Data-driven approach for baseline construction in a demand flexibility future

Samuel Samuel Fernandes, Aditya Kanteti, Eliot Crowe, Jessica Granderson

IEPEC 2025

Problem

- Applying dynamic pricing approaches for utility tariffs, as a means to manage energy consumption relative to time-varying grid constraints, is being explored through a collaboration between CalFlexHub and Olivine in 20 homes.

Problem

- Aggregators have a business need to show individual homeowners the value of participating in a dynamic pricing program, and therefore there's a strong need to identify suitable M&V approaches.

Problem

- Existing M&V approaches include comparison group approaches, using data from participants and matched non-participants. Non-participants are chosen based on their pre-program consumption similarity to the participants.

Problem

- This approach has led to self-selection bias and affected the accuracy of energy savings estimates because the participants in rebate programs tend to have different energy usage patterns than the general population.

Research questions and proposed solution

Given this challenge, our research looked into the considerations for emerging M&V approaches (M&V2.0) for demand flexibility/grid services/GHG impact quantification of the future.

Research questions and proposed solution

– How does one construct a baseline with limited data?

Research questions and proposed solution

– How does one construct a baseline with limited data?

– What are the implications for practitioners?

Research questions and proposed solution

- We leveraged interval data and dynamic price signal information in the post-implementation period, and developed a data-driven (machine learning) approach to find 'untreated' time periods in a dynamic pricing future i.e we constructed a baseline with limited data.

Research questions and proposed solution

- This presentation describes the details of our approach, results, limitations, practitioner considerations and future research questions.

Smart thermostat operation in the context of this study
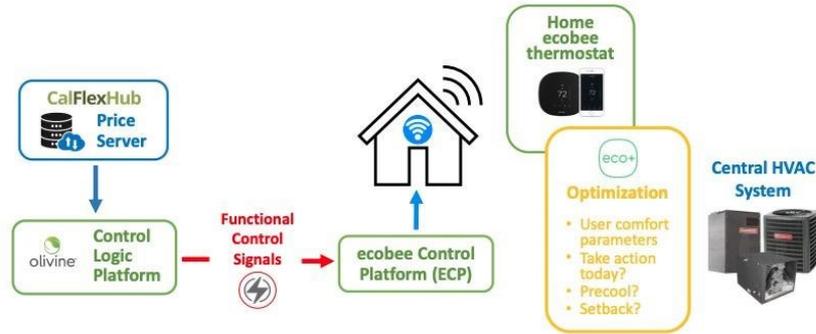
- Precool of 2-4 degrees turned on one hour prior to load shed event window (determined from day-ahead price signals)
- Event window set points were set up to 4 degrees F above original set point for 2-hours

Our dataset

- Interval data : 15-minute interval data load profiles for 20 homes during 2 seasons for the time period:
  - April 2024-May 2024
  - July 2024-August 2024
- Thermostat data : HVAC mode data (5 min interval) for the 20 homes
  - HVAC modes were heat, cool and off
- Dynamic price data: Hourly price data for the time period specified

Data-driven approach

- Our approach involves three steps:

