DNV·GL

Baseline Methodology Assessment

Energy Networks Association

DNV GL doc No: 1025166001
Date of first issue: 2020-12-01
Date of this revision: 2020-12-21

Customer Details

Customer Name:	Energy Networks Association
Customer Address:	4 More London Riverside, London SE1 2AU
Customer Reference:	ON20-WS1AP7-ITT Document for baseline
Contact Person:	Courtney Madden

DNV GL Company Details

DNV GL Legal Entity:	DNV GL Limited
DNV GL Organisation Unit:	Energy Markets & Technology (EMT)
DNV GL Address:	Vivo Building, 4th Floor, 30 Stamford Street, London SE1 9LQ
DNV GL Telephone:	0203 816 4000
DNV GL org. No.:	1503799
DNV GL doc. No. :	1025166001

About this document

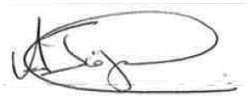
Proposal Title:	Baseline Methodology Assessment
Date of first issue:	2020-12-01
Date of this revision:	2020-12-21

Revision History

Revision	Issue Date	Comments
V0.2	01/12/2020	Draft report
V1.0	11/12/2020	Draft report feedback from ENA ONP
V2.0	21/12/2020	Final Report

for DNV GL Limited

Prepared by:



Angeliki Gkogka
Consultant – Energy Markets & Technology

Approved by:

Hans de Heer
Principal Consultant – Energy Markets & Technology

Table of contents

EXECUTIVE SUMMARY

To date, GB Distribution Network Operators (DNOs) have each developed their own methods to deploy flexibility for distribution network constraint management purposes. As part of such methods, baseline methodologies are an important instrument to validate and quantify the service delivered by a flexibility service provider (FSP). In principle, a baseline is an approximation of the energy consumption or generation that would have occurred if no flexibility activation has been triggered. The ENA Open Networks Project (ENA ONP) defines the baseline as the "Established level of Distributed Energy Resources base load from which a delta is measured to calculate level of service delivered."

DNOs are so far applying different methodologies to establish these baselines. Although DNOs' experience in operating flexibility is still in its infancy, alignment (or even standardisation) of baseline methodologies will benefit both FSPs and the DNOs, by removing barriers for FSPs and providing DNO confidence in the efficacy of flexibility, as well as encouraging greater third-party engagement in the Distribution System Operation (DSO) flexibility markets.

This report reviews all the known baseline methodologies for application in the ENA DNO products: Sustain, Secure Scheduled, Secure Dispatched with week-ahead utilisation instruction, Secure Dispatched with real-time utilisation instruction, Dynamic and Restore.

One of the main tasks for the assessment of baseline methodologies was to understand and agree the ENA DNO product parameters, as these have been developed in the recent years and through ENA ONP. Product parameters have a large impact on the choice of the most suitable baseline methodologies, which is a theme that stakeholders highlighted in the ENA ONP 2020 Flexibility Consultation. Through market assessment and review of existing GB and international baselines, we find that GB DNO Flexibility Products involve some specific parameters that are not widely seen internationally:

- Sustain and Secure Scheduled have very long utilisation instruction notification periods[1] which do not exist elsewhere in the GB and international practices. As such, we could not identify any best practices for these products.
- Real time products (i.e. Dynamic, Restore and, Secure Dispatched – real time), show similarities with balancing products and adequacy products on the notification instruction timing. However, the utilisation period[2] of DSO products is notably longer than what we see in most balancing and adequacy products.

In light of limited GB and international best practices that fit the GB DNO flexibility products, we formed our recommendations mainly based on our experience with flexibility markets and the design fit of the flexibility products, as well as our experience with baseline design in general.

We also took into consideration the current status of GB DSO flexibility markets and the current priorities of DNOs and FSPs, informed by responses to the ENA's public consultation as well as further bilateral stakeholder engagement as part of this assessment. Simplicity and inclusivity of flexibility markets are key priorities at this stage, so that wider market participation in DSO services is facilitated . We also took note that several industry stakeholders, as well as Ofgem and BEIS, have called for consistency across DNO products, ESO products and the wider industry.

On this basis, our recommendations focus on three types of baselining methodologies that are relatively simple, are known in GB markets, and which are currently in use by DNOs and/or in ESO balancing

1 The time period before a Flexibility Service is required by a DNO, in which a DNO may issue a Utilisation Instruction to a Flexibility Provider for the provision of a Flexibility Service (ENA ONP WS1A P3 definition)

2 The time period during which the Flexibility Service Provider provides their service/activates flexibility.

services and/or in the Balancing Mechanism. To facilitate technology-agnostic solutions and allow FSPs to agree a baseline with the DNO based on their needs and types of asset, we recommend an agile baseline methodology solution: Historical baselines (with or without Same Day Adjustments (SDAs))[3] could be used as the default baseline, whilst allowing for all technologies to use a nomination-type baseline, where the historical baseline is not sufficiently accurate or where this is not the preferred option for the FSP.

Table 1: Final Baseline Recommendations for ENA DNO Products

Product	Main recommendations
Sustain and Secure Scheduled	More experience in operating flexibility services needs to be gained by all DNOs before moving to the standardisation of the validation process (including baselines, if applicable). For future standardisation efforts we recommend testing the following technology-specific validation mechanisms, which consist our interim recommendations: • For dispatchable generation assets we recommend setting the baseline to 0. The asset is required to provide capacity when needed, regardless of how it would have otherwise dispatched. To remunerate the activation, we recommend the flexibility DER assets would be compensated for the hours of *delivery x capacity.* • For non-dispatchable generation, we recommend the same approach using de-rating factors. • For load assets, we recommend using a historical baseline which uses data from the previous year, i.e. Mid 8 of 10 for weekdays, mid 2 of 4 for weekends.
Secure Dispatched (week-ahead)	Historical baseline without SDA Mid 8 of 10 for weekdays, mid 2 of 4 for weekends. [4] Nomination to be used for • dispatchable generation • connections with dominant dispatchable generation • if accuracy levels of historical baselines are (too) low
Secure Dispatched (real-time), Dynamic and Restore	Historical baseline with SDA Mid 8 of 10 for weekdays, mid 2 of 4 for weekends. [4] Nomination to be used for • dispatchable generation • connections with dominant dispatchable generation • if accuracy levels of historical baselines are (too) low

[3] For more information on different baseline methodologies and their key characteristics please refer to section 2: Introduction to baseline methodologies of the main document.

[4] To address extreme loads, the mid 8 of 10, excludes extreme loads from both top and bottom load days. (e.g. out of 10 days which are selected for the baseline, the top and the bottom load days are excluded). For more information on different baseline methodologies and their key characteristics please refer to section 2: Introduction to baseline methodologies of the main document.

For more information on the detailed recommendations for each product please refer to section 5.3: Analysis of baselines per product

Summarising our recommendations for different technology types we have the following observations:

Table 2: Recommendations for technology types

Technology	Recommendations
• **Generation**: Both dispatchable and non-dispatchable, fossil-based and renewables	We recommend nomination baselines for both dispatchable and non-dispatchable generation. These recommendations do not apply to Sustain and Secure Scheduled products.
• **Demand (load)**: Electricity consumption	We recommend historical baselines if this provides sufficient accuracy. Otherwise nomination baselines can be used.
• **Storage**: Mainly batteries	
• **Any combination of the above**	

The recommendations provide the basis of the baselining exercise which is undertaken by the Workstream 1A (WS1A) Product 7 (P7) of the ENA ONP. DNOs should further consider their implementation and explore the applicability of the baselines to their products as more learnings and insights become available to them from operating flexibility products.

We recommend the following next steps:

1) For the Sustain and Secure Scheduled products, we recommend DNOs to further gain experience prior to standardising these products.
2) Finetune the baseline methodology against the product design, if the latter evolves over time.
3) As our assessment of the baseline methodologies has been purely qualitative, we recommend enhancing this study by calculating the variance and bias for the proposed historical baselines for a large, representative set of assets.

In the coming years, a greater number of new technologies are expected to seek participation in DSO products. We recommend assessing the suitability of the chosen baseline methodologies for a new technology when a provider wishes to include this technology in its portfolio. If the current baseline methodologies prove not sufficiently inclusive, then a technology-specific baseline methodology can be trialled for this technology and added to the list of available / allowed baselines methodologies. As this will be a developing process, we recommend that the DNOs develop an appropriate governance strategy for changing/augmenting their baselining methodologies.

1 INTRODUCTION

1.1 Context

All Great Britain (GB) Distribution Network Operators (DNOs) have agreed under Open Networks project (ONP) to seek alignment in the development of flexibility products and their processes such as procurement, dispatch and settlement. ENA's project initiation document (PID) 2020 has also indicated that DNOs agreed to seek alignment in the area of baseline methodologies for flexibility products.

To date, GB Distribution Network Operators (DNOs) have each developed their own methods to deploy flexibility for grid management purposes. Although DNOs' experience in operating flexibility is still in its infancy, alignment (or even standardisation) of baseline methodologies will benefit both FSPs and the DNOs, by removing barriers for FSPs and providing DNO confidence in the efficacy of flexibility, as well as encouraging greater third-party engagement in the Distribution System Operation (DSO) flexibility markets. DNV GL considers this work crucial in providing confidence to energy industry stakeholders in the future commercial viability of flexibility services, which will accelerate the transition to a smart, flexible energy system.

