Centre for Net Zero
Powered by Octopus Energy

Quantifying demand flexibility

Towards a standardised approach to baselining

January 2024

octopus energy · enel x · Enedis · nationalgrid

EXECUTIVE SUMMARY

This paper outlines a potential set of common principles for quantifying demand flexibility, considering different methods and when these might be appropriate.

It draws on contributions from Centre for Net Zero, Octopus Energy, Enel X, Enedis and National Grid Electricity Distribution, as well as discussion with a range of parties across industry. We are building on existing work on this topic and calling for further collaboration - ultimately, we aim to work towards a more standardised approach to baselining in future.

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. Baselining is necessary to quantify, and therefore value, demand flexibility. It has a particular significance in "explicit" markets; where flexibility is bought by a market operator, and provided by a market participant. These rely on a way to measure the flexibility delivered, and as flexibility services expand, we need robust methods underpinning them to establish a fair playing field for market participants and consumers.

Four baselining principles, rather than one optimal methodology, allow us to balance trade-offs between competing objectives for different use cases:

1. **Accuracy** – ensuring that the flexibility provided is neither overestimated for the party buying it, nor underestimated for the party providing it.
2. **Fairness** – limiting opportunities to "game" baselines, while ensuring consistency for consumers and a level playing field for all market participants.
3. **Simplicity** – making it straightforward to implement, replicate, understand and verify from available data.
4. **Interoperability** – enabling easy access to different flexibility services, and a common framework for coordinating this.

Five broad baselining archetypes, commonly used by market participants, are considered:

1. **Fixed baseline** – the simplest method of all is to assume the "normal" consumption level of an asset (i.e. with a fixed profile), with any deviation from this regarded as delivery.
2. **Meter-before, meter-after** – using actual load data immediately preceding a flexibility instruction as a proxy for what the consumer would otherwise have used.
3. **Consumption profile** – using recent historical meter data to construct a counterfactual for what the consumer(s) would otherwise have used.
4. **Control group** – using statistical sampling to create a counterfactual for a portfolio of customers, based on data from similar customers who are not providing flexibility.
5. **Nomination** – a flexibility service provider generating a subjective forecast for the dispatch window when flexibility is to be provided.

Our recommended next steps to build consensus around a set of principles and guidelines are:

- **Set up a baselining working group**, made up of both buyers and sellers of flexibility, with a core focus of agreeing on a set of common principles.
- **Establish good practice guidelines** for market operators now and consider potential rules to regulate baselining in future aligned with the agreed principles.
- **Develop a "library of baselines"** to guide those buying flexibility to identify suitable baselines for the mix of customer archetypes in the future energy system.

INTRODUCTION

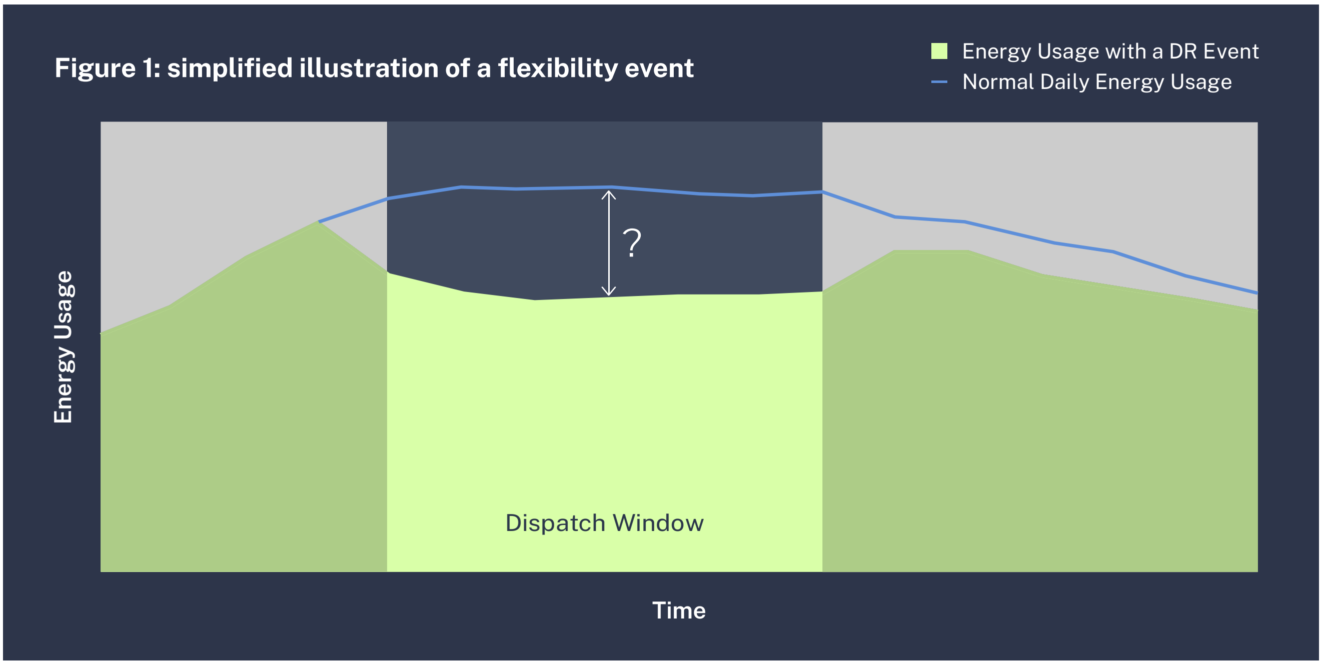
Demand flexibility encourages consumers to shift their electricity consumption to benefit the system - for example, to times when supply is cleaner, cheaper and more plentiful. This becomes increasingly important in energy systems powered by intermittent renewables. All consumers can play a crucial role in harnessing the power of the wind and sun, addressing grid constraints and driving down system costs for everyone. Globally, the International Energy Agency estimates we need 500 GW of demand response brought onto the market by 2030 - a tenfold increase on today.[1]