  - Interval data *deconstruction* using price data
  - Baseline *imputation* using machine learning
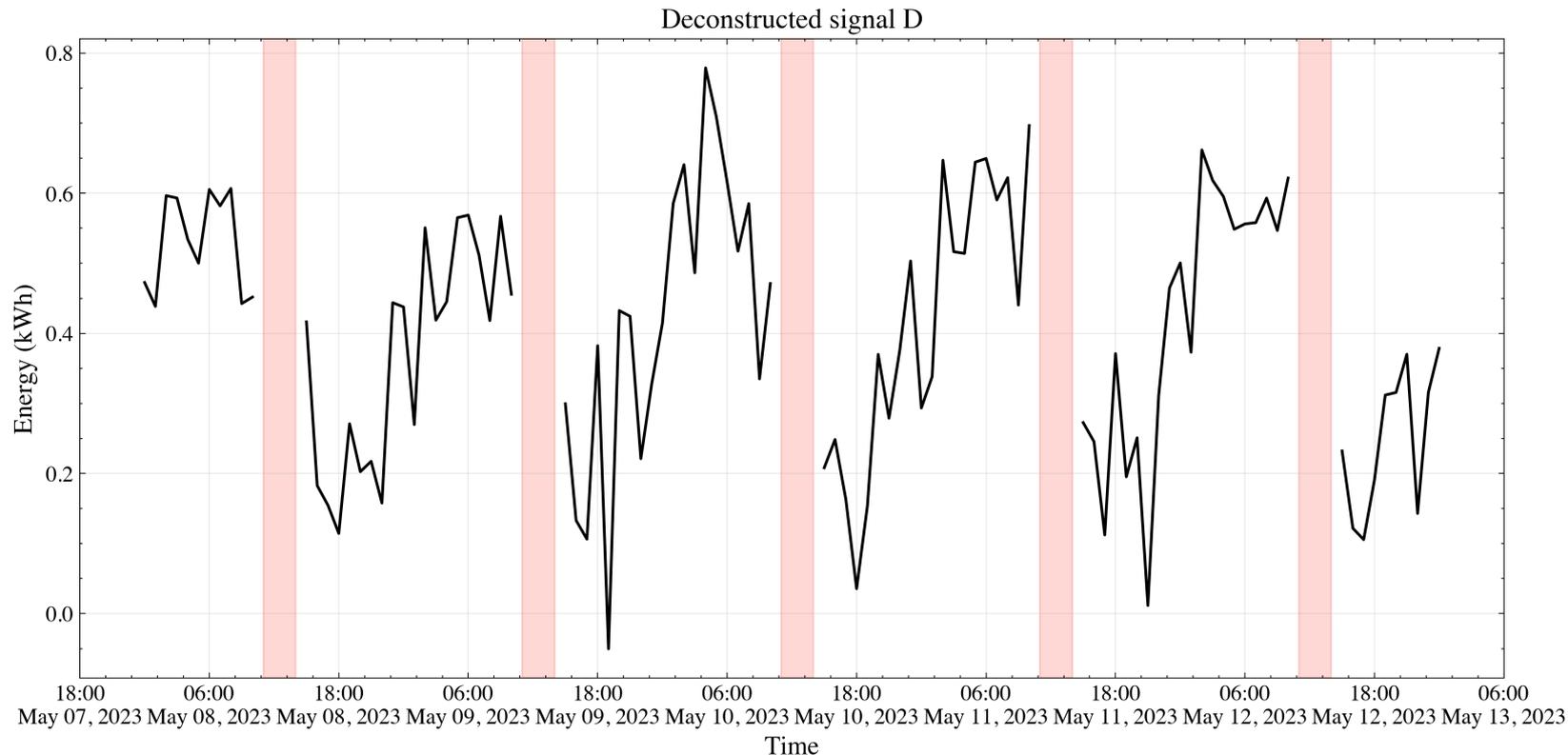 - Savings *evaluation*

Method

Identify periods when price was above 45 cents/kWh (shaded red)

