Novel Randomized M&V Examples

2024-05-24

Table of Contents

[1 Objective 1](#_Toc167439582)

[2 M&V use case 1](#_Toc167439583)

[3 Use case measurement dataset 1](#_Toc167439584)

[4 Novel randomized M&V analysis 1](#_Toc167439585)

[4.1 Switchback experimental design 1](#_Toc167439586)

[4.2 Sequential analysis of the intervention effect 1](#_Toc167439587)

[4.3 Scenario A: A discrete non-routine event 1](#_Toc167439588)

[4.4 Scenario B: Continuous baseline change due to HVAC component degradation 1](#_Toc167439589)

# 1 Objective

The purpose of this document is to provide open-source code for analysts who wish to apply the novel measurement and verification (M&V) procedure described in the associated paper to their own projects or datasets. By following the guidance below, users should be able to:

1. Replicate the novel M&V method proposed in the manuscript:
   * Design a randomized and balanced switchback control implementation schedule consisting of a baseline strategy and an intervention strategy,
   * Conduct sequential analysis to infer intervention effect on a target metric (e.g. energy consumption, carbon emissions, or thermal comfort).
2. Demonstrate the reliability of the novel randomized method and compare it with the conventional method under:
   * Scenario A: Discrete non-routine events
   * Scenario B: Continuous baseline change

By replicating the method proposed in the manuscript, we aim to demonstrate that the novel M&V approach can reach a conclusion faster than the conventional method and provide a robust uncertainty quantification. Additionally, by evaluating savings estimations under various common scenarios, we showcase the reliability of this new method in overcoming the limitations of conventional approaches (e.g. the impact of changes in building performance that are wholely unrelated to the intervention). This example code follows the flowchart depicted in the manuscript and provides a minimal example for demonstration purposes. The analyst should make reasonable adjustments to address their specific use case.

# 2 M&V use case

We apply the same use case described in the manuscript where a software-as-a-service control company offers a more efficient chiller control algorithm. Here, the analyst aims to determine the amount of HVAC energy saved by this control intervention. The overall M&V protocol regarding sampling ratio, carryover effect, blocking design and stopping criteria remain the same as the manuscript. To replicate the analysis, we also fitted a time-of-week and temperature model and normalized the savings using a typical meterological year weather file for the same location. We compared the randomized method with the conventional method using the average treatment effect (ATE) by calculating the mean difference between baseline samples and intervention samples to estimate the energy-saved by the control algorithm. For both the TMY-normalized, and actual-weather savings, we also represent these as fractional savings (or percent savings) by normalizing using the mean of the baseline measurements.

# 3 Use case measurement dataset

For this demonstration, we simply create a hypothetical baseline and intervention measurement set by using a linear function including outdoor temperature (), peak and off-peak hour indicator (, ) as independent variables:

We apply that to two years of measured, actual weather data in Chicago, USA. The term accounts for random noise sampled from a normal distribution . The created measurement set for baseline and intervention controls is shown in figure 3.1. If the analyst follows the conventional M&V method described in ASHRAE Guideline 14 (G14) or the International Performance Measurement and Verification Protocol (IPMVP), then only baseline measurements are measurable/observable during the pre-retrofit period and only intervention measurements are measurable during the post-retrofit period. In other words, the control strategy switches between baseline and intervention one time in the entire study period.

| datetime | t\_out | power |
| --- | --- | --- |
| 2021-01-01 00:00:00 | -1.66666666666667 | 44.9914924400949 |
| 2021-01-01 01:00:00 | -1.54000000000000 | 54.0111057579299 |
| 2021-01-01 02:00:00 | -1.41333333333333 | 36.4702105818566 |
| 2021-01-01 03:00:00 | -1.28666666666667 | 52.3555812024666 |
| 2021-01-01 04:00:00 | -1.16000000000000 | 54.1238962059008 |
| 2021-01-01 05:00:00 | -1.03333333333333 | 61.6213199392728 |