1.2 Scope of work

This work sits under Workstream 1A (WS1A) and Product 7 (P7) of ENA ONP. The objective of the product is to undertake assessment of current work in the industry that relates to the development of a baseline methodology for measuring delivery of services through third party DER and recommend further actions that DNOs could take to achieve a consistent approach.

DNV GL was appointed by ENA ONP to perform the analysis and assessment of existing baseline methodologies against ENA DNO products with the goal of developing recommendations and an alignment approach for ENA ONP to consider.

1.2.1 Purpose of this report

This report was developed by DNV GL and presents the analysis and assessment of existing baseline methodologies under a number of assessment criteria that DNV GL and WS1A P7 product team developed during the course of the project. The purpose of this report is to provide recommendations and next steps for DNOs to further develop aligned baseline methodologies.

This report accompanies ENA ONP's report of WS1A P7 and includes the detailed assessment and considerations that informed the final recommendations.

1.2.2 Structure of this report

Table 1 summarises the structure of this report.

Table 1: Structure of the report

Section 1 – Introduction Provides a summary of the context of this work, the purpose of this report and ENA ONP's inputs.
Section 2 – Introduction to baseline methodologies Provides an introduction to baseline methodologies and their key characteristics.
Section 3 – Consultation and bilateral engagement Provides a summary of the ENA's 2020 Flex. Consultation responses which are relevant to baseline and our insights of bilateral engagement with key stakeholders.

Section 4 – Market Assessment Provides an overview of the standardisation of ENA DSO products and maps existing baseline practices against ENA DSO products.
Section 5 – Key outcomes Describes the key outcomes of our assessment and recommendations for baseline methodologies for the ENA DSO products.
Section 6 - Implementation of baselines Provides practical recommendation on the implementation of baselines with regard to stacking, information exchange as well as monitoring and validation.
Section 7 - Recommendations and next steps Summarises the recommendations and next steps for GB DNOs.

In the report, we have highlighted our recommendations in blue.

1.3 ENA Open Networks Input

To perform our analysis and assessment we worked in close collaboration with GB DNOs and ENA ONP team. We received regular feedback to our outcomes and we also informed our assessment by developments under ENA ONP Workstream 1A, which we have listed below:

- We reviewed the responses to ENA ONP's 2020 Flexibility Consultation to inform our assessment and understanding of stakeholders' priorities and requirements. In addition, based on stakeholders' responses we identified stakeholders for bilateral engagement to further inform our work (section 3).
- We engaged with WS1A P3 team to capture developments of the standardisation of services parameters (sections 4.1 and 4.2).
- ENA ONP WS1A P7 team developed the baseline principles that the baseline methodologies are expected to meet and should be assessed against. Table 2 summarised the baseline principles which form the basis of our assessment criteria.
- The product team then considered what other factors should be included for the scoring and assessment of potential methodologies. Table 3 summarises the scorable sub-factors that also informed the development of our assessment criteria.

Table 2: Principles for baseline methodologies

1	**SIMPLICITY**	The solution is practical and the effort required is proportionate to the outcome.
2	**ACCURACY**	The baseline methodology provides a good representation of the counterfactual behaviour of the provider.
3	**INTEGRITY**	Restricts ability for the DER provider to distort or game the market.
4	**REPLICABILITY**	Can be replicated for forecasting and verification by all relevant parties.

Table 3: Scoreable Sub-factors for baseline methodologies

Inclusive	Is not biased to a particular type of response
Design fit	Meets the needs of the DNO product design specifics.
Stackability	Allows participation in other markets

2 INTRODUCTION TO BASELINE METHODOLOGIES

In principle, a baseline is an approximation of the energy consumption or generation that would have occurred if no flexibility activation has been triggered. The ENA Open Networks Project (ENA ONP) defines the baseline as the "Established level of Distributed Energy Resources base load from which a delta is measured to calculate level of service delivered."

Baselines are grouped by some key characteristics such as the type of data and the estimation method which generally governs how the selected data is used to calculate a baseline for a dispatch day. Typically, a baseline is a load/generation profile with the same time resolution as the product. Metering requirements are set by the product for validating the service delivery, where the resolution should correspond with the nature of the product. If e.g. the violation of thermal limits can lead to a black-out within 5 minutes, a 30-minute interval will not be effective. Since baselines are used for service validation, it should therefore follow the resolution of the product. We acknowledge that some of the GB DNOs use a flat baseline which is the same across the whole utilisation period. Although some flat baselines are considered in this study (e.g. Meter Before-Meter After, see Table 4), we recommend (e.g. for historical baselines) not to apply flat baselines.

Our assessment reviewed all the known baseline types which are used worldwide. Table 4 summarises these baseline methodologies and their characteristics.

Table 4: Baseline methodology group

Methodology group	Characteristics
Meter Before – Meter After (MBMA)	MBMA is a flat baseline set at a pre-activation level determined by one or a small number of intervals (settlements) which is consistent with no delays. The baseline can be the single meter reading or average/median/min/max of meter readings before the activation window. Metering readings of the activation window are compared against the meter readings prior to the activation to calculate the delivered flexibility. The Meter Before/Meter After is a widely used methodology for accurately estimating the level of service delivered under real-time dispatch conditions and short utilisation periods. It is also a preferred baseline for frequent activations as it is not dependent on historical data. As such, MBMAs are very common in balancing products which have these characteristics. Meter Before/Meter After requires DERs with relatively flat load profiles during the utilisation period. If a resource has periods of ramping up or down or general variability, the meter Before/Meter After approach can over- or under-estimate the actual level of load reduction even for the shorten period. MBMA requires sub-metering at DER asset level. **Data:** MBMA baselines have extremely limited data selection rules, essentially a single interval. **Estimation method:** No estimation method is used for MBMA baselines, as they are a flat line at the prescribed load level. Single meter reading is sufficient.
Window before and after (WBA)	This is similar to the MBMA. In this case the window after is also taken into account to compensate for direct rebound effects. It has same characteristics

Methodology group	Characteristics
	and applications to the MBMA. **Data:** This baseline is not dependent on historical data or external data (e.g. weather data) **Estimation method:** No estimation method is used.
Nomination	The baseline is the forecast of the generation or demand profile of the DER asset, as if no flexibility utilisation would take place. The forecast is sent by the Flexibility Service Provider (FSP) to the DNO before gate closure or at another predefined deadline. For example, the physical notifications which are used in the Balancing Mechanism belong to the Nomination baseline types. The DNO can then use this profile to calculate the deviation of the metered data from the planned profile. The FSPs can use different methods to forecast their profile. The nomination is a technology agnostic baseline for real-time utilisation instruction. In addition, as per stakeholders' response it works well when sub-metering is available, but it creates challenges and inaccurate forecasts when sub-metering is not available. **Data:** There are no data requirements. The FSP can select if they want to use any historic data in forecasting their demand/ generation. **Estimation method:** The FSP can choose their forecasting approach. The estimation method is not applicable to this baseline.
Historical baseline (a.k.a. rolling baselines)	This baseline methodology uses historical data to calculate the baseline, usually based on recent data prior to the utilisation day. For example, the 'average interval load of last 5 business days' can be the baseline from which the deviation is calculated. **Data:** Two important aspects of the historical baselines are the data selection and exclusion rules which determine what data will be used to produce the baseline. Historical baselines require usually 5 to 10 days of data. In addition, historical baselines exclude prior-event days and non-similar days (e.g. weekends and holidays). In addition to the exclusion of prior dispatch and non-similar days, many historical baselines have an additional set of data exclusions based on load characteristics. For example, a baseline may drop low load days if they fall below a threshold related to the mean of the selected days (e.g. the low load day is less than 20% of the mean load during the selected days during dispatch window hours). Other baselines rank the included days based on load and exclude a subset of those chosen days, either extreme days (e.g. mid 8 of 10) or just low-load (high 4 of 5). For example, mid 8 of 10 baselines excludes the highest and lowest demand days of the 10 selected and high 4 of 5 excludes the lowest demand day. These exclusions are designed to target

Methodology group	Characteristics
	the recent days that are most likely to predict the dispatch day load. **Estimation method:** Historical baselines are primarily calculated using a simple mean for each interval across the final set of chosen days. A median approach is also used by some historical baselines. DNOs usually provide the baseline which is based on historical data to the FSPs. DNOs use metered historical data and FSPs can replicate the baseline using their own data.
Regression-based	A regression model is used to calculate the baseline. Regression-based baselines use schedule, weather and other variables to explain customer load variability. The regression summarises how load interacts with the variables allowing for the prediction of load levels at any combination of the variables. The baseline is the regression-based predicted load for the dispatch day based on that day's weather and schedule characteristics. **Data:** Regression-based baselines require substantially more data than historical baselines. The make up for the lack of recent data by using advance methods to identify the relevant load characteristics from a large pool of data. Most regression baselines required at last a full year of data. Regression approaches use all of the available data but control for excluded days in the regression structure. In the regression context, controlling for weekends days removed their effect on the weekday baselines while informing the weekend baseline. **Estimation method:** Regression estimation methods are substantially more complicated than the mean used by historical baseline approaches. For example, a regression baseline may include calendar, weather and daylight variables in multiple forms. The specification controls for heat build-up over days, heat gain within the day, hour of light and fraction of dark as well as a range of temperature-time interactions. The regression uses this specification to develop a baseline that reflects the calendar, weather and daylight characteristics of the dispatch day. The regression produces both weekday and weekend baselines. Regression-based methodologies have the potential to be the most effective baseline without same-day adjustment (SDA) (*see below for explanation on same-day adjustments*). The difficulty of developing an accurate baseline without the SDA is a primary potential justification of the additional technical challenge of the regression approach.