Demand flexibility is already happening at scale and it is growing. For example, in Great Britain the Electricity System Operator's (ESO) Demand Flexibility Service saw over a million consumers shift ~3 GWh of electricity from periods of peak demand in Winter 2022-23.[2] As such services expand, we need robust methods underpinning them to establish a fair playing field for flexibility service providers and consumers. There are many ways to deliver demand flexibility, with different methods for valuing its contribution to the system and rewarding consumers.

Implicit demand flexibility is delivered through dynamic pricing to reflect variability on the market and network, such as through Time-of-Use tariffs. **Explicit demand flexibility** is procured through incentive-driven services and can then be traded on markets, similar to generation flexibility. We expect implicit flexibility to grow significantly in future energy systems, particularly through automated response to price signals. However, we also expect explicit flexibility to have an important role to play, especially if targeted at times and locations it is needed most, or to provide greater certainty of response.

Baselining is necessary to quantify, and therefore value, demand flexibility. It has a particular significance in explicit flexibility markets (e.g. flexibility services) as the market operator must use some form of baselining to measure delivery. It is also important in the balancing and settlement process to assign energy flows between aggregators and suppliers. Determination of the change in energy consumption for a given household(s), business(es) and/or asset(s) requires a **counterfactual**: an estimate of the electricity that would have been consumed by participants in the absence of a flexibility instruction.

Figure 1: simplified illustration of a flexibility event



[1] International Energy Agency, Demand response analysis (2022) | [2] National Grid Electricity System Operator (Great Britain), Demand Flexibility Service

Implicit markets avoid the need for regulated baselining approaches as market participants simply respond to price signals, although baselining may be used internally to quantify the flexibility delivered. Explicit markets require baselining when any party is buying flexibility. The market operator must use a baseline to validate delivery and payment in the settlement process; it is therefore crucial that baselining methodologies are fair and do not allow for manipulation. Participation in explicit flexibility services is often managed by a third party on behalf of the consumer, who is rewarded either directly for the flexibility delivered or indirectly as part of a service.

Recent market changes to enable independent (i.e. aggregator) access to the wholesale market (e.g. Grid Code Modification P415 in Great Britain[3]) are also underpinned by baselining. Here it is critical that the baseline methodology accurately accounts for all energy associated with the flexibility delivered. Any inaccuracy would wrongly assign energy between parties and create an unfair playing field between aggregators and suppliers. For example, only accounting for reduction in consumption, with no accounting for energy recovery effects, passes the costs or benefits associated with recovery from aggregators onto suppliers.

Figure 2: Different services and markets which use baselining

Market Type	**Implicit**		**Explicit**			
Participant Type	**Wholesale Energy**	**Network Charges**	**Reserve**	**Constraint Management**	**Frequency Response**	**Capacity Market**
Balancing Responsible Party	No baseline	No baseline	Baseline	Baseline	Baseline	Baseline
Aggregator	Baseline	No access	Baseline	Baseline	Baseline	Baseline

Reaching consensus on baselining is crucial to ensuring flexibility services can scale while maintaining consumer trust. For example, the European Smart Grid Task Force found that the lack of an appropriate baseline is commonly identified as a barrier for access to the market, especially when the methodology is not transparent, standardised or accurate.[4] There are also evidence gaps around the performance of different baselines, particularly for domestic customers and low-carbon technologies, and important questions to consider, such as the level of aggregation required for an acceptable level of accuracy, the role of machine-learning (ML) algorithms, and the extent to which baselines should be technology agnostic.

In this paper we outline a potential set of common principles for baselining, considering different methods and when these are appropriate. We seek to build on existing literature on the topic and good practice already implemented across industry, calling for further collaboration on the issue. Ultimately, we aim to initiate work towards a more standardised approach to baselining in future.

This paper draws on contributions from Centre for Net Zero, Octopus Energy, Enel X, Enedis and National Grid Electricity Distribution, as well as discussion with a range of parties across industry.