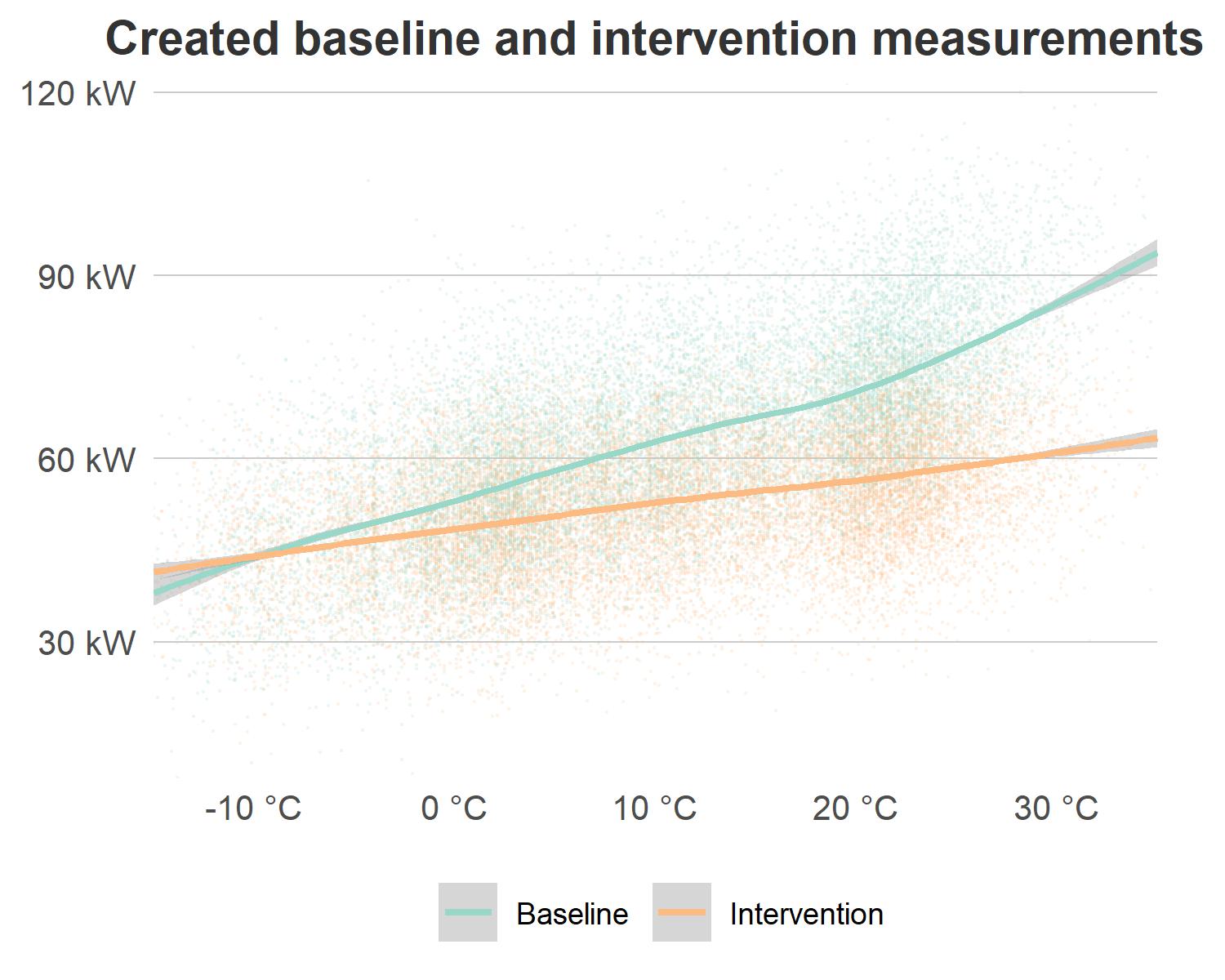


Figure 3.1: Generated hypothetical baseline and intervention measurements across the whole study period using a linear function (data points show hourly energy consumption and line fitted using the loess function).

# 4 Novel randomized M&V analysis

## 4.1 Switchback experimental design

As most control intervention strategies can be implemented interchangeably with the building baseline control, we argue in the paper that randomizing the implementation sequence between intervention and baseline control strategies can balance the impact of confounding variables when estimating the intervention effect, find a result faster, and be more robust to the occurrence of changes in performance that are unrelated to the intervention. A simple way to get started is using a built-in function ’*sample()’* in R package *‘base’* to randomly select at each sampling interval which strategy to implement. The advantage of this function is easy to implement for a short test run (e.g. 1 week) on a single unit. But it is not recommended for more complicated tests, and we will address these limitations below.

In statistics, average treatment effect (ATE) estimation depends on the statistical power of unbiased randomization and balanced sample size. This means to accurately estimate the energy-saving effect of the intervention control strategy, we should design an unbiased randomization that has an equal amount of measurements for each level of the confounding variables. Therefore, a simple randomization can still be biased, for instance, by yielding a poor randomization where the baseline - by chance - is sampled more often on weekends and vice versa. This can be avoided by designing a block randomization, which is a type of experimental design that first divides the testing units into blocks and then randomly assigns which strategy to implement for each unit within each block. Thus in practice, each block type corresponds to one confounding variable, and since strategies are sampled at the block level, this means each confounding variable is covered by all testing strategies. For example, building energy consumption is known to be affected by occupancy, time, and outdoor weather conditions. If the control service company wishes to bill the customer based on the savings solely from the control retrofit itself, the M&V analyst should decompose the measured energy savings and adjust for the influence from differences in weather.

For example in this document, we assume two strategies are being compared over a blocking period of 12 weeks. To do block design, we use a function in R called *‘blocks()’* from the *‘blocksdesign’* package. We further bonded other functions and inputs to make the package more versatile (see code in .rmd file). Once successfully called, the function will return a list consisting of schedule output and schedule summaries. The detailed schedule can be viewed by calling the schedule key name.

| season | date | weekday | strategy |
| --- | --- | --- | --- |
| 1 | 2021-01-01 | Fri | 1 |
| 1 | 2021-01-02 | Sat | 2 |
| 1 | 2021-01-03 | Sun | 1 |
| 1 | 2021-01-04 | Mon | 2 |
| 1 | 2021-01-05 | Tue | 1 |
| 1 | 2021-01-06 | Wed | 1 |
| 1 | 2021-01-07 | Thu | 1 |
| 1 | 2021-01-08 | Fri | 2 |
| 1 | 2021-01-09 | Sat | 1 |
| 1 | 2021-01-10 | Sun | 1 |
| 1 | 2021-01-11 | Mon | 1 |
| 1 | 2021-01-12 | Tue | 1 |
| 1 | 2021-01-13 | Wed | 1 |
| 1 | 2021-01-14 | Thu | 2 |

To check whether the blocking experimental design is balanced, we can call three built-in summary table outputs:

1. Weekday summary: calculates how many days are sampled for each season weekday and strategy combination.
2. Consecutive days summary: calculates how many consecutive days are sampled for each strategy.

| season | weekday | strategy | n |
| --- | --- | --- | --- |
| 1 | 1 | 1 | 6 |
| 1 | 1 | 2 | 6 |
| 1 | 2 | 1 | 6 |
| 1 | 2 | 2 | 6 |
| 1 | 3 | 1 | 6 |
| 1 | 3 | 2 | 6 |

| strategy | consecutive |
| --- | --- |
| 1 | 80 |
| 2 | 81 |

## 4.2 Sequential analysis of the intervention effect

This section replicates the sequential analysis described in the manuscript using the newly created measurement dataset. The results are shown in figure 4.1.

