QUANTIFYING DEMAND FLEXIBILITY AT HOUSEHOLD LEVEL

ANALYSIS OF BASELINING METHODOLOGIES

JANUARY 2024

Centre for Net Zero

Powered by **Octopus Energy**

MEASURING DEMAND FLEXIBILITY

Demand flexibility incentivises consumers to shift their electricity consumption to benefit the system - for example, to times when supply is cleaner, cheaper and more plentiful. This will become increasingly important in energy systems powered by intermittent renewables.

Where flexibility is explicitly bought by a market operator, and provided by a market participant, we need **robust ways to measure the flexibility provided**.

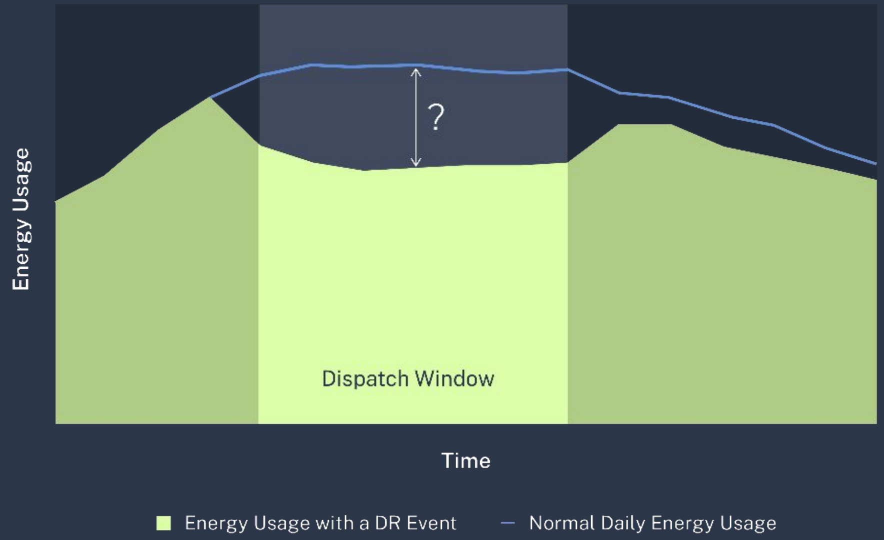
WHAT IS BASELINING?

Baselining is a way of quantifying, and therefore valuing, flexibility by measuring the change in energy consumption for a given household(s), business(es) and/or asset(s).

We define a baseline as an estimate of the electricity that would have been consumed by participants in the absence of a requested 'flexibility event'. This involves creating a **counterfactual**.

This can be used to **reward consumers** -either directly for the flexibility delivered or indirectly as part of a broader service.

This analysis focuses specifically on **baselining individual households** to accurately remunerate those customers for the flexibility provided during events.



Centre for Net Zero and others have published a new paper exploring different methods and use cases for baselining, calling for collaboration on approaches.

OUR ANALYSIS

We aim to understand how accurately different 'consumption profile' baselines (i.e. based on recent historic consumption data) in Great Britain perform in remunerating households for flexibility events.

We consider industry standard baselines from a variety of markets, using over 2 years of smart meter data. We look at flexibility events that are dispatched a few times a year for a couple of hours, with day-ahead notice, similar to ESO's Demand Flexibility Service in Winter 2022/23.

We use a sample of 10k Octopus Energy customers, randomised and accounting for:

- Low carbon technology (LCT) ownership
- Type of day (weekend/weekday, season)
- Time of day (settlement period, overnight/grid peak, etc)
- Weather

To consider performance, we look at errors by:

- mean absolute error (MAE), as an indication (in kWh) of how "wrong" the baseline is per settlement period for a household
- mean absolute percentage error (MAPE), as an indication (as a %) of how "wrong" the baseline is per settlement period, aggregated over the sample.

Our analysis makes some key assumptions:

- Future household demand is similar to historical demand
- Smart meter data is available and reliable at half-hourly resolution
- There are a few interspersed flexibility events these households have participated in; if events become very frequent, unaffected historical data may become outdated or insufficient.

RULE-BASED ALGORITHMS

Centre for Net Zero

Powered by **Octopus Energy**

RULE-BASED ALGORITHMS

We consider a range of rule-based algorithms used for baselining electricity consumption.