[3] Ofgem, Balancing and Settlement Code (BSC) P415: Facilitating access to wholesale markets for flexibility dispatched by Virtual Lead Parties (October 2023) | [4] European Smart Grids Task Force, Demand Side Flexibility: final report (April 2019)

BASELINING PRINCIPLES

Baselining will never be a perfect science as it involves making assumptions about "what would have happened". There is no one optimal method, but different techniques which involve making trade-offs between competing objectives. In doing so, we should aim to balance these principles:

Accuracy

Ensuring that the flexibility provided is neither overestimated for the party buying it, nor underestimated for the party providing it.

Simplicity

Making it straightforward to implement, replicate, understand and verify from available data.

Fairness

Limiting opportunities to "game" baselines, while ensuring consistency for consumers and a level playing field for market participants. While perfect accuracy is not possible, inaccuracies should not systematically favour one party over another.

Interoperability

Enabling easy access to different flexibility services, and a common framework for coordinating this.

Following these principles will create flexibility markets that are reliable and trusted both by those who are buying flexibility, and those who are providing it. They are intended to build upon existing work on baselining methodologies, such as the Energy Network Association's Open Networks Project in Great Britain, the European Smart Grids Task Force at EU level, and the North American Energy Standards Board.[5]

[5] Energy Networks Association, Baseline Methodology Assessment (2020); European Smart Grids Task Force, Demand Side Flexibility: final report (April 2019); NAESB WEQ 015, Measurement and Verification of Wholesale Electricity Demand Response (December 2008)

BASELINING ARCHETYPES & USE CASES

Based on a range of approaches used by market participants, we see five broad archetypes:

1. Fixed baseline

The simplest possible approach is where the market operator assumes a consumption profile (usually technology specific) and any deviations from this in the metered position can be considered delivery.

Example: Distribution Network Operator (DNO) markets in Great Britain.[6]

2. Meter-before, meter-after

Using actual load data immediately preceding an instruction as a proxy for what the consumer would otherwise have used. This very simple method can be suitable in evaluating flexibility when the duration of response is seconds or minutes.

Example: Frequency Control Ancillary Services in Australia.[7]

3. Consumption profile

Using recent historical consumption data to create a counterfactual. This method is particularly useful for sites with regular patterns or where variations (e.g. due to weather) can be sufficiently well-captured by in-day adjustments. This method can be used frequently, but if events become too frequent, there might not be sufficient data once previous activations are excluded.

Example: Capacity Market in Great Britain.[8]

4. Control group

Using statistical sampling to create a counterfactual or a portfolio of customers, based on data from similar customers who are not providing flexibility. This method can be used to remunerate a flexibility service provider for the total flexibility provided by their portfolio. Provided that sufficient data for statistical sampling is available, it is also valuable in cases where data quality varies in the portfolio.

Example: In France, grid operator Enedis's "panels method" (discussed further below), although used more regularly in other industries (e.g. clinical trials).

5. Nomination

A flexibility service provider generating a subjective forecast of flexibility for the dispatch window. This method is well established in energy markets and works effectively so long as providers are incentivised or required to submit accurate predictions.

Example: Balancing Mechanism in Great Britain uses the principle of physical notifications (nominations) to determine the delivery of flexibility, and crucially participants must submit these constantly and are obligated to follow them if not dispatched.[9]

[6] Energy Networks Association, Open Networks Project, DNO flexibility Services, p.29 (2019) | [7] Australian Energy Market Operator, Market Ancillary Service Specification (2023) | [8] National Grid, Capacity Market DSR Testing Process (June 2017) | [9] National Grid Electricity System Operator, What is the Balancing Mechanism?

These five baselining archetypes are discussed further below; they are not exhaustive and approaches will evolve over time. There are alternatives not considered in detail here, such as:

- **Maximum baseload or "drop to" approach**, which uses historical meter data to generate a flat level of electricity demand that the customer is rewarded for remaining at or below.
- **Zero baseline or generation approach**, whereby the baseline is simply set to zero, commonly used to measure flexibility provided by sub-metered behind-the-meter generators which would not normally generate (i.e. back-up diesel).
- **Non-meter data**, whereby a baseline is constructed using non-meter data where the relationship to energy consumption is well understood. An example is using ambient temperature to infer what domestic consumption would have been on a cold day.

Baselining approaches need to adapt to changes in the energy system. Traditionally, market operators have adopted consumption profile baselines using relatively simple algorithms. This approach works well for more traditional demand response, such as from industrial consumers or behind-the-meter diesel generators, but is likely to be less suitable as demand flexibility providers evolve, including to domestic consumers and different low-carbon technologies. Some notable issues which arise with consumption profile baselines are:

- **Variable loads** – while industrial sites may have highly regular patterns of operation, many residential loads are more variable, due to human behaviour being less predictable (e.g. spontaneous energy use or different responses to external temperature). Baselining approaches need to account for this while mitigating against gaming.
- **Regularly activated loads** – while industrial loads are typically activated infrequently (at high cost), other assets (e.g. EVs) can be activated regularly, across multiple markets, making it harder to construct a counterfactual from historical data during inactivated periods. In such cases, alternatives will need to be considered, such as nomination baselines.
- **Energy recovery** – demand response is increasingly provided by demand shifting, rather than demand destruction (e.g. from diesel generators), with any shifted energy use recovered afterwards. Baselining will need to account for energy recovery effects.