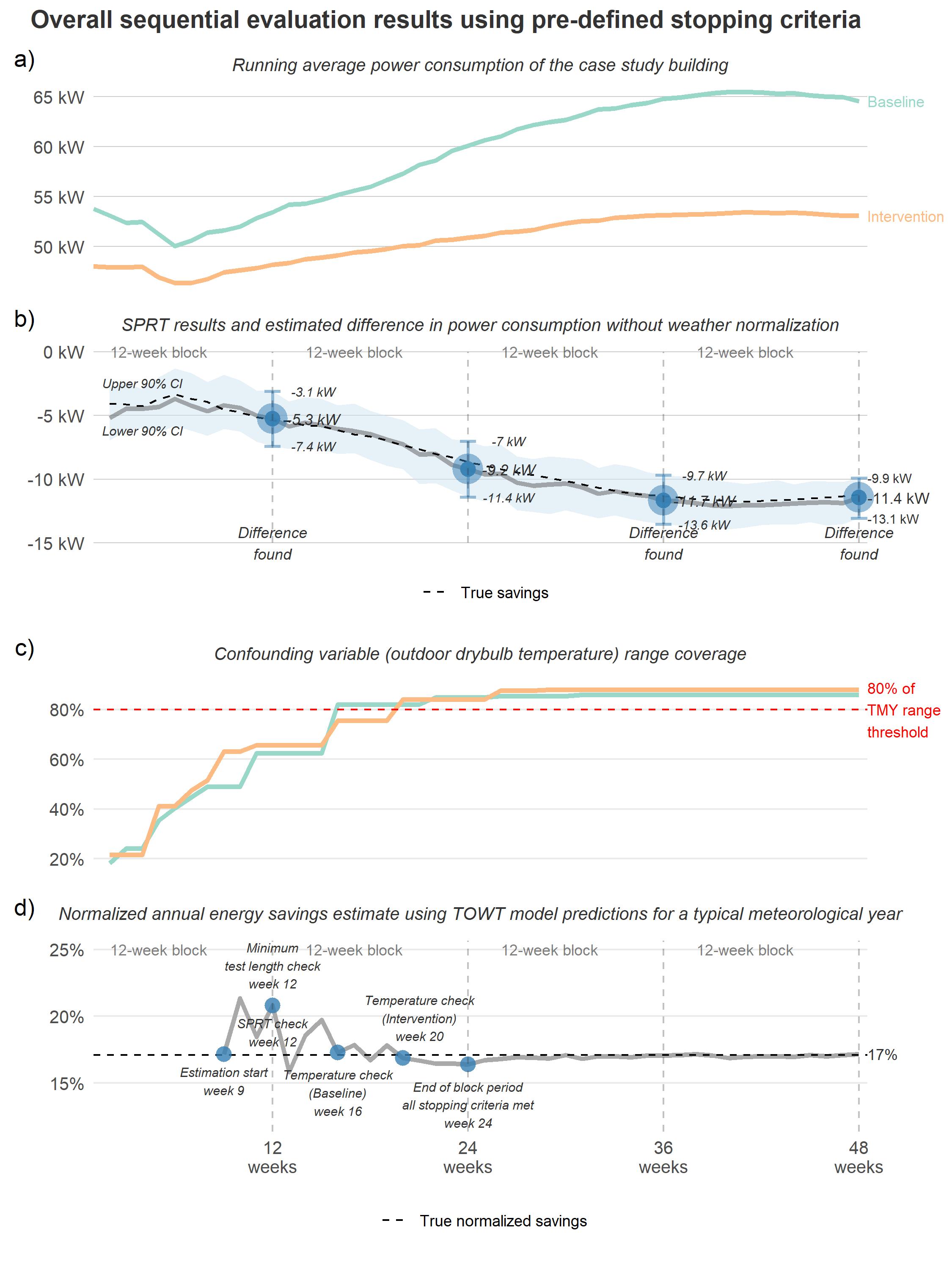


Figure 4.1: Comprehensive results summary of the proposed M&V method applied to the case study throughout a total of 48 weeks

As figure 4.1 indicates, the running average of power consumption in subplot a) shows more savings from the intervention strategy during warmer outdoor weather in general. This also corresponds to the sequential probability ratio test (SPRT) results shown in subplot b). The energy savings as the difference between baseline samples and intervention samples were estimated over actual weather encountered during the study. Subplot b) also shows a minimal detectable saving was observed early in week 12. It also highlights how consistently close the estimated interventon effect is to the true effect (without the effect of noise) throughout the study period. However, subplot c) indicates that outdoor weather has not reached the 80% threshold range calculated with reference to the typical meteorological year (TMY) weather file, we need to continue the M&V process in order to be able to normalize to a full range of weather conditions. At week 16, the accumulated temperature measured in sampled baseline days covered 80% range and at week 24, sampled intervention days satisfied the criterion. Therefore, we report a 9 kW savings at the end of this blocking period (i.e. week 24). If the analyst decides to continue sampling, the plot further shows the uncertainty associated with saving estimation further decreased throughout the remaining 24 weeks. Subplot d) plots the sequentially estimated normalized savings on the TMY weather conditions and outlines the check-point for satisfying each critical stopping criterion. It shows although the saving was detected early by the statistical test at week 12, the normalized annual estimation still largely fluctuates at the point. After covered sufficient weather conditions, the annualized savings stabilize at 17% which corresponds to the true TMY normalized savings as shown in subplot d).

Considering now the scenario where the analyst had applied conventional M&V following ASHRAE Guideline 14 or IPMVP, it would take 2 years to reach the same result (12 month baseline, 12 month intervention). The point at which the new method was able to accurately estimate TMY normalized savings was week 24. Using conventional M&V methods, this would merely be the mid-point of baseline collection assuming no interruptions or delays and they would need to wait for another one year and half to estimate weather normalized savings. Even if the analyst performed conventional M&V using a 6 month baseline and 6 month intervention period (1 year total), the result would be less accurate and still take longer than the method proposed above. Attempting to perform conventional M&V using less than that (e.g., 24 weeks) would yield nonsensical data as the baseline and intervention periods would span very different weather data ranges. Yet this randomized M&V method was able to achieve a valid, robust result in this time.