Short look-back (< 1 week)	**Previous similar day** Most recent similar settlement period on a similar day type
Medium look back (1 - 3 weeks)	**UKPN recent history** ~1 week of smart meter data from similar day type **P376* (with and without in-day adjustment)** ~2 weeks of smart meter data from similar day type.
Long look back (4 - 6 weeks)	**Four week average **** 4 weeks, using 4 data points for each half hour **Capacity Market** 2-6 weeks, with higher weighting for more recent similar days

Algorithms face a trade-off between using more historical data to capture variability in demand and therefore improving accuracy, and setting too high a bar for customer data requirements and therefore affecting participation.

* used for National Grid ESO's Demand Flexibility Service in 2022-2023 (with in-day adjustment) and 2023-24 (without in-day adjustment)
** Used for OE's Big Dirty Turndown Trial (2021)

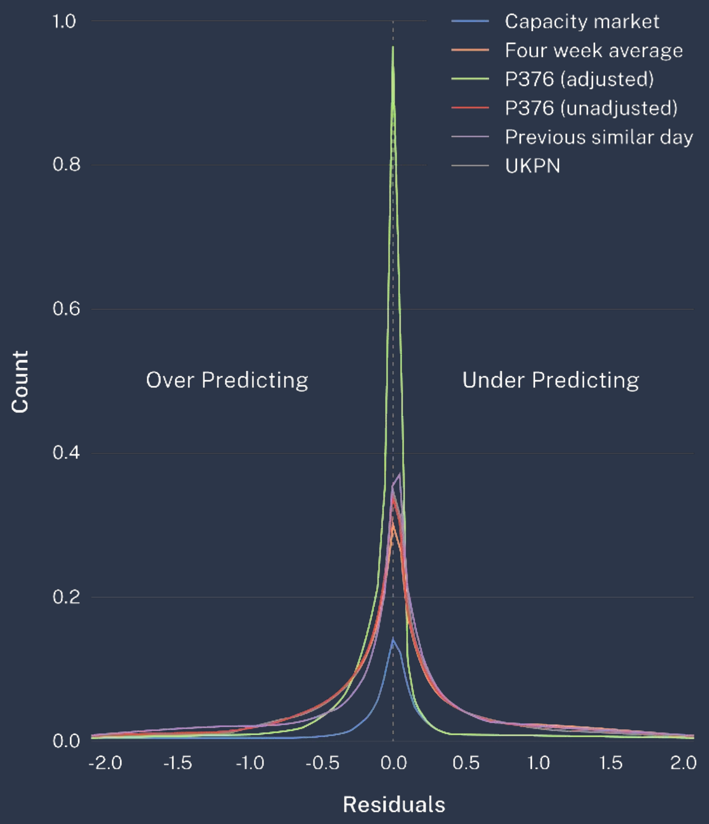
RULE-BASED ALGORITHMS
PERFORMANCE AT HOUSEHOLD LEVEL

- **Errors at the individual household level are high**: average consumption of 0.23 kWh compared to the smallest overall error of 0.13 kWh.
- **Distribution of errors are similar** regardless of the algorithm, as shown in the graph.
- **Errors depend more on the time of day** than methodology: errors overnight, when electricity demand is typically low, are significantly lower than errors in the evening, when flexibility events are currently needed to manage peaks in demand.
- **Adding more historical consumption data improves accuracy, to a point**. Using 1 week of data (e.g. UKPN) rather than the previous similar day reduces the overall error by ~14%. Capacity market captures a lot of variability but offers smaller additional improvement.

Summary of errors at household level

MAE (kWh)	Overnight	Morning	Evening	Other	Overall
Previous similar day	0.09	0.15	0.21	0.17	**0.16** (±e-5)
UKPN recent history	0.09	0.13	0.18	0.15	**0.14** (±e-5)
P376 (adjusted)	0.10	0.14	0.2	0.16	**0.15** (±e-5)
P376 (unadjusted)	0.09	0.13	0.18	0.14	**0.14** (±e-5)
Four week average	0.09	0.14	0.18	0.15	**0.14** (±e-5)
Capacity market	0.09	0.13	0.17	0.14	**0.13** (±e-5)
Average consumption	0.17	0.22	0.3	0.22	**0.23**

Distribution of errors at household level



RULE-BASED ALGORITHMS
PERFORMANCE AT AGGREGATE LEVEL

- **Using roughly 2 weeks worth of data generally results in lower errors** - for example, UKPN and P376.
- **Errors differ by the season** - regardless of algorithm, ‘shoulder months’ like April and November have higher errors.
- **Errors can become exaggerated when the method uses more historical data.** Data from more than a month ago becomes outdated: heating and electricity consumption behaviours may have changed.