Baselining approaches should be in line with the use case. Accuracy is often traded off against simplicity -for example, buyers of frequency response need a completely different response rate and accuracy from buyers of constraint management products. Often, it is also convenient to separate the arrangement between the flexibility service provider (i.e. the energy supplier or third party aggregator) and the system, from that between the flexibility service provider and the end customer. These can be decoupled and use different baselines. An important consideration in determining a suitable baseline is the notice of dispatch and the duration of response, as well as the frequency at which flexibility is dispatched.

Baselines for commercial services also differ from those used in forecasting for network planning purposes. Network operators can take approaches to baselining, such as the "drop to" method, which reflect assumptions made in their planning timeframes. As a network operator procures flexibility in more operational timeframes, these assumptions will change. While the use case is different, baselining in network planning should still aim to balance the principles outlined above.

1. FIXED BASELINE

The simplest approach to creating a counterfactual is to use a fixed baseline where the assumed behaviour of an asset, or group of assets, is set in a static profile, to which the metered output is compared. This approach can work well for highly controllable loads which might have high capital costs to establish control (i.e. build-out of integrations and automated systems) but, once established, low cost to dispatch. For example, the average consumption of an EV fleet can be used as an effective counterfactual.

This approach has been adopted by some DNOs in the creation of local flexibility markets because they can think more long term (i.e. on the timeframe of infrastructure build) to invest in such systems. A pre-fault peak turn down service, where the provider shifts highly controllable load off-peak everyday (e.g EVs), can be used to address local congestion constraints. As a consumption profile baseline, which uses recent meter data, would not show any change in consumption behaviour in this case, the fixed baseline is a better means of rewarding participation.

2. METER-BEFORE/METER-AFTER

The “meter-before/meter-after” approach is a flat baseline set based on consumption immediately before an event, using a single meter reading or average/median/min/max of a few readings. Meter readings during the event are compared against the meter readings prior to the dispatch window to calculate the delivered flexibility.

Meter-before/meter-after is widely used to estimate the level of service delivered under real-time dispatch conditions and short utilisation periods. It is also a preferred baseline for frequent dispatches as it is not dependent on historical data. It is more commonly used in the reserve and frequency regulation markets, compared to the other baselining archetypes discussed here.

This method becomes less accurate as dispatch durations increase beyond a few minutes as it is no longer reasonable to assume that consumption would not otherwise have changed. Where there is a ramp-up or ramp-down period, or the customer is given notice of a dispatch, the “meter-before” period needs to be set before the customer could possibly respond, otherwise the baseline will be biased or “gameable”.

3. CONSUMPTION PROFILE BASELINES

There is a range of methodologies that use historical electricity consumption to create a counterfactual for a consumer. These can be rule-based algorithms, as well as statistical models or ML algorithms. Other approaches use more complex algorithms to provide a more accurate estimate, particularly at the level of an individual consumer, usually with these common traits:

- **Similar days in recent history are used**, e.g. baselines for working days are calculated using consumption from previous working days. This underpins a fundamental assumption in baselining that future demand profiles are similar to recent historical demand.
- **Previous days with flexibility events are excluded**, as the demand on those days is not reflective of load in the absence of flexibility events. This can be problematic in programmes which are dispatched very frequently, as there may not be enough uncontaminated days in recent history.
- **Recent historical meter data is available and reliable** at the necessary temporal resolution, which may not always be the case in an operational setting.

One example of a consumption profile baseline using a simple rule-based algorithm is the “high 5-of-10 methodology”, shown in Figure 3 below. This identifies five days of the 10 most recent similar days with the largest daily consumption, excluding non-working days and days with flexibility events, and averages consumption in the equivalent time period from the identified five days to create a baseline. The underlying assumption here is that flexibility events are more likely on high consumption days; where that’s not the case, a 10-of-10 methodology may be more appropriate.