Methodology group	Characteristics
Calculated baseline (e.g. weather-based)	This method involves calculation based on external parameters, without relying on historical data. These baselines are not very common as they are applicable to technologies and assets with a demand or generation profile which can be calculated based on formulas and inputs which do not use historical data. Some examples are as follows: • Generation of a wind turbine base on wind speed and capacity. • Generation of a solar PV based on solar panels characteristics, radiation, sunlight weather information etc.
Control group (a.k.a. peer group)	The baseline is calculated using as inputs measurements of similar customers who do not participate in the flexibility service. **Data:** This baseline methodology uses data of assets/ technologies/ customers which are of the same type as the participating DER technologies. The main requirement is that these assets do not participate in flexibility products. The methodology uses data of the same utilisation period as the flexibility event. **Estimation method:** The baseline is calculated based on a simple estimation method, using a simple mean or median.

In addition to baseline methodologies of Table 4, we also see a variant of historical and regression-based baselines which includes a same-day adjustment of the baseline. The following section describes the key characteristics and methods of the same-day adjustments.

Same-Adjustment (SDA) Method and Period

Adjustment method and period are additional components of historical and regression baselines. Adjustment method refers to adjustments that can be made to the initial baseline to make it a better fit for the dispatch day load/generation. Adjustments use the most up-to-date information to inform the final position of the baseline. The adjustments bring the baseline into line with the pre-dispatch intervals on which the adjustment is based. This means that the baseline starts the dispatch period relatively close to actual load and will only diverge if the load shape from the baseline is different from the actual profile of the day.

Adjustment period refers to the specific intervals that are used to make the adjustment.

Figure 1 visualises how this mechanism works in practice. A customer with a weather-sensitive load profile is shown, with the meter data displayed in green. A historical baseline calculation (in this example a mid 3 of 10, displayed in blue) without same-day adjustment is constantly below the measured values (probably the current day is much colder than the preceding days). The adjustment window is set to 8:00AM – 12:00AM, during which the average (relative or absolute) difference between the baseline without SDA and measurements is determined. During the DR event (utilization period), the baseline is adjusted according to this difference. This results in the adjusted baseline (in pink), which, based on a visual inspection, approaches the counterfactual more closely. This has direct influence on product validation, as shown by the difference in calculated load drop, which corresponds to the activated energy.

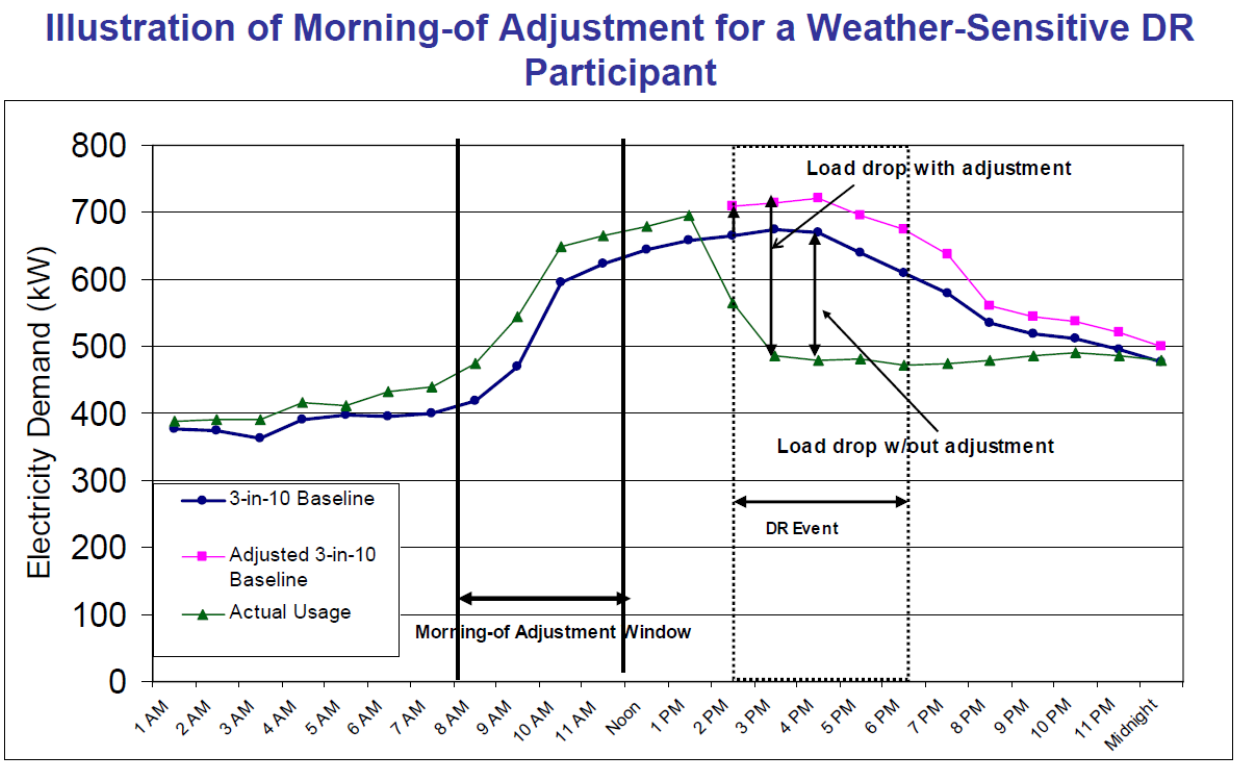


Figure 1: Visualisation of Same Day Adjustment mechanism

In simulations and based on our experience, the evidence is conclusive that same-day load-based adjustments improve the accuracy of baselines. Load-based adjustments also carry a degree of risk because the adjustment process relies on a handful of load intervals to adjust the whole baseline. SDAs can be more susceptible to gaming as few intervals can have a big effect on the baseline. Beyond gaming, there are reasonable pre-dispatch load characteristic that may also affect the baseline in ways that would make the baseline less accurate for both system and customers. For these reasons adjustment approaches that do not use pre-dispatch load are also available.

Additive and Scalar same-day load adjustments

Adjustments are made either using an additive or scalar approach. Both approaches have the effect of bringing the baseline in line with actual load on average, during the pre-dispatch adjustment period. During the dispatch period the adjustments differ:

- For the additive adjustment, the baseline is shifted by this same amount for each interval. For example, the difference between the measurement and the baseline value in the adjustment period is added to the baseline value.
- In contrast, for the scalar approach the magnitude of the adjustment will vary depending on the level of baseline load to which the percentage adjustment is applied. For example, the baseline is multiplied by the ratio d/b, where d is the measurement in the adjustment period and b the baseline value for the adjustment period.

Drop-to mechanism

Although baseline methodologies are suitable for most of the flexibility products, other mechanisms have also been used, such as the **drop-to mechanism** which is used to validate that a flexibility asset has provided the contracted capacity. The drop-to mechanism also known as *firm load* in case of demand

response, identifies load levels below which an asset will stay during an activation period (Figure 2). Rather than attempting to explicitly measure load reduction from a counterfactual baseline, the drop-to mechanism assures that an asset is not contributing to system load beyond the specified amount. The key characteristics of a drop-to mechanism are:

- Drop-to typically is used for large customers to address local emergency conditions.
- Drop-to is suitable for capacity-only products. When energy is remunerated, the activated energy needs to be calculated for which the drop-to mechanism is not suitable.
- Drop-to mechanisms generally do not, on their own, support a quantified settlement process. Ex post application of drop-from baselines or regression techniques can provide estimates of the actual load reduction supplied at the time of activation. Similarly, a deemed load reduction value can be calculated as the average of customer load net of the drop to value for relevant historical intervals.

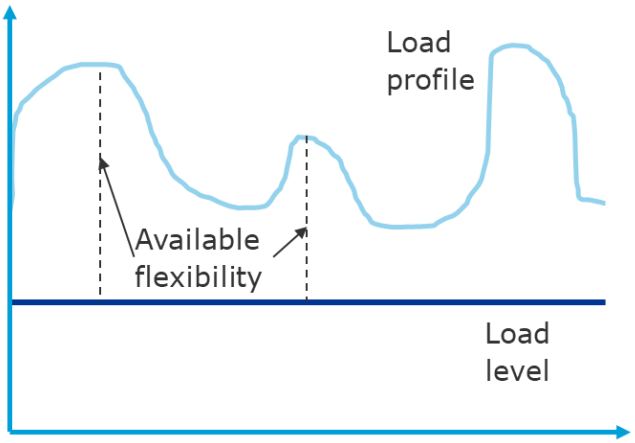


Figure 2 Drop-to mechanism example

3 CONSULTATION AND BILATERAL STAKEHOLDER ENGAGEMENT

This section summarises our insights and key messages which we derived from the 2020 Flexibility Consultation and the bilateral engagement that we performed during September and October 2020.