Energy Consumption with Price Signal Overlay
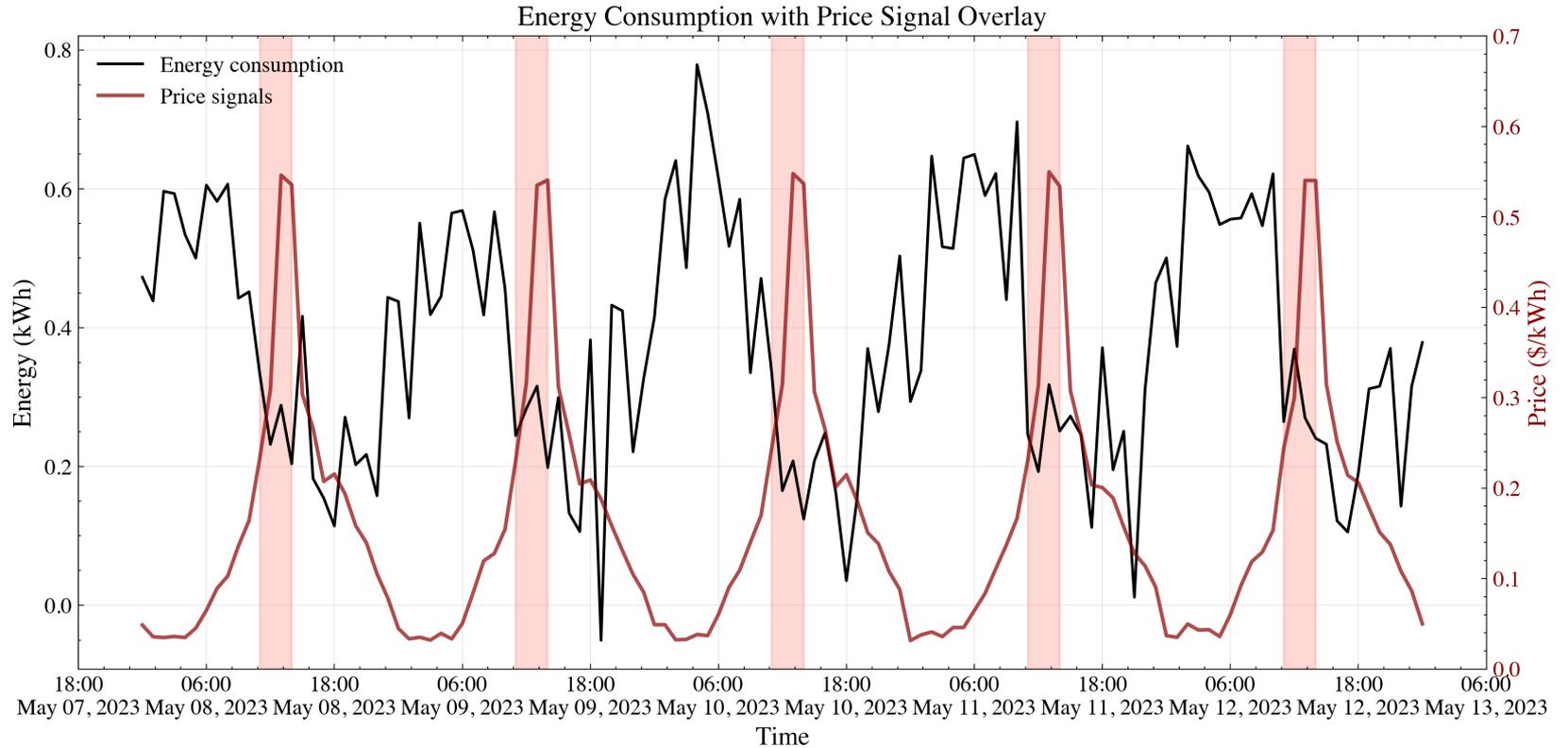
Method

Remove parts of the interval data where smart thermostat is known to have implemented controls, to get deconstructed signal *D*


Deconstructed signal D

Method

Impute the deconstructed signal using two machine learning algorithms. Visual inspection reveals good predictions.

Original data with imputed baseline using KNeighborsRegressor

Method

Impute the deconstructed signal using two machine learning algorithms. Visual inspection reveals good predictions.