## 4.3 Scenario A: A discrete non-routine event

One common and typical non-routine event in an energy-saving M&V project is a change in occupancy, or a substantial change in occupant behavior. This is because occupancy has a substantial influence on building energy consumption but it is typically not closely related to intervention. In reality, it is very rare to have occupancy measurements (compared to other independent variables such as outdoor temperature and therefore) when a analyst fits a regression model (e.g. TOWT), thus, the analyst normally must assume that occupancy remains unchanged throughout the study. ASHRAE G14 and IPMVP only require the analyst to check for the coefficient of the variation of the root mean square error (CVRMSE) and normalized mean square error (NMSE) of a fitted baseline model. Thus, if occupancy changes after the retrofit implementation, the analyst can still obtain an ‘accurate’ baseline model with qualified CV(RMSE) but it is biased due to the unrelated change event being overlooked. The novel M&V method, on the other hand, resolves this by sampling randomly and frequently throughout the full distribution of occupancy over the full study period, and thus the results are robust to the change in occupancy.

Consider a scenario (Scenario A), where a tenant in a high rise office building moves out during the M&V study period, vacating one floor of the building. We expect the impact of non-routine events such as these to be balanced in sampled baseline days and intervention days if we sample randomly with equal probability. This means if occupancy is notably different over say, a 4 week period, the randomized M&V will sample approximately 2 weeks with baseline strategy and 2 weeks with intervention strategy. Therefore, the calculated energy savings can still represent the intervention’s true effect. Under conventional M&V, either the baseline or the intervention estimate will be affected by this difference in occupancy. To illustrate one example of this scenario, we modeled a situation where the 12-month baseline occurs under stable occupancy, and after the retrofit implementation, the energy consumption decreased by 10 kW for each metered interval from April to September because of decreased occupancy (i.e. tenants moving out leaving one floor vacant) and that the analyst was not aware of this change in occupancy.

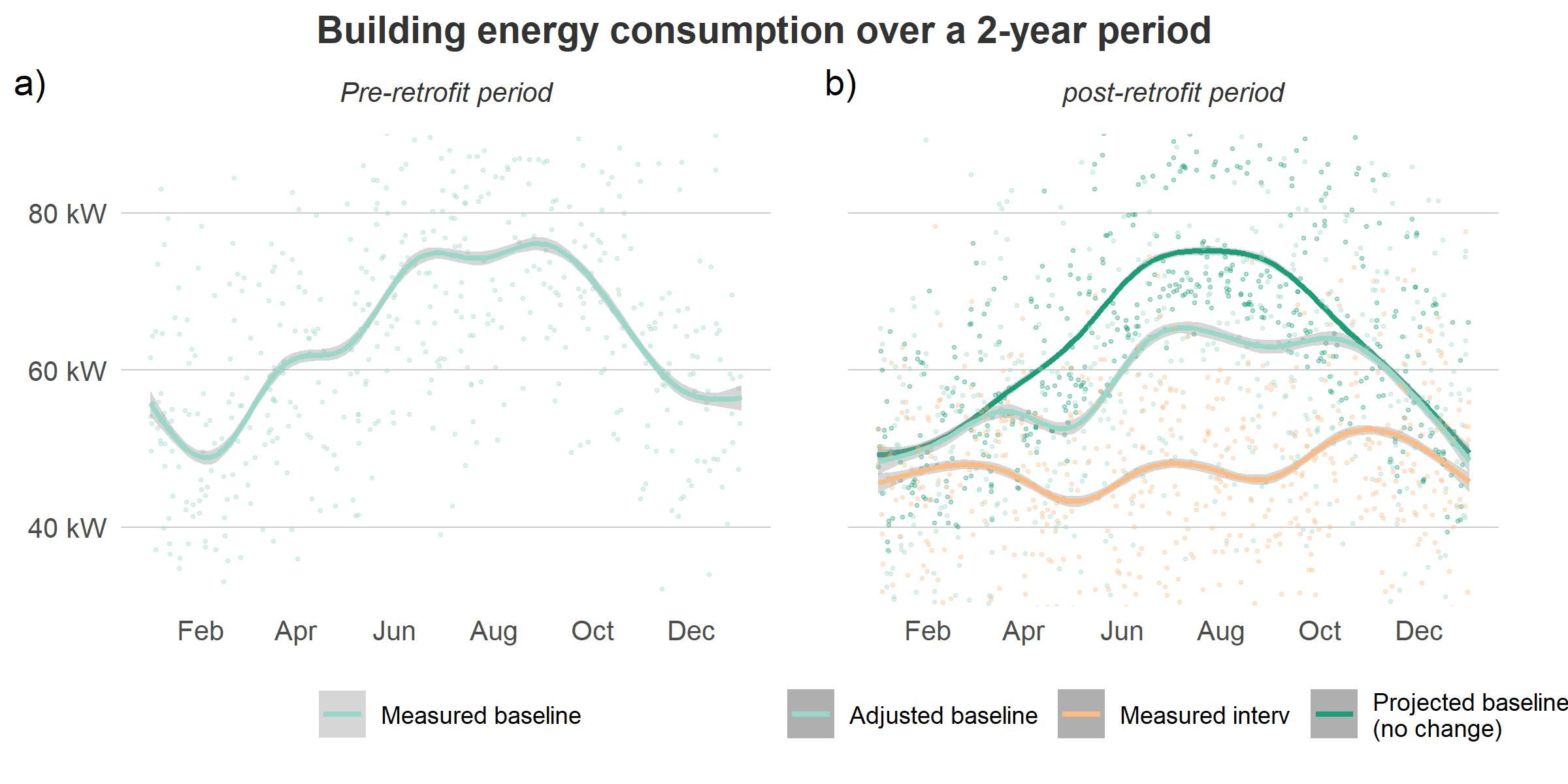


Figure 4.2: Energy measurement and modeling results for a 2-year conventional M&V project. Left: 500 random samples of measured baseline. Right: 500 random samples of measured intervention, projected baseline from modeling results without considering occupancy change and adjusted baseline measurement for occupancy change. (All lines fitted using loess function)

Figure 4.2 shows that if the analyst follows the steps described in G14 and IPMVP, they would start baseline measurements 12 months prior to the control retrofit implementation (subplot a) and fit an outdoor weather and time-dependent regression model using the collected metered data as the baseline. In other words, the model is only valid for stable occupancy. Then, during the intervention period, an unexpected change in occupancy occurs, which makes the baseline model no longer representative, but without a reasonable means of adjusting it. Here, conventional M&V would overestimate the savings by the intervention.