The Flexibility Consultation 2020 from the ENA Open Networks closed in September 2020. Following the closure, we worked with ENA ONP product team in processing, analysing and reporting on consultation responses to identify their impact on WS1A P7.

We also engaged directly with the relevant stakeholders to ensure that we have a full understanding of their responses, as well as to ensure that the project captures and benefits from stakeholders' expertise in full. We engaged bilaterally with four stakeholders and we also facilitated a workshop with the Association of Decentralised Energy (ADE) members.

ENA has already published their response to the consultation answers, so we will not repeat existing outcomes. We are presenting below the key messages and observations from stakeholder that we considered in our assessment approach and recommendations, including the outcomes of both the consultation and our bilateral engagement.

Table 5: **DNV GL's approach to stakeholders' responses**

Stakeholders said…	DNV GL did…
Give consideration to all methodology types, not just historical data.	We worked with DNOs and stakeholders (through bilateral engagement) to understand the new technologies that participate in flexibility services and faced challenges related to baselines. For example, bilateral engagement showed that historical baselines will not work for certain new-built assets, while nomination baselines submitted 1-hour ahead will create challenges for assets with variable load/generation. We considered these challenges in our assessment and final recommendation, suggesting an agile baseline methodology solution.
Lack of simple and clear baselines;	Simplicity was set as one of the priorities for assessing baselines. Simplicity and consistency are both reflected in our final recommendations to ENA ONP.
Challenges driven by metering configurations	We talked with stakeholder to understand their challenges and potential solutions. Our engagement showed that, not having the possibility to apply submetering[5] for service validation, will be a challenge for certain baselines, including nomination baselines. Following engagement with DNOs, we have clarified that sub-metering will be allowed in all DSO products, if this is preferable to the FSP.
Current processes do not support revenue stacking	In this report in section 6.1 we explain how stackability has been considered in our recommendations and we provide guidance on how recommended baselines will facilitate revenue stacking.
Simplicity and Inclusivity are key priorities	Simplicity and inclusivity are the highest weighting criteria assessment.
Alignment across the industry/ESO	We engaged bilaterally with ELEXON to discuss current developments of relevant modifications (P376, P375) and lessons learnt. We also look into new ESO products and baseline methodologies to identify applicability to DSO products. Our recommendations are also considering alignment among ENA DNO products, where this is possible.

[5] Typically, all sites have a main meter at boundary level which meters the flow of all assets present at this site (Meter Point Administration Number – MPAN). Sub-metering is the concept where there is a meter at the asset level which measures the power flow related to the specific asset only. .

Stakeholders said…	DNV GL did…
Future-proofing methodologies	Our current recommendations follow a simple approach which allows FSPs to agree different technologies to agree their baseline with the DNO. In the future and as more baselines are added, this can be revised.
Consultation should be undertaken with a wide range of stakeholders both before and after recommendations are finalised.	Stakeholder engagement has been ongoing through ENA's Advisory Group. Feedback received will feed directly into the product recommendations. We engaged bilaterally with 4 different stakeholders. We run a workshop with the Association of Decentralised Energy which was attended by a large number of stakeholders.
Be open and transparent	Our approach and recommendations are available to everyone through this report.
Consider susceptibility to gaming and accessibility	Integrity is one of the assessment criteria. We considered accessibility through our criteria of "Inclusivity", "Replicability", "Robustness to data". (see section 5.1)
Our bilateral engagement with stakeholders demonstrated that few best practices/precedents are available for GB DSO products, particularly where long utilisations and instruction notifications are concerned	When we engaged with ENA's stakeholders (including the ADE workshop), we asked them for international practices and examples that could further expand our database. We were particularly interested in products with long utilisations and long instruction notification period. Our engagement confirmed the outcomes of our market assessment that GB and international practices are limited and may not be suitable examples for GB DSO products. As such, we based our recommendation not only on best practices but mainly on our experience and the design fit for the products.
Controversial views on nomination baselines. Some stakeholders are strongly in favour of nomination methodologies on the basis that nomination baselines can be more inclusive for certain assets especially when there is lack of historical data or for variable loads. Other stakeholders highlighted that nomination baselines are technology-dependent, working well for scheduled generators and controlled industrial processes but not for most other asset types.	We took both views into considerations when we assessed and recommended baseline methodologies. Section 5 includes the details of our approach. Our final recommendations are agile indicating that a nomination baseline can be used when historical baselines are not sufficiently accurate and if nomination is preferred for certain assets by FSPs.
The role of residential portfolios and their profile should be considered. DNO products currently use a flat baseline of behaviour over a month, which is challenging when it comes to domestic users, whose behaviour tends to vary.	DSO products are open to all technologies and user segments. Our agile approach developed a view on how inclusivity for all technologies can be achieved, without making the baseline methodology overly complex at this point in time.

4 MARKET ASSESSMENT

This section describes key outcomes of our analysis of ENA DSO products and mapping GB and international practices against these products. In this section we also explain the impact of standardisation of ENA DSO products as well as of current and future technologies in the assessment of baseline methodologies.

4.1 ENA DSO Products

One of the main tasks for the assessment of baseline methodologies is to understand and agree the ENA DNO product parameters, as our final recommendations related to the ENA DSO products. Product parameters have a large impact on the choice of the right baseline methodologies, which is a theme that stakeholders highlighted in the ENA ONP 2020 Flexibility Consultation. The parameters with the largest impact on baseline assessment are:

- Metering configuration;
- Type of remuneration;
- Utilisation instruction notification period;
- Utilisation period;
- Frequency of use of the flexibility product;
- User and technology segment;
- Pool or asset level activation; and
- Controller of the assets (i.e. DSO or the FSP)

Following discussions with DNOs and WS1A P3 representatives we finalised the products and the parameters that we are going to assess against the baselines, and we have summarised them in Table 6. The table does not show the following parameters as these were common across all products:

- Metering configuration: Sub-metering or MPAN metering will be allowed in all products, depending on FSP's preferences. These parameters are not standardised yet.
- User segment and technologies: ENA DSO products are open to all user segments and technologies
- Pool or asset level activation: Pool activation is possible across all products.
- Controller of the assets (i.e. DSO or the FSP): The FSP activates the flexible resources in all products.

To perform our assessment, we grouped the products under 4 baseline product categories mainly based on the utilisation instruction notification period. We also considered the utilisation period and the frequency of use. Where there was not alignment in these parameters, notification period took precedence.

As such, we split Secure Dispatched product to 2 baseline product categories on the basis of different utilisation notification periods (i.e. week ahead and real time) and we grouped Sustain and Secure Scheduled.

Table 6 summarises the ENA DSO products, their parameters and their baseline product categories.

Table 6: ENA DSO products and key design parameters

Category	Product	When required?	Metering Resolution	Type of remuneration	Earliest utilisation instruction notification	Latest utilisation instruction notification	Typical utilisation period	Frequency of use
A	Sustain	Scheduled forecast overload	HH metering	Utilisation and Availability	Scheduled in advance Years ahead	3 months ahead	Not defined, typically 3 to 24 hours	High: 5 deployments per week
A	Secure Scheduled	Pre-fault and peak shaving	metering requirements vary across DNOs		Contract stage	3 months ahead	not defined, typically 3 hours or more	Medium: 2 deployments per week
B	Secure Dispatched (week-ahead)				10 days ahead	3 days ahead		
C	Secure Dispatched (real-time)				30 minutes	Real Time		
D	Dynamic	Network abnormality or planned outage			15 minutes	Real time	Not defined, Typically several hours (it could take up to days)	Rarely, in case of faults
D	Restore	Network abnormality		Utilisation (it can also be availability only, it depends on DNOs)	15 minutes	Real Time	Not defined, Several hours to days, minimum 3 hours	Rarely, in case of complete loss of supply

4.2 Standardisation of ENA DSO products

The standardisation of ENA DSO products and their parameters has a large impact on the assessment and consistency of baseline methodologies across DNOs. We acknowledge that the standardisation of ENA DSO products is work in progress under WS1A P3.

ENA WS1A P3 is working towards standardisation of 10 product parameters, which do not include all the parameters which relate to the baseline methodology. As such, there are still uncertainties around product parameters and how DNOs will use different products.

The main uncertainties that are relevant for the selection of baseline methodologies are:

- Metering configuration: Although DNOs mentioned that sub-metering and Meter Point Administration Number (MPAN) metering will be both an option for the FSP, the parameter is not standardised yet and sub-metering requirements are not defined. For our assessment, we assumed that sub-metering will be allowed. The use or not of sub-meters affects the baseline methodology as some baselines (i.e. nomination) are not appropriate for MPAN configurations.
- Metering resolution: Metering resolution varies across DNOs. Metering resolution should be considered when designing a baseline methodology as it relates to aspects such as data exchange requirements, accuracy of a baseline, replicability and stackability of flexibility services. By default the time resolution of the baseline should align with the metering resolution requirements of the product (please refer to section 2 for further explanation on this topic).
- Utilisation instruction notification period: DNOs use different notification periods for the Sustain and Secure products. The lack of standardisation adds complexity when assessing baseline methodologies and hinders the development of a consistent baseline methodology across DNOs. For example, the notification periods of the Secure product vary from months to real-time, while the utilisation instruction of the Sustain product takes place at contract stage, which is also different across DNOs.
- (Maximum & Typical) Utilisation period: The typical and maximum utilisation period also vary across DNOs and are not part of the standardised product parameters. Utilisation period is particularly relevant when we assess the accuracy of baselines and the stackability of products.
- We have noticed that DNOs use the existing products for different reasons and under different circumstances on the grounds that GB DNO flexibility services are at their infancy. We expect that as DNOs gain more experience with the operation of flexibility services, then further alignment and consistency will be achieved.