Original data with imputed baseline using GBR
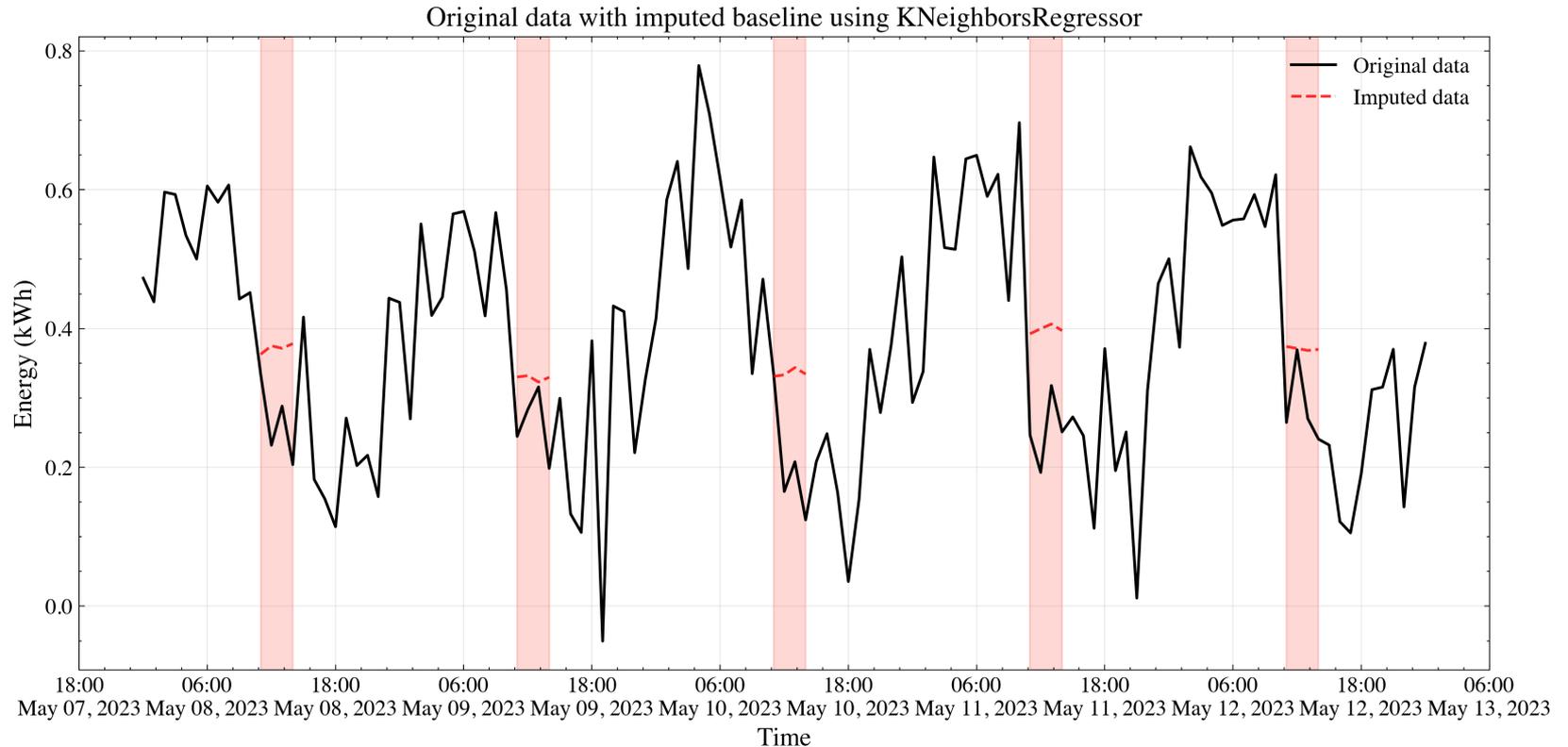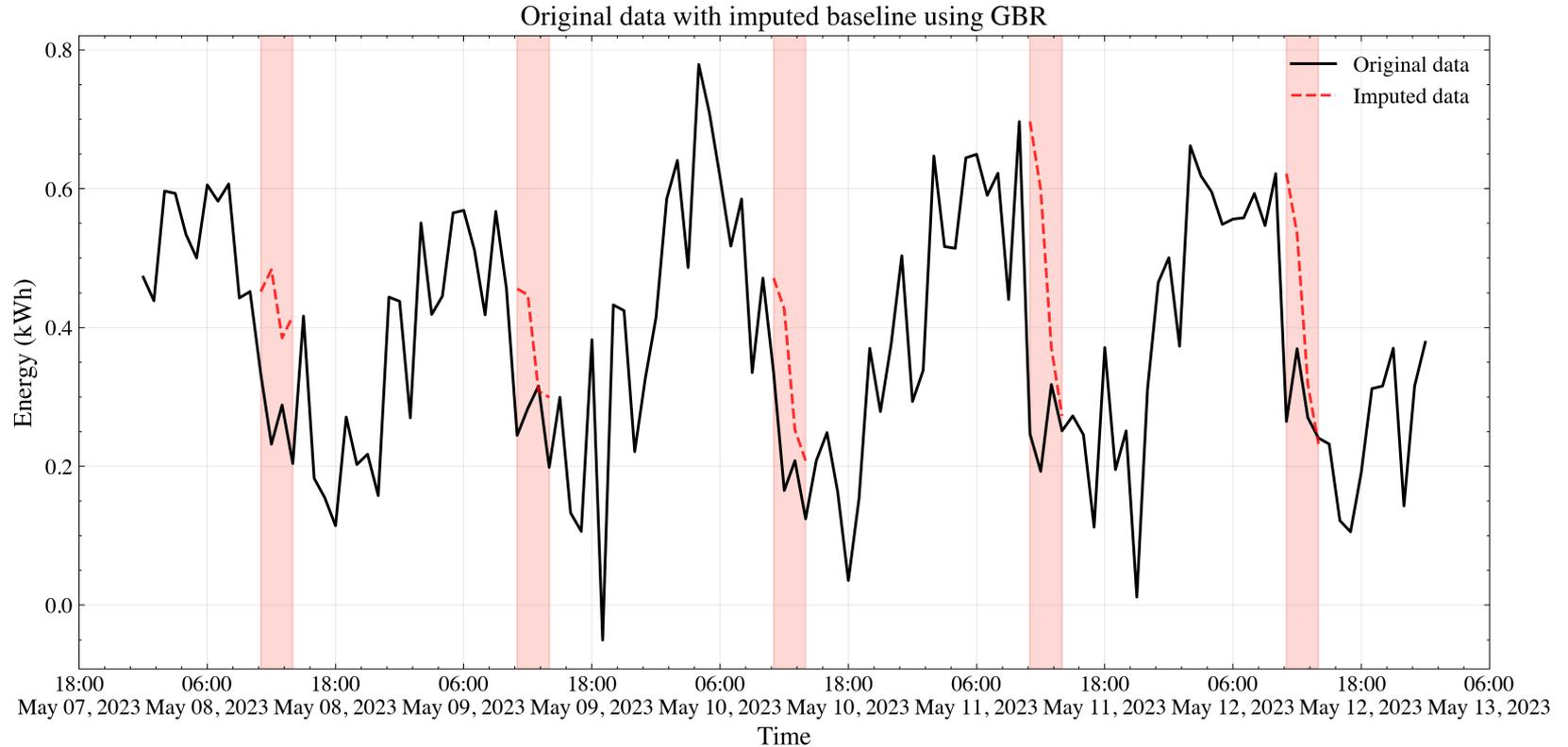
Method

- Case study analysis compared two machine learning algorithms:

- Gradient Boosting Regressor:
  - Builds an additive model in a forward stage-wise fashion;

- KNN Regressor:
  - KNN regression is a non-parametric method used for predicting continuous values.

Findings

Table 1: Seasonal Savings for the Models Used

Season	Model	Mean Savings (%)	Min Savings (%)	Max Savings (%)
Spring	GBR	9.3	8.6	10.3
Spring	KNN	10.1	9.4	11.0
Summer	GBR	9.9	9.6	10.7
Summer	KNN	10.6	9.9	11.1

Conclusions

- We developed an approach to find 'untreated' time periods in a dynamic pricing scenario, i.e construct a baseline with limited data.

- We applied our approach on a test dataset of smart thermostat implementation and were able to reconstruct baseline and estimate savings using this approach.

Conclusions

- Data-driven approaches offer promise to practitioners of addressing the challenge in finding 'untreated' time periods in a dynamic pricing scenario

- These preliminary results have demonstrated a concept to estimate savings without having a completed separate data period with no interventions