Instead, if the analyst follows the steps of the novel randomized M&V described in the manuscript, they would configure both control strategies in the BAS and switch back and forth between them depending on a pre-determined, randomly selected schedule. Figure 4.3 shows that even if the M&V test period spans the period during which occupancy changes, and the analyst has no prior knowledge about when and how long the occupancy has changed, this method will still find a reasonable estimate of the true effect of the intervention. This is because it is randomly and frequently sampling over the full distribution of occupancy experienced by the building.

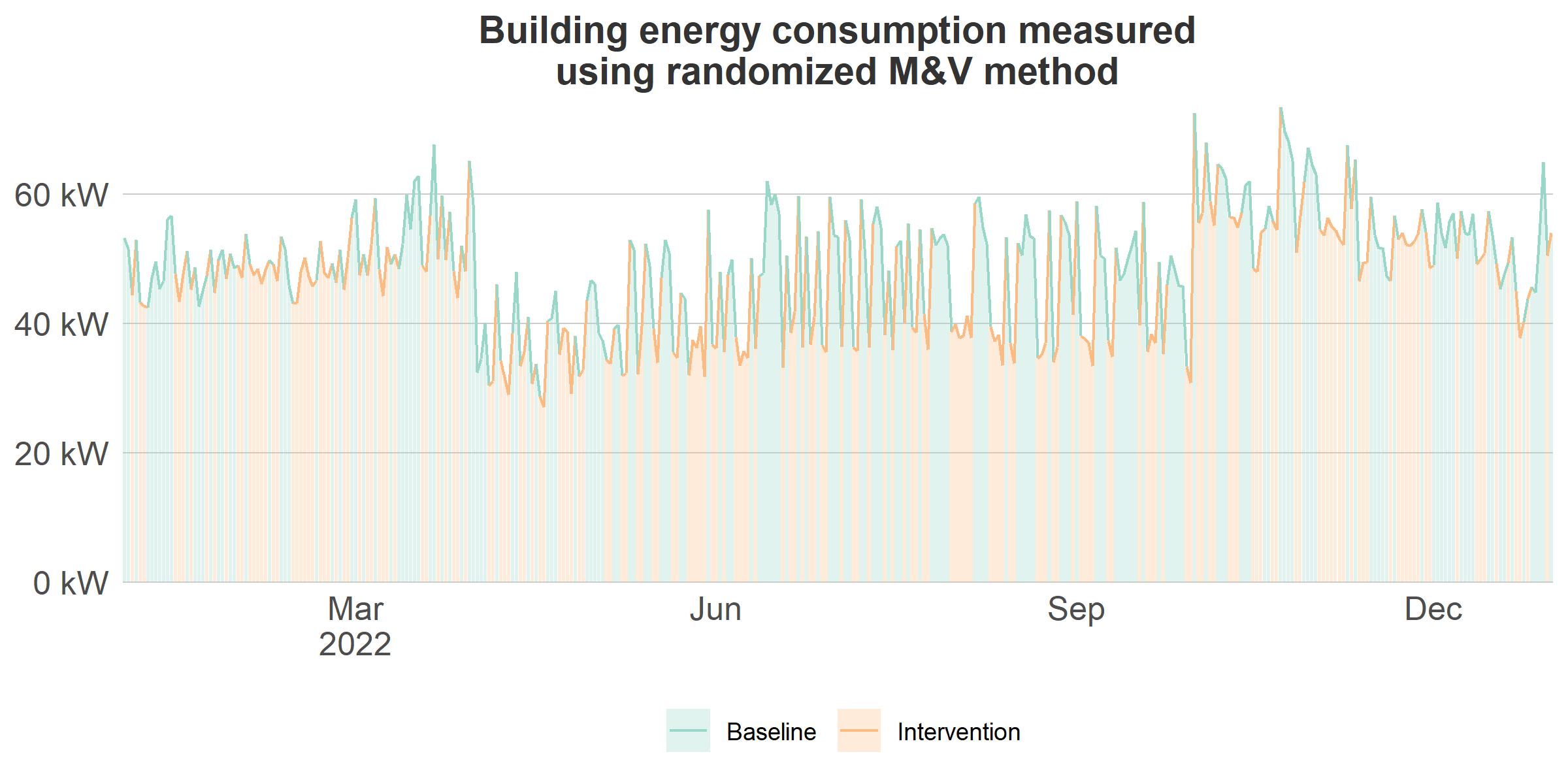


Figure 4.3: Averaged energy measurements at sampling interval (daily) using novel M&V method