To manage these uncertainties, we used the parameters of Table 6 which we agreed with the DNOs and WS1A P3 representatives. Our recommendations reflect these parameters; nevertheless, our assessment approach took into consideration that further changes will take place in the future. As such we recommend a simple yet robust and future-proof approach to baselining.

4.3 Mapping GB and international practices against ENA DSO products

To map GB and international practices against ENA DSO products, we had first to review their product parameters and identify similarities with ENA DSO products.

We compiled a data base containing 90 different products from 9 countries: GB, France, Switzerland, Finland, Netherlands, Belgium, USA, Canada and Australia. Since DSO constraint management is quite a novel service, the vast majority of the services that we included in our analysis are balancing, adequacy or wholesale services.

Figure 3 summarises the type of products per country (and state) that we reviewed and shows that our list included only 1 DSO Constraint Management product

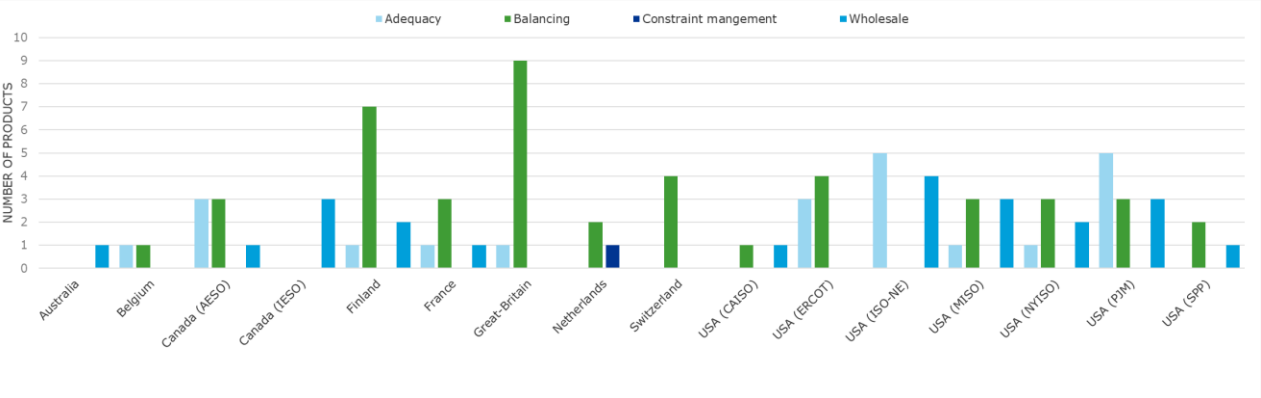


Figure 3: GB and International products grouped by market

To map the DSO services into the international products, we looked into two product parameters: utilisation instruction notification and maximum/typical utilisation period.

Figure 4 shows that:

- Wholesale products have the longest utilisation instruction notification periods, generally between day-ahead and few hours ahead.
- Adequacy products typically receive their dispatch notification between 1 and 3 hours before utilisation.
- Balancing products have the shortest utilisation instruction notification periods, typically ranging from 15 minutes to seconds.

With regard to the utilisation period, our analysis shows the following outcomes:

- Most balancing products show short utilisation periods, typically ranging from 15 minutes to 1 hour.
- Adequacy product show utilisation periods of the order of hours, however, in practice these services are generally called upon for a duration of less than 2 hours.
- Wholesale's utilisation periods are in general as scheduled or dispatched. Although some of them can take from minutes to several hours, the most common utilisation period is less than 1 hour[6].

[6] In wholesale market-based solutions activation is determined by a bid clearing price in the real-time energy market.

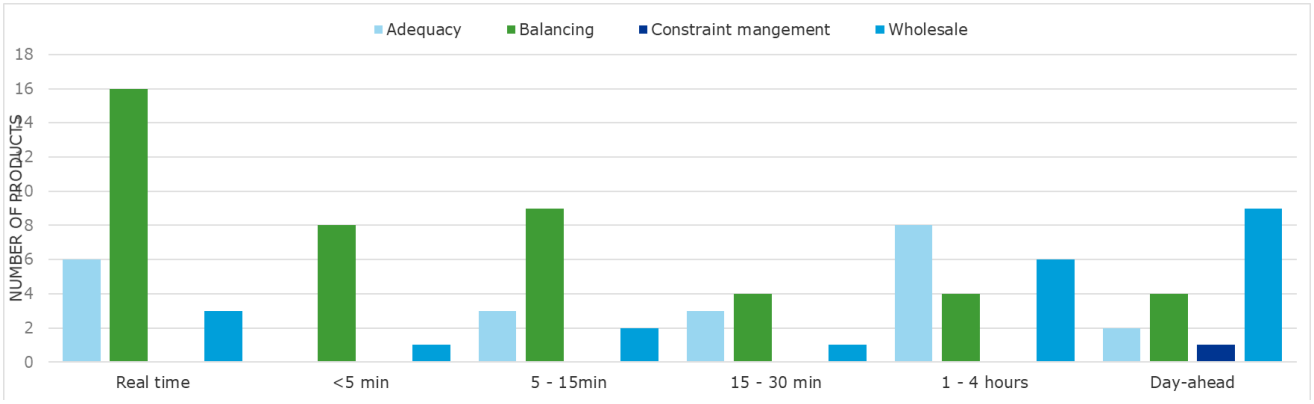


Figure 4: Utilisation instruction notification

Our mapping exercise of ENA DNO products against GB and International products show that:

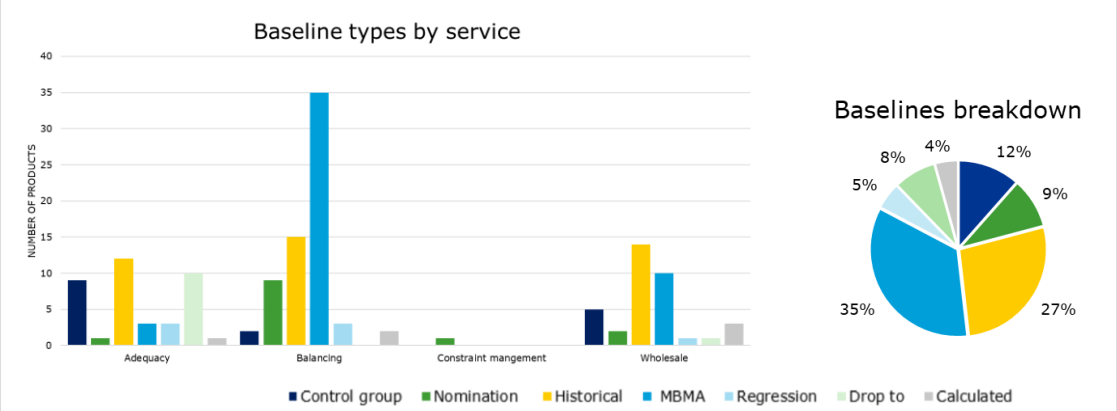
- Sustain and Secure Scheduled have very long utilisation instruction notification periods which do not exist elsewhere in the GB and international practices. As such, we could not identify any best practices for these products.
- Real time products (i.e. Dynamic, Restore and, Secure Dispatched – real time), show similarities with balancing products and adequacy product on the notification instruction timing. However, the utilisation period of DSO products is notably longer than what we see in international practices. Although we have identified adequacy, wholesale and balancing products with longer utilisation periods, they usually do not surpass 4 hour periods. There are only 4 products with utilisation periods longer than 8 hours:
    - There is 1 adequacy product (Canada) with utilisation period longer than 8 hours, but this product is capacity payment only and uses a drop-to mechanism, which cannot be used in products which remunerate both availability and utilisation.
    - GB's Capacity Market has an indefinite utilisation period. Capacity Market does not use any baselines for generations and uses historical baseline for Demand Side Response (DSR) assets.
    - ESO's FFR (high) has an indefinite utilisation period, and short utilisation instruction notification. The performance is measured by the differences between demand/generation at the start of a sample period and its actual demand/generation. A new methodology, which is under trial, proposes using a forecast that captures the response of naturally variable baseline (i.e. nomination).
    - Finally, in ESO's Dynamic Containment contacts are for 24hours, for full delivery on the service day. As this product is under soft launch at this stage, it is worth noting that nomination baselines will be used (Physical Notification).