Summary of errors at aggregate level

	MAE (kWh)	MAPE (%)
Previous similar day	103 (±2.2)	6.43% (±0.09)
UKPN recent history	79 (±1.6)	5.00% (±0.07)
P376 (adjusted)	83 (±1.4)	6.05% (±0.09)
P376 (unadjusted)	92 (±1.8)	5.82% (±0.08)
Four week average	105 (±2.1)	6.99% (±0.1)
Capacity Market	98 (±2.0)	6.16% (±0.09)

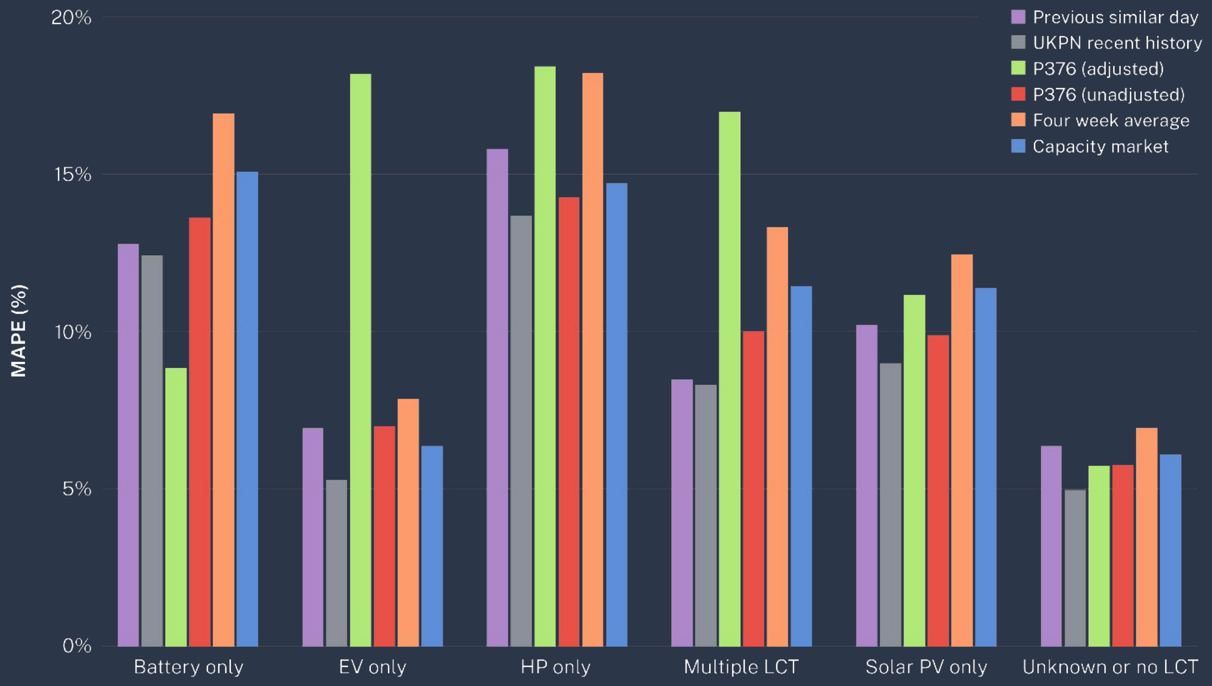
Distribution of MAPE in the sample by month (%)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Previous similar day	7	6	6	7	7	7	5	5	7	6	6	9
UKPN	6	5	5	8	4	4	4	4	4	4	6	7
P376 (adjusted)	6	5	5	7	6	6	5	6	6	7	7	7
P376 (unadjusted)	6	6	5	9	4	4	4	4	6	5	8	9
Four week average	7	7	7	10	5	5	3	4	5	5	11	13
Capacity market	7	6	6	8	4	4	4	4	5	5	9	12

RULE-BASED ALGORITHMS
PERFORMANCE BY LCT OWNERSHIP

- **Households with no LCTs generally have lower errors** than those with multiple LCTs, across all baselines. Households with batteries and heat pumps have particularly high errors - on average, more than doubling the MAPE.
- **Household with LCTs may have higher errors** possibly due to increased electrification leading to higher consumption. This may also be due to smaller sample sizes or the impact of weather on consumption (e.g. heat pump performance) that is not accounted for by rule based algorithms.