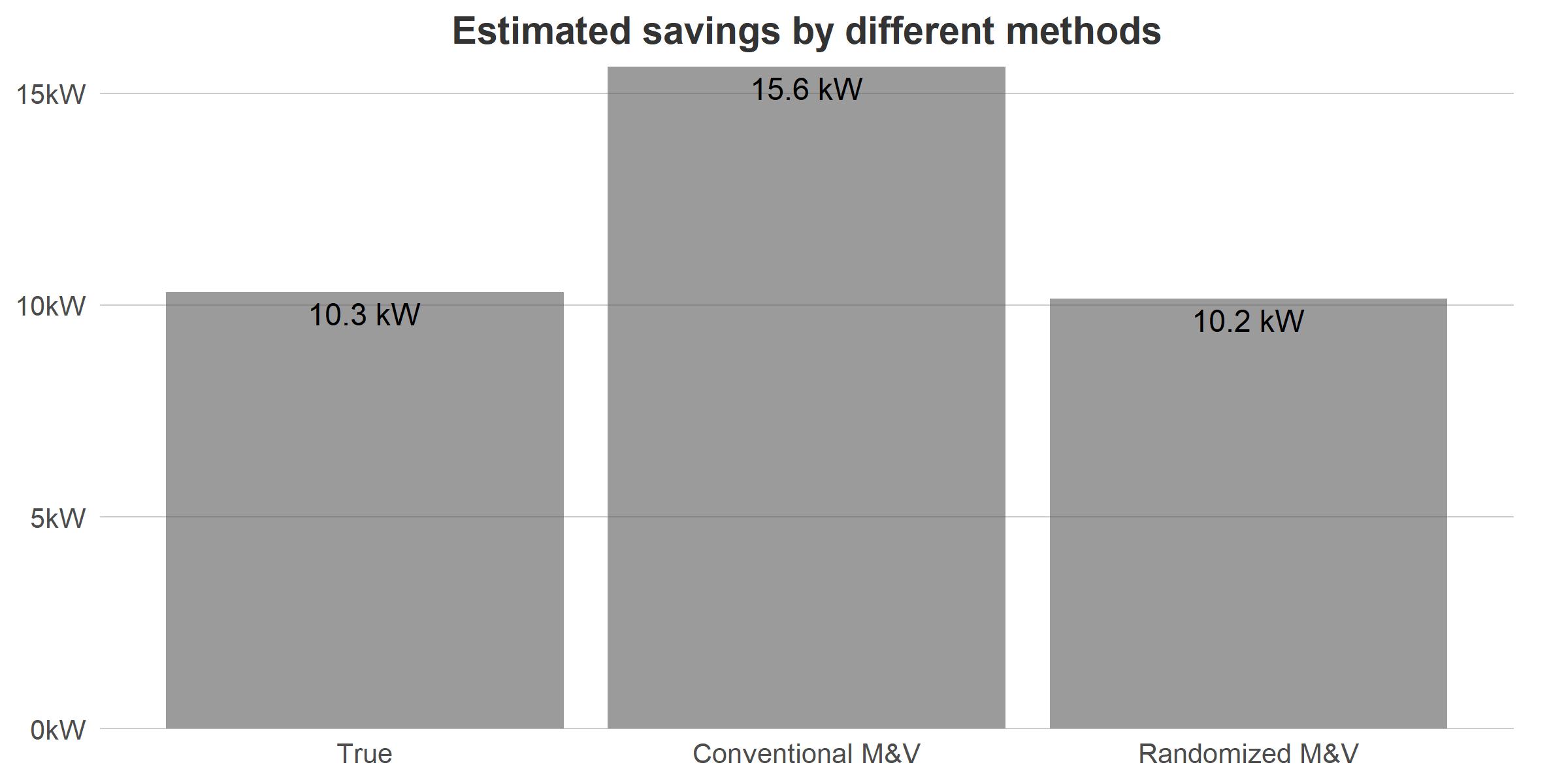


Figure 4.4: Comparison of savings estimation between conventional method and novel method proposed in the paper.

## 4.4 Scenario B: Continuous baseline change due to HVAC component degradation

Building performance can also change intermittently or continuously over time, which can be more difficult to detect. For example, gradually replacing older, lower performance lights as they fail with LEDs. Another example is a filter becoming increasingly clogged in an air handling unit, increasing fan power due to increased static pressure to overcome the increased resistance of air passing through the filter. In this scenario, if the building manager changes the air filter before starting the M&V and the analyst chooses to use the conventional pre-/post-type analysis, the energy measured 12 months later in the post-retrofit period should correspond to a higher baseline due to increased fan energy as the particle accumulates in the filter. Here, the conventional M&V method will underestimate the intervention energy-saving effect. If the analyst chooses to use the novel random sampling M&V, the results will be far more robust to this issue.

In this scenario, we simulated a situation where the fan power gradually increased due to accumulated particles in the filter. We assume the fan power accounts for 25% of the HVAC electricity consumption and due to clogging, system static pressure increases gradually by 50% over the course of the 2 year study period. The relationship between fan power and static pressure can be roughly estimated as , which means the fan power would also increase from 100% to 150%.