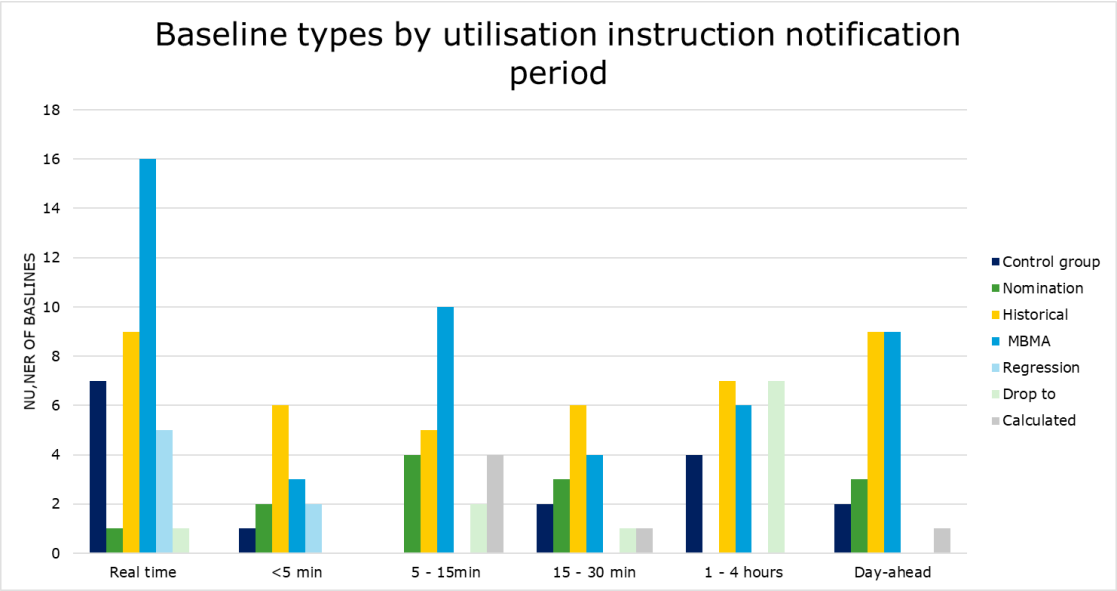
All things considered, we cannot suitably map ENA DNO products against GB and international practices. This observation aligns with the messages from stakeholder during our bilateral engagement. Nevertheless, we have taken into account the above-mentioned examples in our recommendations.

In addition, we provide below our observations of our market assessment which have only partially informed our recommendations (see also Figure 5 for reference):

- MBMA is the most common baseline and mainly used in balancing services. MBMA is a common baseline for short and fast reacting balancing services and the most common baseline for the shortest utilisation instruction (i.e. less than 15 minutes).

- Historical baselines are the second most common baselines. They are met across all products and are the most popular baselines across products with longer utilisation periods such as wholesale and adequacy products which typically use historical baselines.
- Regression-based, calculated and control group baselines are less common; these baselines are usually preferred when the common baselines are not suitable. Notably, regression baselines are used for longer utilisation periods, as they can increase accuracy using external variants.
- Nomination baselines are more common in balancing products, with short notification periods, Nomination have also been used for relatively longer utilisation periods. It worth noticing, that nomination is used in the only DSO constraint management product which is included in our list.
- Finally, a key observation is that a large number of products (~ 40) provide a choice of baselines to the FSPs. This is mainly driven by the fact that some of the common baselines (historical and MBMA) are not always suitable for certain assets/customers or technologies, particularly as new assets enter the flexibility markets.





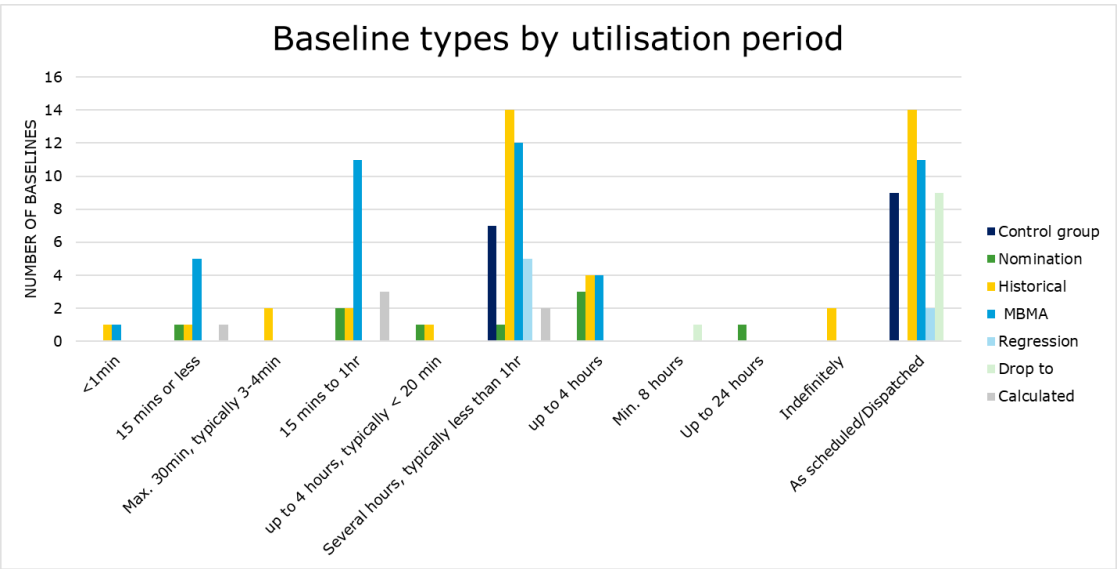


Figure 5: Key statistics of the market assessment[7]

7 "As scheduled/ Dispatched" is a common requirement in wholesale products in the US. Typically the utilisation periods of these products are minutes to 1-2 hours.

4.3.1 Alignment with the ESO

Alignment with ESO products, albeit not expressed in the key principles, nor explicitly in the assessment criteria for baseline methodologies (please refer to section 5.1), has been a key element in both our analysis and recommendations. This is highlighted through the following elements:

- **Examining of best practices**: Many ESO (viz. TSO / ISO) products have been analysed, including their associated baseline methodologies, such as STOR, Dynamic Containment, FFR services and GB's Capacity Market. Products with similar characteristics compared to the different DNO products are considered "best practices" and provide important input to our baseline assessment and selection.
- **Baseline methodology assessment**: In analysing the costs and benefits for the FSP, we have taken into account potential synergies when baseline methodologies are aligned with those that are currently applied for any of the ESO products. The main example is the nomination baseline methodology type that resembles the physical notification used in the balancing mechanism and in the Dynamic Containment.
- **Process considerations**: In our assessment, we also describe the information exchange associated with the baseline calculation. For example, using the sub-meter that is currently used for ESO services, for DNO services, creates synergy for both the DNO and FSP.

However, due to substantial differences between ESO and DNO products, there are also limitations in the synergy that can be achieved. For example, although the Sustain product may strongly resemble the national capacity market, the long notification time will necessarily lead to a different baseline design, as will be pointed out in the next sections. Also, the long utilisation periods render the baseline methodologies used in several ESO products unsuitable for DNO products.

4.4 Current and future technologies

All DSO products are technology-agnostic, meaning that they are open for all technologies to participate and do not include explicit requirements (or) favouring certain technologies.

It is important to understand, however, that certain technologies have a better match with certain products than others. For example, products with long utilisation periods will favour (dispatchable) generation assets to storage facilities. Yet this is not driven by the technical product definition, but by the economics of the assets looking to participate.

With many new (DER) technologies entering the market in large numbers, e.g. EV chargers and heat pumps, it is not yet clear which technologies have sufficient economical fit with the DSO products.

The baseline methodologies that are assessed in this study cover the following technologies:

- **Generation**: Both dispatchable and non-dispatchable, fossil-based and renewables
- **Demand (load)**: Electricity consumption
- **Storage**: Mainly batteries
- **Any combination of the above**: At the customer side, often different technology types are combined behind the meter**.**

The baseline methodologies also cover different customer segments: **Industrial**, **Commercial** and **Residential**.

In most of the international best practices and the current DSO products, only a limited set of technologies are / have been participating. This is mostly limited to generation (CHP/gen sets) and industrial load. The corresponding baseline methodologies are typically designed for these types of technologies, and may not necessarily facilitate some of the future technologies. We take this into account using the following principles:

- We understand that the best practices are used and proven for a small set of current technologies, and may not support all future technologies;
- We want to ensure that standardisation of baseline methodologies for the future DSO products should at least allow the current technologies to continue participating; and
- We should develop a view how inclusivity for future technologies can be achieved, without making the baseline methodology overly complex at this point in time.

5 KEY OUTCOMES

5.1 Assessment criteria and prioritisation

ENA WS1A P7 team developed the principles which formed the basis for the qualitative assessment of existing baseline methodologies and cover the key aspects that baseline methodologies are expected to meet. The product team then considered the additional factors that should be included for the scoring and assessment of baseline methodologies. These principles and factors are displayed at the first column of Table 7.

Based on baseline principles we developed the assessment criteria against which the baseline methodologies have been assessed (Table 7).