Mean absolute percentage error by LCT type



MACHINE LEARNING ALGORITHMS

Centre for Net Zero

Powered by **Octopus Energy**

MACHINE LEARNING ALGORITHMS

We consider a range of ML algorithms, which have been back-tested with data from up to 40k households over two years. We estimate these models will need to be retrained once a month and generalise reasonably to households not used in training.

The ML algorithms below, in order of increasingly complexity, use the following inputs

- Past 4 weeks consumption (same settlement period and day of week)
- Settlement period ('one-hot-encoded')
- Temperature (min, max, mean of the grid supply point)

Linear regression (LR)

Widely used, relatively easy to explain and interpret.

But requires a lot manual feature engineering to find a "good" model.

Implementations with and without temperature are included.

Multilayer perceptron (MLP)

May find nonlinear interactions between features without having to model it.

But harder to explain and interpret than LR.

Implementations with and without temperature are considered.

XGBoost

State-of-the art in time-series forecasting, using consumption each day in the most recent week, with little feature training required.

But hard to explain and interpret.

Both implementations include temperature but XGBoost (rolling) includes additional features such as rolling mean of consumption.

ML algorithms face an additional trade-off between models which have high forecasting capabilities and/or can account for different customer archetypes, and an approach which can be explained to consumers, implemented by providers and maintained easily.

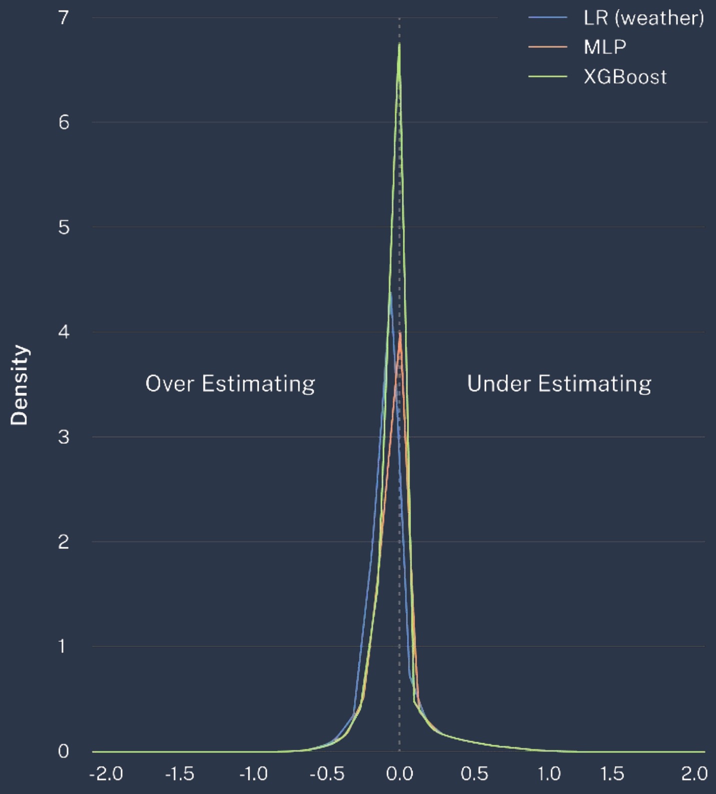
MACHINE LEARNING ALGORITHMS
PERFORMANCE AT HOUSEHOLD LEVEL

- **Despite the added complexity, in most cases there is no improvement in forecast accuracy at the household level** - the best-performing ML algorithm has the same error level as the best rule-based algorithm.
- As with rule-based algorithms, **errors are more dependent on the time of day than the methodology**, as shown in graph.
- LR tends to overestimate consumption, whereas MLP tends to underestimate.

Summary of errors at household level

	Overnight	Morning	Evening	Other	Overall
LR	0.17	0.14	0.19	0.15	0.16 (±e-5)
LR with weather	0.17	0.14	0.19	0.15	0.16 (±e-5)
MLP	0.09	0.13	0.17	0.14	0.13 (±e-5)
MLP with weather	0.11	0.12	0.17	0.13	0.13 (±e-5)
XGBoost	0.09	0.13	0.17	0.13	0.13 (±e-5)
XGBoost (rolling)	0.09	0.13	0.17	0.14	0.13 (±e-5)
Capacity market *Rule-based algorithm comparison*	0.09	0.13	0.17	0.14	0.13 (±e-5)
Average consumption	0.17	0.22	0.3	0.22	**0.23**

Distribution of errors at household level



MACHINE LEARNING ALGORITHMS
PERFORMANCE AT AGGREGATE LEVEL

- **LR has particularly high errors** as one model is not able to capture all the sources of variability; adding weather has some benefit in aggregate but not nearly enough to make it a competitive option
- At aggregate level, **even our state-of-the-art ML algorithm (XGBoost) is not more accurate than a simple rule-based algorithm** (e.g. UKPN).