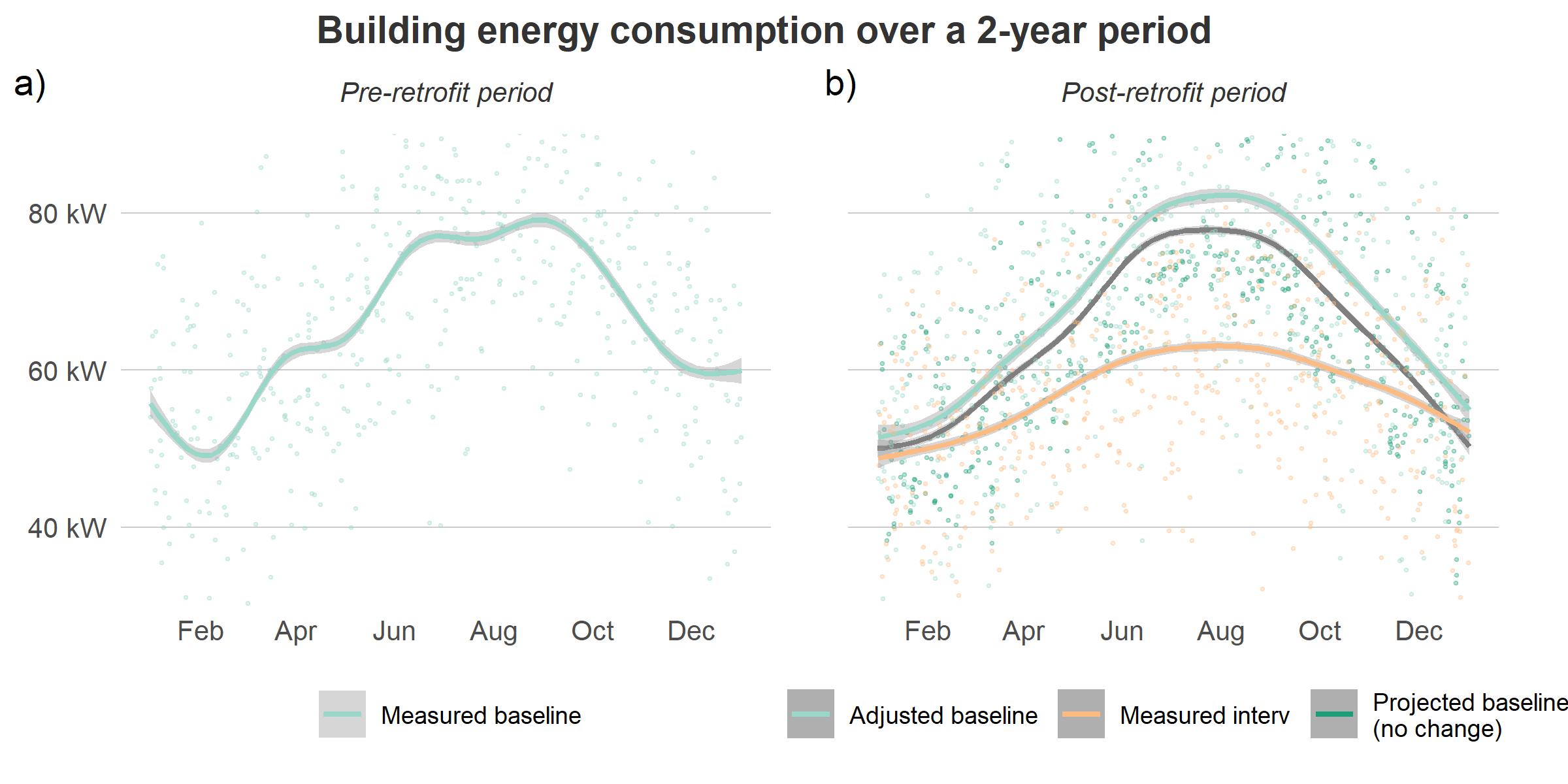


Figure 4.5: Energy measurement and modeling results for a 2-year conventional M&V project. Left: 500 random samples of measured baseline. Right: 500 random samples of measured intervention, projected baseline from modeling results without considering filter clogging and adjusted baseline measurement for filter clogging. (All lines fitted using loess function)

The above figure 4.5 shows both metered energy consumption and modeled baseline for a 2-year M&V project in this scenario. The figure shows that if using the conventional method, the analyst would start a 12-month baseline and fit a regression model using the collected measurements. Subplot b) shows that the projected baseline model will under-estimate the fan energy increase compared to a correctly adjusted baseline in the post-retrofit period. This will bias the energy savings estimate associated with the intervention, yielding a lower energy savings. However, if the analyst chose to apply the random switching method shown in figure 4.6, it will still accurately estimate the intervention effect.