As a next step we prioritised the assessment criteria. Our experience in the design of baseline methodologies shows that priorities vary across flexibility products, markets, organisations and countries. Whilst accuracy of the baseline is a key principle for mature markets, simplicity and inclusivity may be prioritised at markets which are at their infancy. Therefore, we tailored our prioritisation to GB DNO flexibility markets, based on the following evidence:

- Following discussions with DNOs and WS1A P7 we understand that the priorities for DNOs are to enable further market participation, remove barriers for FSPs, provide clarity to FSPs and streamline the processes across DNOs where this is possible.
- In addition, our review of the consultation responses has indicated that stakeholders are interested in simplicity, consistency, transparency and inclusivity, so that all technologies can participate in the flexibility services.
- Bilateral engagement with stakeholders has further evidenced that simplicity and inclusivity of technologies and DER assets are key priorities for them. Bilateral engagement has also indicated that stackability would increase market participation and is an important factor for FSPs to enter flexibility markets, if they can preserve revenues from other value streams.

Table 7 shows the principles, the assessment criteria and the prioritisation that we applied. In our scoring exercise we weighted the criteria based on their prioritisation from 1 to 4, with "1" reflecting the lowest prioritisation. We then proportionally converted the prioritisation into a weighting factor; the sum of all prioritisation scoring was 23; the score of each criterion was proportionally converted into a weighting factor so that the sum of all weighting factors is 100% *(e.g. weighting factor of Replicability = 2/23 = 8.7%)* Section 5.3 includes the outcomes of the scoring of each baseline methodology against the criteria for each product category.

Table 7: Assessment Criteria Prioritisation

Principles	Criteria and weighing factor	Description	Priority
Simplicity	DNO imple-mentation costs (13%)	Are the costs for implementing and operating the administrative processes proportionate for the DNO?	High
Simplicity	FSP imple-mentation costs (17.4%)	Are the costs for implementing and operating the administrative processes proportionate for the FSP?	Very High
Replicability	Replicability (8.7%)	Is the baseline reproducible by the DNO, FSP, and third-party validator for settlement (verification) purposes?	Medium
Design fit	Robustness to data (4.35%)	Are there high requirements on data to calculate the baseline? Do data quality issues undermine the baseline quality?	Low
Accuracy	Variance (8.7%)	Does the Baseline Methodology provide an accurate estimate of the flexibility load impact at a level expected by DNO and FSPs, or does it show a relatively high variance?	Medium
Accuracy	Bias (8.7%)	Does the Baseline Methodology provide an unbiased estimate of the flexibility load impact at a level expected, or does it show a relatively high bias? [8]	Medium
Integrity	Integrity (8.7%)	Does the Baseline Methodology avoid or minimize the risk of gaming and strategic behaviour?	Medium
Inclusivity	Technology agnostic (17.4%)	Is the Baseline Methodology technology agnostic and not biased to a particular type of solution, technology and provider?	Very High
Design fit	Design fit – parameters (4.35%)	Can specific parameters of the service design be met?	Low
Stackability	Stackability (8.7%)	Does the Balancing Methodology allow the FSP to combine the delivery (Availability and/or Utilisation) of DNO products with other markets?	Medium

We scored the baselines against the assessment criteria based on the assessment framework which is shown in Table 8. Section 5.3 includes the outcomes of the scoring.

Table 8: Assessment Criteria Framework

Criteria	5 - Excellent	4- Good	3 - Average	2 - Fair	1 - Poor
DNO imple-mentation costs	Simple, straightforward calculation, e.g. MBMA	Relatively simple calculation	Slightly complex calculations	Rather complex calculations	Highly complex calculations, e.g. regression
FSP imple-mentation costs	as above	as above	as above	as above	as above
Replicability	Easy to be replicated by different parties.	Relatively easy to replicate by other parties. Same data should be used.	It can be replicated by different parties, dependent on data	Difficult to be replicated by different parties, due to complex	Very difficult to be replicated by different parties due to complex calculations, use

[8] Lack of bias: no systematic tendency to over- or under-state reductions

Criteria	5 - Excellent	4- Good	3 - Average	2 - Fair	1 - Poor
				processes or lack of information.	of certain data or intelligence.
Robustness to data	No dependency on data	Availability of data is not an issue	Availability of data can be an issue, dependency is low	Availability of data can be an issue, dependency is average	Availability of data can be an issue, dependency is strong
Variance	Small variance for all technologies	Small variance for most technologies, medium variance for some profiles (e.g. volatile load).	Small, medium and high variances depending on technology	Some small and medium variances, mostly high variance for different technologies.	High variance for most technologies
Bias	No bias throughout the service window	Small bias in certain parts of the service window	Small or no bias on average, yet higher bias at specific moments of the day / service window	Some bias on average, yet higher bias most moments of the day / service window	Strong bias throughout the service window
Integrity	No gaming options foreseeable	Some gaming options, with sufficient mitigations options	Likely gaming options with good mitigation / some gaming options with little mitigation	Likely / obvious gaming options with some mitigation	Obvious gaming options with little/no mitigation
Technology agnostic	Fully technology agnostic	Technology agnostic with few exceptions	Favouring specific technologies, not excluding others	Favouring specific technologies, excluding some others	Favouring specific technologies, excluding others
Design fit - parameters	Excellent fit to product design	Good fit to product design	Average fit to product design	Fair fit to product design	Poor fit to product design
Stackability	Stacking on Capacity Market (CM) and Wholesale (WS) always possible	Stacking on CM and WS in most situations possible	Stacking on CM and WS sometimes possible	Stacking on CM and WS in exceptional situations possible	Stacking on CM and WS never possible

5.2 Exclusion and addition of baselines

Our initial analysis considered all the baseline types that are presented in Table 4. Whilst progressing with our analysis and assessment, we excluded and added variants of the main baseline types.

Exclusion of Calculated and Control-Group baselines

In our assessment process, we have first put emphasis on baseline methodologies that can be (in principle) used for all technologies, trying to keep complexity low. For this reason, we have decided to initially exclude the calculated baselines and the control-group baselines, as these baselines are only applicable to a small set of technology types. For example, a calculated baseline which uses weather and wind data could be applicable for wind generation; a control-group baseline would be suitable for EVs chargers. These methodologies (in their specific implementation), however, cannot be applied wider. When no suitable technology-independent baseline methodologies can be found, the calculated baselines and the control-group baselines will be considered again.

Addition of Same Day Adjustments (SDA)

Same Day Adjustment of a baseline implies that the baseline is adjusted based on certain measurements during the delivery day to improve the accuracy of the baseline. This could e.g. account for unforeseen changes and variable loads (e.g. weather conditions). SDAs are particularly relevant for historical baselines where historical data may well represent the load shape, but may not always show the right level, as historical data insufficiently reflect same-day conditions. As SDAs can improve both the variance and bias of a baseline methodology considerably (and may score different on these criterions), we have included two subvariants of the historical baseline methodology types in our assessment: with and without SDA.

Although SDAs are suitable for historical baselines, we consider them unsuitable for flexibility products with utilisation instruction notification periods longer than a day. Although they could improve the accuracy they would certainly deteriorate the integrity. More specifically, we recommend that the FSP should not be capable of manipulating their baseline after the utilisation has been instructed. As such, we have ruled out methodologies with SDAs for Category A and Category B products.

Exclusion of Meter Before Meter After (MBMA) and Window Before After (WBA)

As per Table 6, the utilisation periods of all DNO products are relatively long (from hours to days). Therefore, we have excluded the methodologies of MBMA and WBA. As we have observed in the market assessment, MBMA and WBA methodologies are suitable for shorter notification periods (e.g.in the scale of minutes). Using MBMA and WBA in GB DNO products would significantly deteriorate the accuracy of the baseline methodology to unacceptable levels.

Although some products may accept offers/portfolios with shorter utilisation periods, defining a separate methodology only for these specific cases would create additional complexity.

The use of MBMA and WBA could be revised in the future if flexibility products move to shorter utilisation periods. Nevertheless, the effectiveness of SDAs is similar to these two baseline methodologies and provide similar accuracy results. As such, SDAs provide a future-proof solution and we do not expect the need for MBMA/WBA solutions.

5.3 Analysis of baselines per product

This section describes the scoring of the shortlisted baseline methodologies against each product, our approach and provides further details on the design parameters of the highest scored methodologies.

5.3.1 Category A – Sustain & Secure Scheduled

The Sustain and Secure Scheduled products are unique products due to their long utilisation instruction notification period which can be even 1 year in advance. In addition, these two products have long utilisation periods: 1 of the DNOs has required 30 days service delivery for the Sustain product and 16 hours service delivery for the Secure Scheduled. Because of these two parameters and the way that DNOs use Sustain product (i.e. reinforcement deferral), it resembles a local capacity market. However, DNOs will pay both for availability and utilisation.

Secure Scheduled is used pre-fault and for peak shaving. Although its applicability is different from Sustain Product, our assessment and recommendations are the same due to the long utilisation instruction notifications and the utilisation periods.

Considering the product parameters, generation assets with non-variable profile will make a better fit for the Sustain and Secure Scheduled. Although the products are open to all technologies, in practice only few technologies are able to compete on an economical basis due to the long notification and long utilisation

periods; this product therefore mainly attracts dispatchable generation. Depending on product specifications renewable generation can also contribute to this product.

Our engagement with the DNOs has also shown that there is still uncertainty on how and when these two products will be applied; more experience will provide further clarity and alignment across DNOs. There is, for example, unclarity if energy efficiency programs can participate in this product, and if load that is permanently aborted can participate.