Summary of errors at aggregate level

	MAE (kWh)	MAPE (%)
LR	293 (±4.0)	22.0%
LR with weather	271 (±3.9)	20.4%
MLP	89 (±1.21)	5.9% (±0.07)
MLP with weather	133 (±1.70)	9.6% (±0.13)
XGBoost	79 (±1.06)	5.4% (±0.06)
XGBoost (rolling)	79 (±1.05)	5.5% (±0.07)
Four week average *Rule-based algorithm comparison*	105 (±2.1)	7.0% (±0.1)
UKPN recent history *Rule-based algorithm comparison*	79 (±1.6)	5.0% (±0.07)

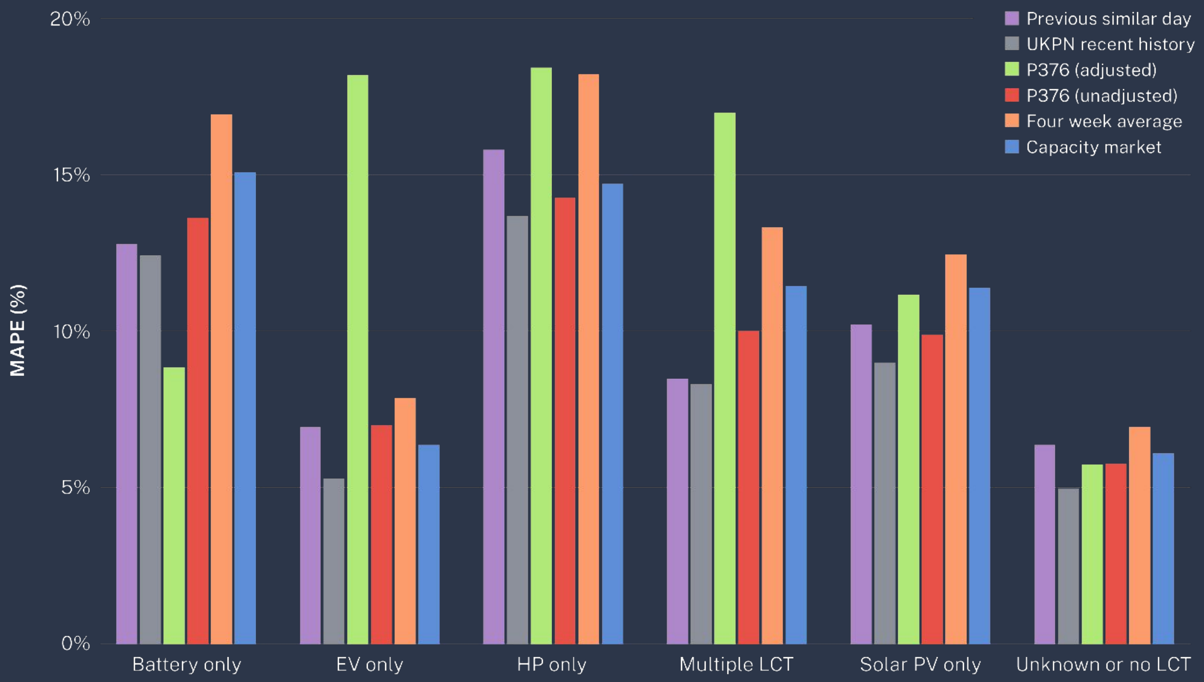
Distribution of MAPE in the sample by month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Four week average	7	7	7	10	5	5	3	4	5	6	11	13
UKPN	6	5	5	8	4	4	4	4	4	4	6	7
LR	16	19	18	24	26	26	27	30	24	23	16	17
LR (weather)	15	27	21	19	16	29	21	20	26	18	14	19
MLP	6	6	6	7	6	5	5	5	6	5	8	8
MLP (weather)	10	11	7	7	12	13	9	13	8	9	8	9
XGBoost	6	4	5	6	5	5	4	4	5	7	7	8
XGBoost (rolling)	6	5	5	7	6	5	3	4	6	7	7	7

MACHINE LEARNING ALGORITHMS
PERFORMANCE BY LCT OWNERSHIP

- **Households with no LCTs generally still show the lowest errors**, but the effect is less pronounced than for rule-based algorithms. LR also has very high errors for households with no LCTs.
- Compared to rule-based algorithms, **ML methods show the potential to reduce errors for households with LCTs** - for example, both MLP and XGBoost give lower errors for households with multiple LCTS compared to rule-based algorithms.