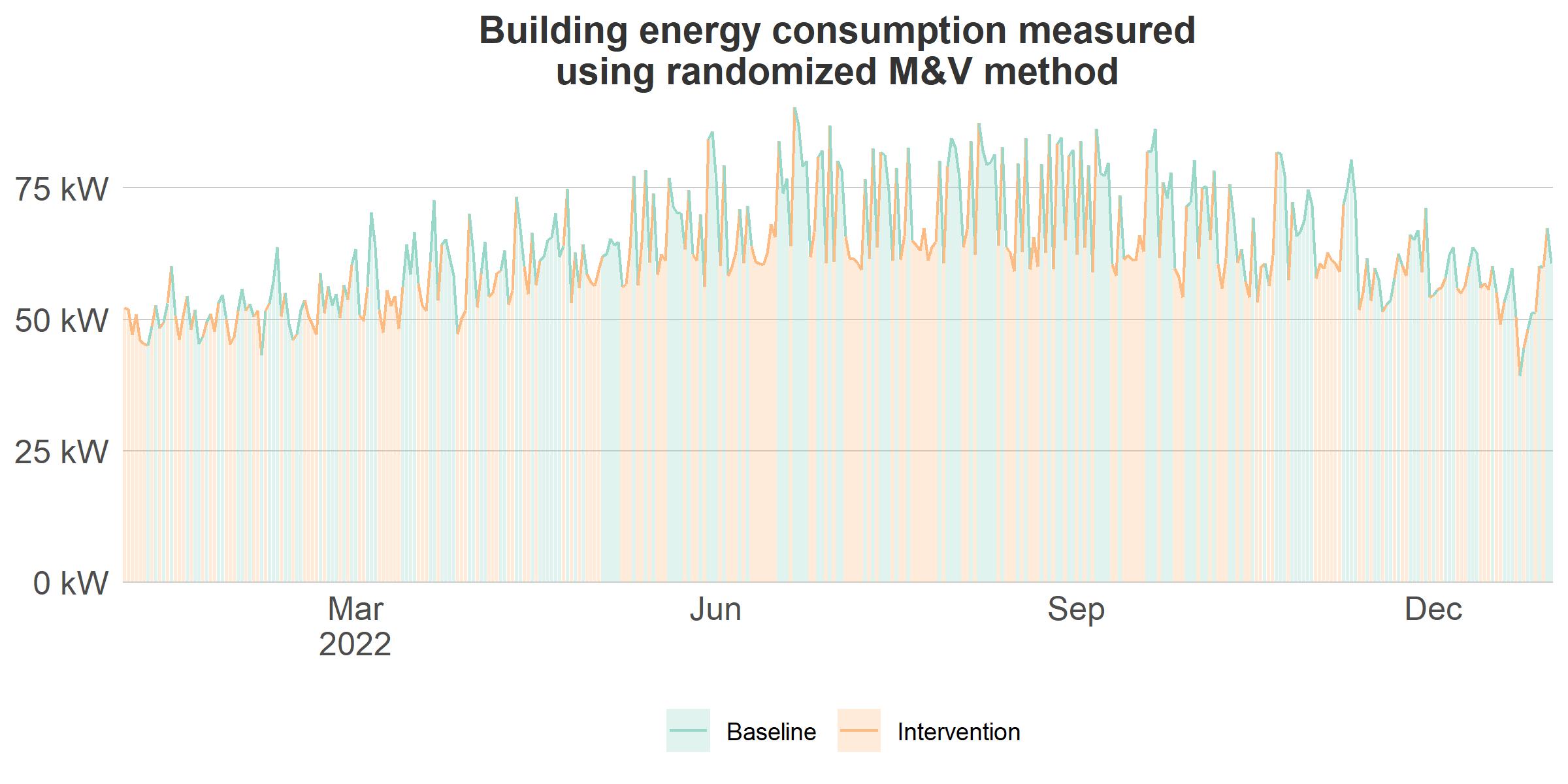


Figure 4.6: Averaged energy measurements at sampling interval (daily) using novel M&V method

Figure 4.7 implies that conventional M&V significantly underestimates the intervention energy-saving effect due to a lack of baseline measurements in the post-retrofit duration. Additionally as expected, when using the randomized M&V, savings are estimated much more accurately.

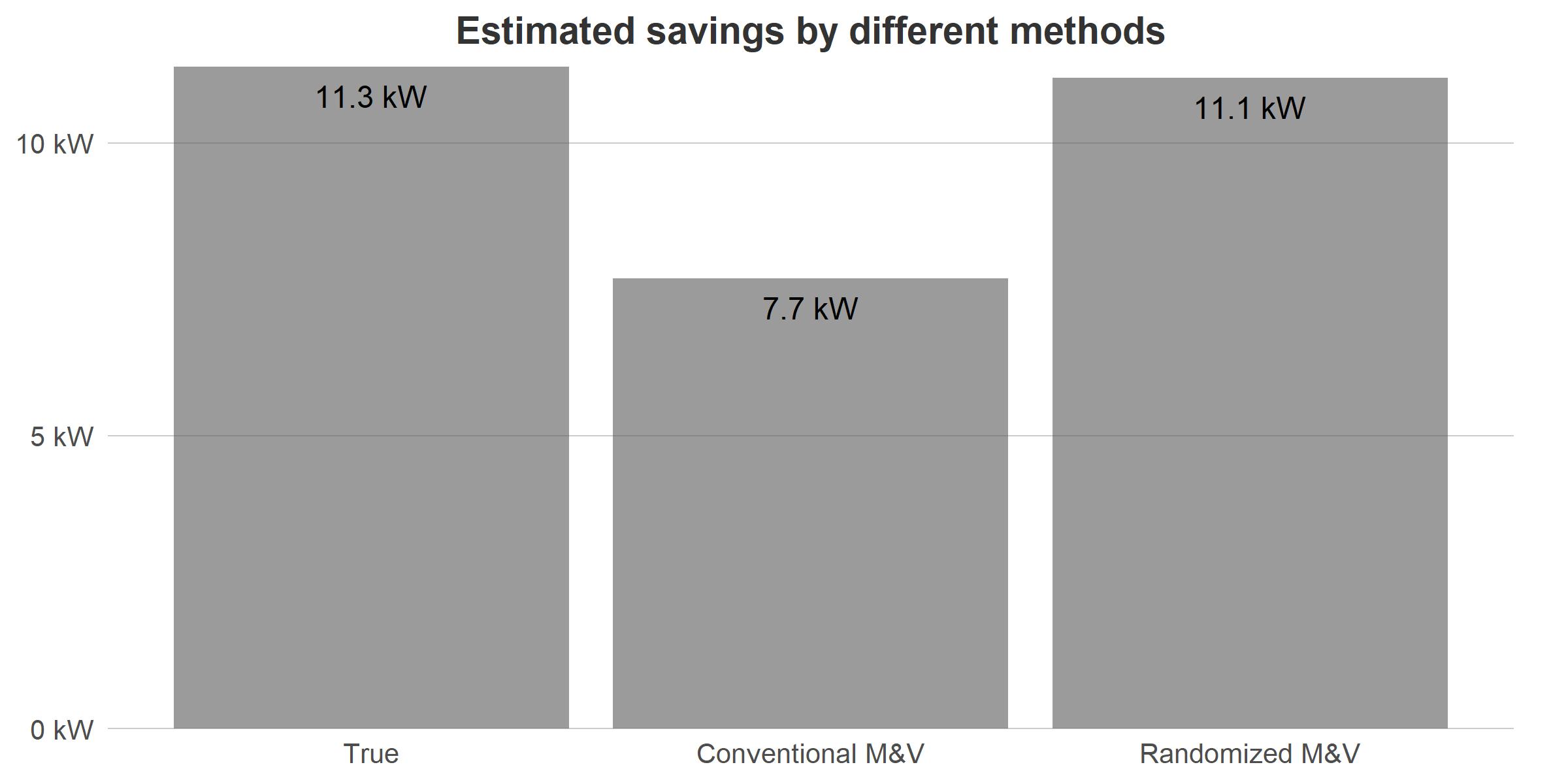


Figure 4.7: Saving estimation comparison between true savings and the two methods (conventional M&V method: no prior knowledge of filter clogging and does not measure baseline, and the novel randomized M&V method: no prior knowledge of filter clogging but measures updated baseline).