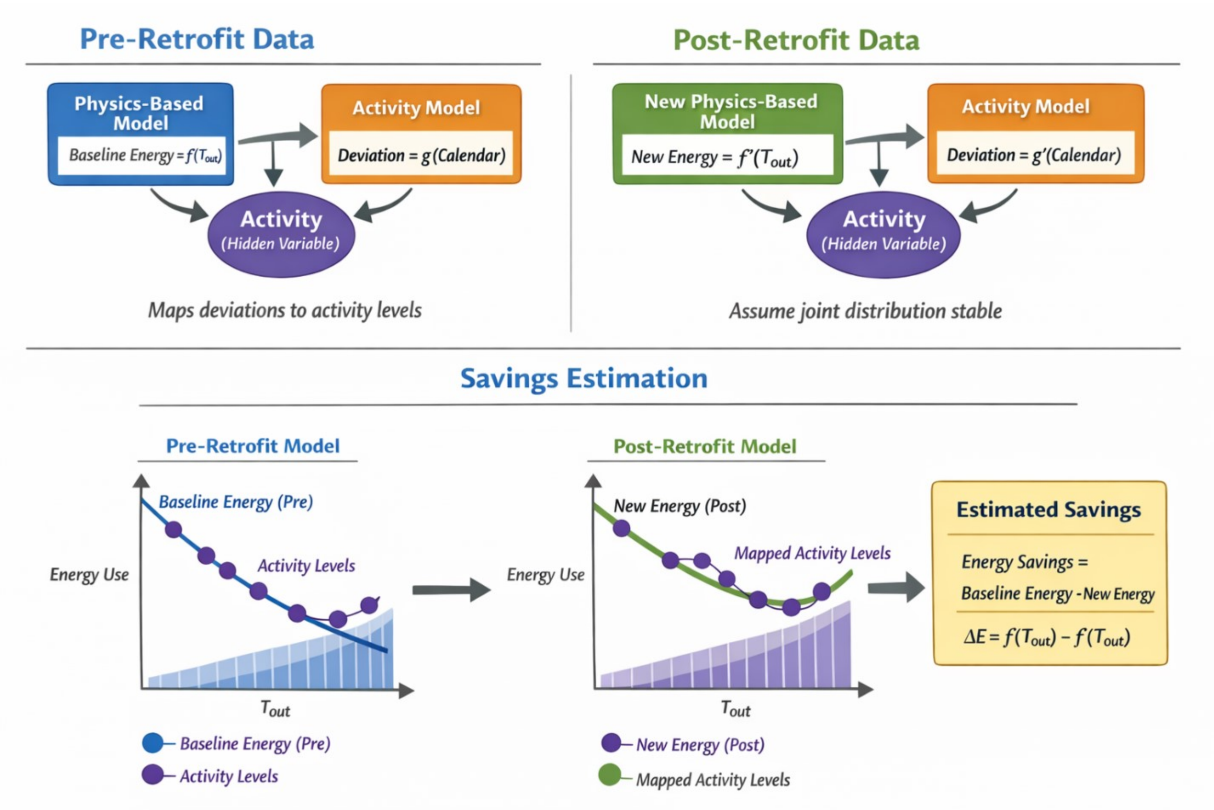
During the reporting period, a new physics-based model is fitted on post-retrofit data. The estimation of the activity follows the same approach with an **additional assumption**: the joint distribution of activity and outdoor temperature is assumed to be stable before and after the retrofit. This assumption is not arbitrary, but reflects the fundamental requirement for data-based M&V whatever the chosen method for estimating the counterfactual energy consumption is:

> The retrofit changed efficiency (temperature response), not occupant behavior or operations.

If this assumption does not hold at least for a period after the retrofit that would allow the quantification of what actually changed in the building's systems, there is an identifiability problem: activity shifts will mask the effect of the upgrade. Here, activity shifts are defined as changes in the way an existing system is used (such as a different number of heating/cooling hours) or changes in the context of the systems' operation (such as a higher or lower number of occupants during peak hours).

Estimation of energy savings

Meval estimates energy savings by comparing the outputs of the pre- and post-retrofit physics-based models augmented by an activity feature that maps the activity levels after the retrofit to similar levels before. In this way, calendar features **never cross** from pre- to post-retrofit periods. Instead, they are utilized only for estimating the hidden activity levels, which are comparable between the pre- and post-retrofit periods:



Dealing with activity shift events

A benefit of the **Meval** model is that given an activity shift event, it can estimate the new activity levels by answering the following question:

> What activity trajectory best explains observed energy under unchanged building physics?

In other words, if the building's efficiency has not changed, any change in energy use under the same outdoor conditions must come from how the building is being used. This change should be reflected in the updated activity levels.

The inverse problem is formulated as:

$$
\begin{gathered}
E_t = \beta_0 + \beta_h H_t + \beta_c C_t \\
H_t = \max(0, T_h - T_t) \\
C_t = \max(0, T_t - T_c) \\
A_t = \exp\left(\alpha_{cal(t)} + u_t\right) \\
y_t \sim \mathcal{N}\left(A_t \cdot E_t, \sigma_y\right)
\end{gathered}
$$

where:

- $\beta_0, \beta_h, \beta_c$ are fixed parameters estimated on pre-event data
- $\alpha_{cal(t)}$ represents calendar effects as random effects
- u_t is a smooth autoregressive process
- σ_y the energy consumption noise

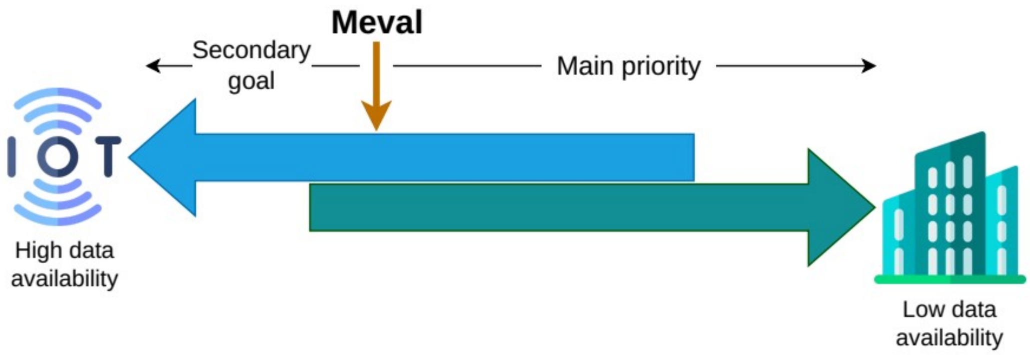
Since $\beta_0, \beta_h, \beta_c$ are fixed parameters, all unexplained systematic changes must go through the unknown activity.

Events that affect both efficiency and activity levels at the same time make identification and correction very difficult. There is an ongoing work on **Meval** that aims at defining the inputs that are required from users and the way these inputs will be used to deal with such combined events.

Positioning of Meval

At its current stage, the **Meval** tool operates on hourly building data, which provides the level of granularity required to reliably identify operational states and detect changes in building behavior. Looking ahead, the tool will become a platform by being integrated directly with IoT sensors and Building Management Systems (BMS), enabling real-time data acquisition and a more detailed representation of building dynamics.

However, a key priority remains the inclusion of buildings where such high-resolution data is not available. The underlying methodology supports this objective by allowing the use of user-defined assumptions in place of missing data, particularly for the baseline period. Then, high-resolution data from the reporting period can be used to verify the extent to which these assumptions align with what is observed.



This flexibility will ensure that robust M&V can be performed even in data-constrained environments, significantly expanding the applicability of the approach beyond digitally mature buildings.