Figure 3: High 5-of-10 consumer profile baseline methodology

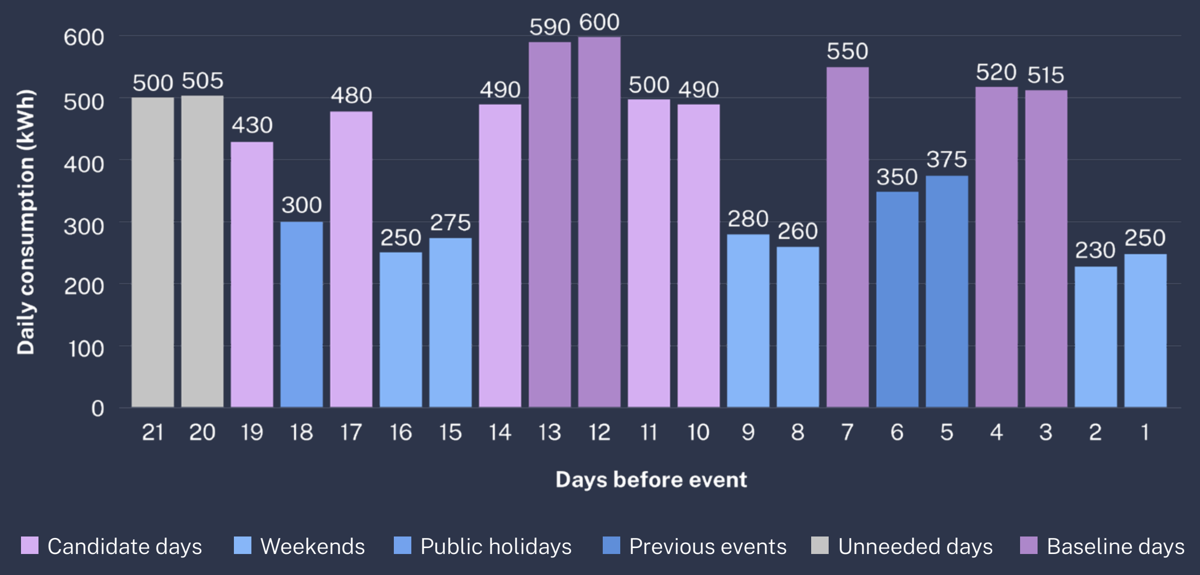
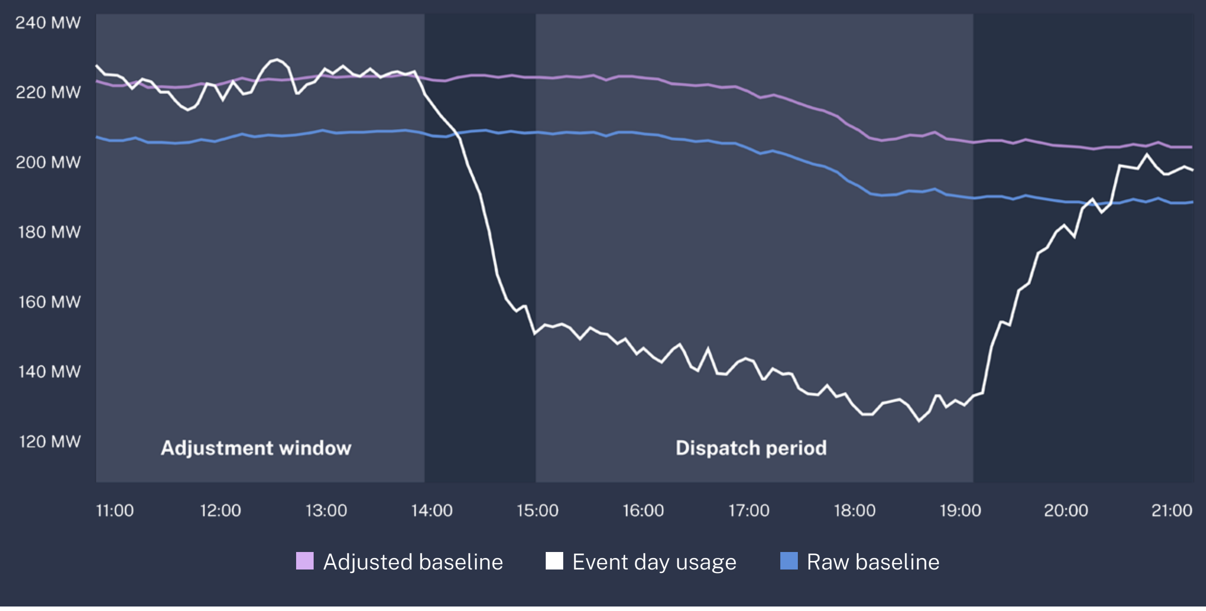


Figure 4: Example of in-day adjustment to a consumer profile baseline



The impact of changes in weather on consumption can be accounted for by using an in-day adjustment (see Figure 4). For example, if a cold spell was followed by a warmer day, the consumption of households with heat pumps may be lower regardless of flexibility activation. Correcting the baseline by using data from earlier on the same day as the flexibility activation can capture this expected change in consumption.