Mean absolute percentage error by LCT type



SUMMARY OF ALGORITHMS CONSIDERED

RULE-BASED

Errors at household level can be significant, and largest during grid peak - accuracy depends more on the time of day than the algorithm.

Roughly two weeks' worth of data results in lower errors. For example UKPN performs well for aggregated day-ahead forecasting and is more consistent throughout the year. Capacity market baseline works well at household level but errors increase in shoulder months, possibly due to changing consumption patterns.

Some sort of weighting towards more recent consumption may improve baseline accuracy.

There are high errors in the 'shoulder months' regardless of algorithm (April and November, December)

Errors are higher for households with LCTs, especially batteries and heat pumps. These may benefit from more complex algorithms.

MACHINE LEARNING

MLP and XGBoost are able to compete with rule based algorithms on accuracy. LR is unable to capture all the sources of variability, even when adding weather

ML algorithms add complexity for little gain. It may be possible to improve accuracy further, but this may come at the cost of how easily it can be explained to customers and implemented by flexibility service providers.

Temperature does not add much value for household level forecasting.

The advantage of ML algorithms may come for specific archetypes, i.e. households with LCTs and automation.

Possible improvements are more frequent retraining of the model, or adding input features:

- Lags of the same day
- Non-temperature weather data
- Rule-based forecast

KEY TAKEAWAYS

Baselines are sensitive to a number of key factors

Historical consumption data	More historical data, to a point, improves accuracy. Using roughly two weeks worth of data generally results in lower errors - more than that can increase errors as data becomes outdated. If averaging over more historic data, more recent data should be upweighted.
Time of day	Errors overnight, when electricity demand is typically low, are significantly lower than errors in the evening, when flexibility events are currently needed to manage peaks in demand.
Season	Regardless of algorithm, shoulder months like April and November have higher errors, possibly due to changing patterns in heating and electricity consumption.
LCT ownership	Households with no LCTs generally have lower errors than those with them. ML algorithms may add value for households with LCTs or some level of automation, which are likely to increase in future. LCT-specific features are worth considering to improve accuracy.
Period of day or time of year	Models may underpredict in key windows, which may depend on household characteristics.

There is a key trade-off between accuracy and simplicity

The key factors (left) should be accounted for, but one baselining methodology for every customer may not be appropriate. **There is a balance to strike** between improving accuracy and adding unnecessary complexity.

Rule-based methods are simple and perform reasonably well to baseline household level consumption today.

More complicated deep learning / gradient boosting methods may be more accurate, but harder to explain to customers, and hard to get flexibility service providers to implement. These may be more useful for internal remuneration.

NEXT STEPS

Towards a standardised approach to baselining

Centre for Net Zero will push for **industry-wide standardisation** of baselining to establish a fair playing field for market participants and consistency for the end consumer. Our joint paper begins to set out baselining principles, archetypes and use cases.

We are calling for a **working group** with a core focus of agreeing a set of baselining principles and guidelines for flexibility service providers, network and system operators.

This can look to develop a **library of baselines** which can guide market operators and participants to identify suitable baselines, considering the customer archetypes expected in a future energy system and how to adapt over time.

Suggested areas for future work

Baselines used during **extreme weather events** and at certain times of year, such as 'shoulder months'. Further work should also be undertaken to account for changes in **behavioural patterns over time.**

The **utility of applying ML algorithms**, especially for households with LCTs or automation.

Accurate baselines when there is **insufficient historical consumption data** or none at all.

Setting **thresholds for minimum constituent volumes** in cases where a higher level of aggregation is sufficient.

Validating baselining approaches using **control groups**

Confronting potential issues around **overlapping flexibility services**, and visibility and flow of data.

Please get in touch if you want to discuss the this analysis: info@centrefornetzero.org.

You can find out more about our range of ongoing research on our website: centrefornetzero.org

Centre for Net Zero
Powered by **Octopus Energy**