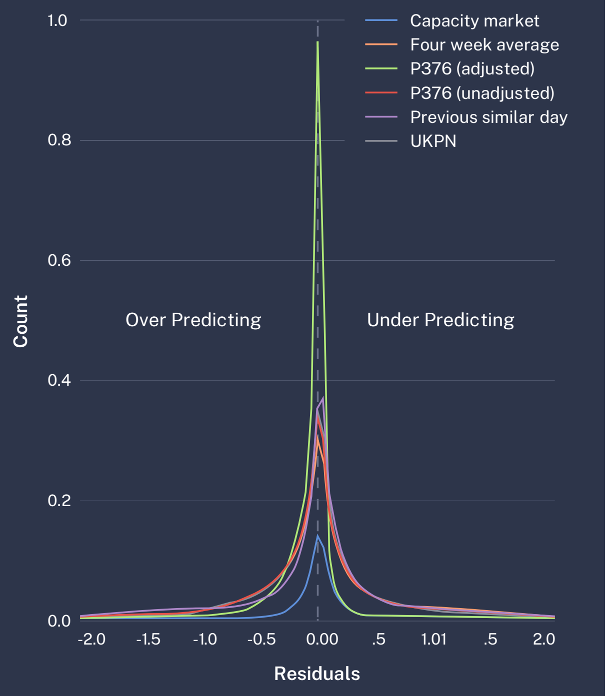
While in-day adjustments can help to correct for the effect of external factors on consumption profiles, they are susceptible to gaming if the adjustment window falls after notice of a flexibility event is given. **In principle, a methodology should never allow a way for the customer to influence the baseline once they know they are going to be dispatched.** Therefore, there is a trade-off between providing consumers with sufficient notice of the flexibility event (e.g. day ahead), which is likely to be simpler operationally, and allowing for in-day adjustments nearer to the dispatch period, which can increase accuracy but should not be allowed after notification of the event. In Great Britain, the Demand Flexibility Service allowed in-day adjustments for residential loads in Winter 2022-23, but the system operator has removed this feature for Winter 2023-24.[10]

Centre for Net Zero (CNZ) has conducted analysis which considers the performance of a range of consumption profile baselines, using over two years of smart meter data. This analysis specifically considers the use case of flexibility events that are dispatched a few times a year for a couple of hours, with day-ahead notice – similar to the Demand Flexibility Service in Great Britain. Looking at both rule-based and ML algorithms, it evaluates the accuracy of baselines to proportionally remunerate participating households for their flexibility.

Different rule-based algorithms have a similar distribution of errors, as shown in Figure 5. Overall, errors can be material at household level: average consumption of 0.23 kWh compared to an average error of 0.13 kWh. This reduces at aggregate level: an algorithm which uses roughly one to two weeks' worth of historical data has a mean absolute percentage error of 5%.

[10] National Grid ESO Demand Flexibility Service Participation Guidance (August 2023)

Figure 5: Distribution of errors for six different consumer profile baselines (CNZ analysis)



A number of factors appear to affect the accuracy of the baselines for this specific use case:

- **Historical consumption data** - more historical data, to a point, improves accuracy. For example, an algorithm using a week of data reduces the overall error by ~14% compared to one using just the previous similar day. However, at the household level, the increase in accuracy when using four or more weeks' worth of historical data compared to using two weeks of data is small. In fact, at the aggregate level it increases the error, possibly because consumption behaviours, such as heating, have changed in that amount of time. Using roughly two weeks of data generally results in lower errors.
- **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** - looking at the aggregate level, CNZ analysis showed that errors differ by 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 LCTs, and households with batteries and heat pumps have particularly high errors.

ML algorithms show a similar distribution of errors to rule-based ones, which again depends more on the time of day than the methodology. However, despite the added complexity, in most cases there is no evidence in this analysis of improvement in forecast accuracy at the household level. The advantage of ML algorithms may come when we look at improving accuracy for specific types of households. Households with LCTs and some level of automation, which we expect to increase in future, may benefit from having more complex algorithms.

Overall, CNZ analysis points to some emerging conclusions for consumption profile baselines for residential customers in the type of flexibility events considered (infrequent, behavioural response):

- **When averaging data, more recent data should be preferred** - roughly two weeks of historical data might be sufficient - and if using more historical data, more recent data should be up-weighted.
- **Baselines are sensitive to a number of key factors**, including LCT ownership, period of day and time of year.
- **There is a trade-off between simplicity and accuracy**; rule-based methods are simple and perform reasonably well, while ML algorithms may have value for some households (e.g. with LCTs), but at the expense of simplicity, interpretability and ease of implementation.
- **For households without LCTs, simple rule-based algorithms appear to be sufficient** for remuneration with a sufficient level of accuracy.

4. CONTROL GROUP BASELINES

When a large enough portfolio of customers exists, it is possible to create a control group as a counterfactual. Approaches to designing a control group vary - at its simplest, average consumption within a different group of customers can be used. As seen in the "panels method" used by Enedis, a model infers a reference load curve using a control group that has similar characteristics and is expected to behave in a similar manner to the participation group in the absence of a flexibility event.[11] The actual behaviour of the control group provides a baseline from which to measure the flexibility delivered by the participation group.

Figure 6: Summary of control group approach (Enedis "panels method")

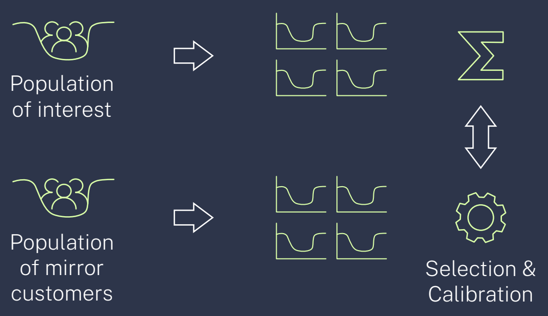
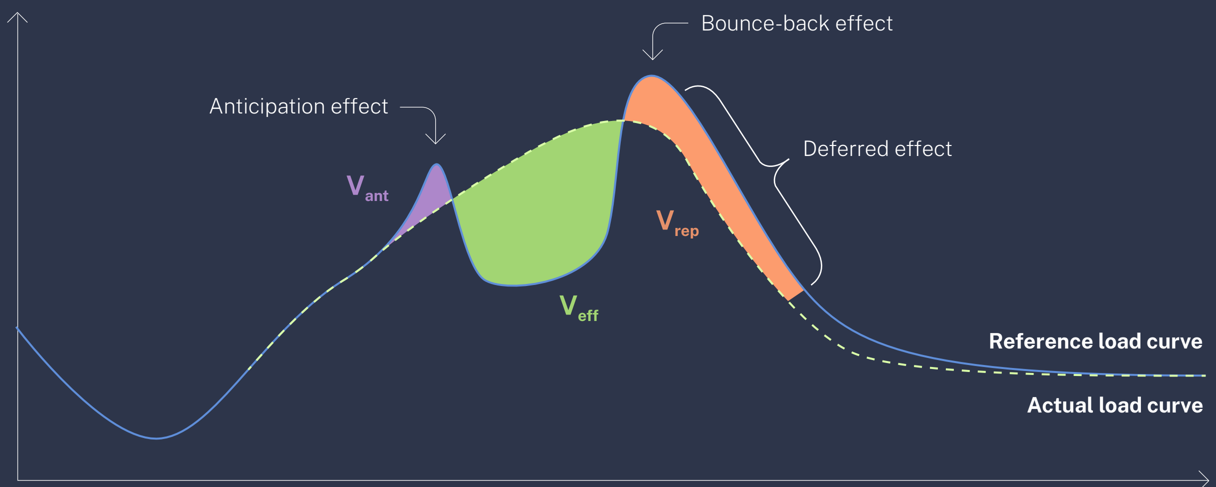


Figure 7



[11] Jean-Baptiste Noël and Olivier Chaouy (2020) Better assessment of load reduction by sampling and machine learning, CIRED - Open Access Proceedings Journal, Volume 2020, Issue 1, 2020

Comparing the estimated reference consumption with the actual consumption of the population of interest enables us to assess its accuracy. This shows a satisfactory average half-hourly time step accuracy of 2-3%. Results appear to be robust, whatever the diversity of days, time of days and populations simulated.

Using control groups to create a baseline also enables us to understand other unintended impacts of participation, such as an increase in load in anticipation of a flexibility event and the “bounce back effect” or “deferred effect” indicating load shift (see Figure 6).

However, there are some challenges to overcome when using this approach:

- **Accounting for bias:** one or several characteristics of the population of interest may be biased compared to the typical distribution of customers. The control group must be carefully designed by introducing the same bias during the construction phase so that it has customers with the same distribution of auxiliary information as in the population of interest.

- **Data limitations for times of day:** where there are regular flexibility events at the same time of day, the data available to train the model may be limited. It therefore needs a way to infer a reference load curve on times of the day it has never learnt from.

- **Contamination by other flexibility events:** it may be difficult to ensure the control group is not taking part in another flexibility event or service, as these increase in future.

- **Gaming at portfolio level:** it may still remain open to gaming if the flexibility service provider selectively sets the control group to exaggerate flexibility. To minimise this risk, it is preferable to identify the control group outside of the aggregator’s control.

There will also be circumstances where a robust control group is not possible to put into operation - for example, if the flexibility service provider does not have access to enough participants. Even where the portfolio is large enough, choosing a subsample of the group of interest to act as a control group has a cost, as it reduces the amount of flexibility that can be provided.

5. NOMINATION BASELINES

There are some unpredictable electricity loads for which historic consumption data, or comparisons with other customers, do not provide enough information to calculate a reliable baseline. For example, some industrial sites work to complex production plans that do not follow a regular daily or weekly pattern. Similarly, loads which are highly controllable, and therefore constantly adapt their consumption schedule (e.g. in response to wholesale prices) may not apply well to baselines which use historic consumption data.

In such cases, the customer may be able to provide a better estimate of the baseline, using their own knowledge of their production plans, plugged-in vehicles or user requests. However, participants are free to set these nomination baselines how they see fit (and the method can change over time). This is a fundamental difference to the approaches discussed above, which include agreed methodologies and results that can be verified by anyone given the same inputs.

Nomination baselines are well established in energy markets, including use in practically all wholesale markets globally, and suit conventional generators well. In Great Britain, they are used in the Balancing Mechanism where large generators submit Physical Notifications (PNs) to the system operator one hour ahead. The principle works well here because there is a sizeable incentive for participants to submit accurate nominations (i.e. not deviate from their PNs unless instructed). Crucially, participants provide PNs constantly so if a unit deviates from its PN this must be justified (e.g. a technical fault) and consistent deviations will result in effective exclusion from the market.

However, if implemented poorly, the subjective nature of nomination baselines can make them open to gaming. For example, in Great Britain wind generators have been found to artificially increase their nominated output when they forecast curtailment actions, to earn greater constraint payments. If actions can be easily forecasted, and therefore “gamed”, another approach may be required -in the example above, a non-meter data baseline of wind speed is being considered.[12]

Where nomination baselines are used, the market operator must continually monitor for baseline manipulation and fraud detection. To aid this, sufficient data must be readily available or appropriate incentives should be in place to submit accurate baselines; e.g. fines or through disqualification from the service.

It is important to note that nomination baselines are not synonymous with schedules dispatched to physical hardware. In cases where asset response can be expected to reflect its schedule, such as conventional generators which serve no additional purpose beyond providing power into the system, it is possible to duplicate schedules as nomination baselines. In other cases, such as assets with more variable loads, there is a stochastic relationship between schedules and behaviour and so this would not be a sensible approach; any regulation requiring schedules to be submitted and adhered to would deter participation, representing poor market design. Nomination baselines can nevertheless work well in both cases.

When designing the timeframe at which participants submit their nomination baselines, it is important to trade-off accuracy against the ability to game. Nominations made closer to real time are likely to be more accurate, as there is less uncertainty in the information available to the participant, but reduced latency also increases the potential for adjusting the baseline for unfair advantage.

[12] National Grid ESO, Power Available (2021)

RECOMMENDATIONS

AGREEING A SET OF BASELINING PRINCIPLES AND GUIDELINES

We believe there is considerable value in identifying and agreeing upon a set of common principles for baselining.

To achieve cross-industry consensus, we recommend **setting up a baselining working group,** made up of both buyers and sellers of flexibility, with a core focus of agreeing on a set of baselining principles across Europe for flexibility service providers, network and system operators.

The working group should, in line with the agreed principles, produce **good practice guidelines** for market operators to follow now, with the aim of establishing a fair playing field and mitigating against gaming risks. It can also **consider potential rules to regulate baselining** in future. For example, some emerging guidelines in this paper are:

- When selecting a baselining methodology for a particular use case or service, there is a need for some consistency and central oversight, rather than having a completely free choice.
- For any baseline methodology, the regulator or market operator must continually monitor for baseline manipulation.
- Baselining should never allow the consumer to influence the baseline once they know they are going to be dispatched. This means in-day adjustments to consumption profile baselines should never fall after notification is given, and a “meter before, meter after” should always set the “meter-before” period before the customer could possibly respond.
- When constructing a control group for use in baselining, it is preferable to identify a population outside of the aggregator’s control.
- Baseline methodologies should be technology-specific given different asset classes require different approaches.

Furthermore, the working group should look to develop a **library of baselines** to guide market participants in identifying suitable baselines. The group could develop processes for adding additional baselines to the library, and determining the conditions and situations for which this is necessary. Doing this collectively avoids the need for each individual market to reinvent the wheel.

This paper calls for a process to begin to agree a consensus around baselining, which should ultimately be incorporated into guidelines and regulations, including the EU Network Code for demand response and the Great British DSO flexibility markets. We will, of course, need to ensure we strike the balance between a consistent approach and allowing for innovation, while also enabling our approach to evolve over time as the evidence base grows.

Use of improved baselining techniques can make participation in flexibility services more transparent, improve asset visibility and lead to more seamless sharing of data -ultimately unlocking flexible energy systems powered by cheaper, cleaner energy.

AREAS FOR FUTURE RESEARCH

Some technical areas that warrant further investigation by the baselining working group are:

Baselining during extreme weather events and at certain times of year, such as "shoulder months", as part of further work to understand behavioural patterns over time.

Removing barriers around participation in multiple flexibility services which may affect baselines, allowing flexibility providers to stack revenue from the various services but requiring greater visibility and flow of data.

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

Identifying methodologies and use cases for technology-specific baselines.

Workable approaches to use when there is insufficient or no historical data.

Accounting for energy recovery in baseline methodologies, where relevant.

Potential thresholds for minimum aggregation levels in cases where individual baselines are insufficiently accurate, but only aggregated data is needed for settlement.

Measures to ensure adequate data is available for regulators and market operators to establish best practice (e.g. requiring constant submission of nominations not just during activation periods).

Approaches to validating consumption profile baselines, including the use of control groups.

Centre for Net Zero
5th Floor, UK House, 2 Great Titchfield Street
London W1D 1NN

info@centrefornetzero.org
@CentreNetZero
company/centre-for-net-zero
centrefornetzero.org

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