The fact that non-dispatchable, renewable generation can participate using de-rating factors, shows that baseline methodologies, which in general are used for calculating the activated volume, are not applicable. In any flexibility product, the activation is calculated as the deviation from the "normal" or counterfactual generation or load profile (baseline) and availability payments ensure that this flexibility is available to the DSO when required. In the case of the Sustain product, renewable generators are not curtailed, so any proper baseline methodology would generate a profile very close to the measurements, not leading to an energy payment. For non-dispatchable generation and in order to account for intermittent generation, de-rating factors can determine the expected level of capacity that a renewable generation asset can secure.

We have considered drop-to mechanisms for this product. We believe this could be relevant for certain technologies (esp. load), yet this is mainly interesting for capacity-only products. Since these products also include an activation remuneration, the activated volume needs to be calculated, which would require some sort of baseline methodology next to the drop-to mechanism. Therefore, a drop-to mechanism does not provide added value for this product.

Finally, our market assessment of international best practices has indicated that the parameters of these two products are not met in any other GB and international products. As such we have developed our recommendations based on our experience and the design fit for the products:

- We believe that Sustain product is not yet sufficiently crystallized to standardize the verification methodology, nor the associated baseline methodology. None of the DNOs has used Sustain product yet, although most of them have tendered for 2020 or following years. Similarly, only one DNO has procured Secure Scheduled, although they haven't dispatched the service yet. As such, we recommend that more experience is gained by DNOs before moving to standardisation.
- For future standardisation efforts we suggest considering the following approach:
    - For dispatchable generation assets we recommend setting the baseline to 0. The asset is required to provide capacity when needed, regardless of how it would have otherwise dispatched. To remunerate the activation, we recommend the flexibility DER assets would be compensated for the *hours of delivery x capacity.*
    - For non-dispatchable generation, we recommend the same approach using de-rating factors. We recommend that the de-rating factors are the same as these which are used in the Capacity Market to achieve further alignment and consistency with the ESO.
    - For load assets, we recommend using a historical baseline which uses data from the previous year. To increase consistency, the same historical baseline methodology as recommended for secure dispatched (week ahead) can be applied.

5.3.2 Category B – Secure Dispatched (week ahead)

5.3.2.1 Scoring

Based on our initial analysis, which is described in section 5.2, we scored 3 baseline methodologies against the assessment criteria and week-ahead Secure Dispatched parameters: nomination baselines (BLs), historical BLs and regression BLs. Table 9 displays the scoring results. The total scoring of each baseline uses the sum-product of the weight of criterion by the score of each baseline against this criterion (e.g. Total scoring of nomination = 3 x 13% + 3 x 17.4% + 2 x 9% etc.)

Table 9: Scoring of baseline methodologies against week-ahead Secure Dispatched[9]

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression-based**
Total Scoring		3.0	3.3	3.1
Simplicity	**DNO implementation costs (13%)**	3	3	1
Simplicity	**FSP implementation costs (17.4%)**	3	4	4
Replicability	**Replicability (8.7%)**	2	5	2
Design fit	**Robustness to data (4.35%)**	4	3	2
Accuracy	**Variance (8.7%)**	3	2	4
Accuracy	**Bias (8.7%)**	3	2	4
Integrity	**Integrity (8.7%)**	2	4	4
Inclusivity	**Technology agnostic (17.4%)**	3	3	3
Design fit	**Design fit – parameters (4.35%)**	3	3	3
Stackability	**Stackability (8.7%)**	5	4	4

Table 10 further justifies the scores of the baselines in the table above.

Table 10: Scoring justification for week-ahead Secure Dispatched

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression-based**
Simplicity	**DNO implementation costs**	DNOs will need to perform monitoring and validation activities	We assume that the DNOs will calculate the historical baselines and provide the BL to the FSP before utilisation.	We assume that the DNO provides the BL to the FSP. Regression BLs are very complex to implement.
Simplicity	**FSP implementation costs**	The FSP provides the BL. Simplicity depends on the tools and data that the FSP will use to provide their forecast of demand/generation. It	This BL is simpler for the FSP, as the DNO provides the BL. The main requirement for the FSP is to provide measurement data (in case of sub-metering) and activation data to DNO.	

[9] Colour coding of principles indicate their prioritisation as per Table 7. Colour coding of baselines' scoring against the selected criteria demonstrates the relative scoring of each baseline for the specific criterion in each row (i.e. dark green is the highest score and red the lowest).

Principles	*Criteria*	Nomination	Historical baseline	Regression-based
		also depends on the frequency that nomination baselines should be sent to the DNO.		
Replica-bility	**Replica-bility**	It is difficult for other parties to replicate the BL, as they need to know how FSPs perform their forecast. FSPs can use a wide range of methodologies to calculate their planned profiles.	It is easy for other parties to replicate the BL, as long as they all use the same data. The only drawback is to ensure that appropriate data is available (i.e. raw data versus validated data).	It is relatively difficult to be replicated, as all parties need to use exactly the same methodologies.
Design fit	**Robust-ness to data**	Although this depends on the forecast approach, in principle nomination has smaller dependency on data compared to the other two BLs.	Greater dependency on quality of data. Less data required.	Greater amount of data required for a useful regression but less dependency on quality of data.
Accuracy	**Variance**	Due to long utilisation periods and long notification periods accuracy is challenging, especially for variable loads.	The high frequency of activation creates challenges as only data from non-event days data should be used. Same comments on utilisation periods and notifications as nomination.	Regression baselines improve the accuracy as they can predict extreme condition load shapes based on the relationship between load and weather in less extreme conditions, to the extent that regression captures the dynamics that drive extreme conditions.
Accuracy	**Bias**			
Integrity	**Integrity**	Due to the week ahead notification and high activations frequency, nomination is highly susceptible to gaming.	Both historical and regression baselines reduce the gaming opportunities as they are based on data and methodologies which cannot be easily manipulated.	
Inclusi-vity	**Techno-logy agnostic**	Due to days ahead utilisation instruction, all BL score average in inclusivity. Long notification periods are particularly challenging for non-dispatchable generation and variable loads.		
Design fit	**Design fit–para-meters**	Although "design-fit parameters" is a valid criterion to include in the assessment, we did not use it due to lack of sufficient experience across DNOs operating the products and due to of variants of the key design parameters.		
Stacka-bility	**Stacka-bility**	Stackability is facilitated by all baselines, since no other flexibility trading has started at the time of the utilisation instruction. Nomination scores higher as prior event days are known to the FSP, which makes nomination easier for stacking and for forecasting stackability requirements.		

The three BLs score average for this product and lower than they score for the other products (sections 5.3.3 and 5.3.4). This relatively low score lies in the product characteristics and particularly the week ahead utilisation instruction. In the future and as DNOs get more experience in operating flexibility

services, DNOs could re-assess the suitability of these baselines and particularly their accuracy. If all baselines are not sufficiently accurate, then DNOs may decide to move to day-ahead utilisation instruction or real-time, or to consider regression-type baseline methodologies.

5.3.2.2 Shortlisted baselines

At this stage our recommendation is to shortlist Nomination and Historical baselines (without SDA) for Secure Dispatched (week-ahead). Although the Regression-based baseline has similar scoring, for consistency reasons across all products we recommend excluding it.[10]

Historical baselines

As discussed in Section 2, historical baselines are baselines constructed with historical data, based on an explicit set of data selection and exclusion rules as well as simple estimation methods. The most common historical baseline is the 10 of 10 baseline applied to weekdays. The 10 days are selected from the most recent non-activated weekdays prior to the day of notification and no additional exclusion apply. The simplicity of the 10 of 10 baseline balances the inclusion of enough days to produce a stable baseline while not requiring too long a lookback period.

Variations on the basic 10 of 10 baseline structure seek to improve the baseline to address specific load characteristics. For example, a 5 of 5 baseline limits the data requirement and targets days closer to the event day. Both characteristics might be preferable for a product with extremely frequent activations.

For weather-sensitive loads and in case that activations are most likely to occur during extreme temperatures, excluding lower load days from the day pool can produce a load shape that is better shaped for those events (e.g. high 5 of 10 or high 3 of 5). Finally, to address extreme loads, the mid 8 of 10, which excludes extreme loads from both top and bottom of the day, makes the baseline less sensitive to unexpected loads in the day pool. Generally, the ranking of the day pool for the purpose of applying exclusions is performed on load, during the expected activation period on baseline days. With a variable length activation period, the ranking should be done on a pre-determined, multi-hour peak load period to apply exclusions based on peak period load characteristics.

If ENA intends to identify a single historical baseline, then the baseline should be the standard 10 of 10 or a close facsimile. Like the 10 of 10, the mid 8 of 10 offers the greater stability of more than 5 days but also includes the outlier protection of the exclusion of extreme days. Baseline lookback periods are commonly limited to 6 weeks. If a product activates customers less than two-thirds of the time, then there will be a sufficient number of non-activated days in the prior 6 week-period to support a mid 8 of 10 baseline.

Most products allow for advance notice of planned outages. Indicating inactive days in advance and removing them from potential activation improves the reliability of the product and lowers customer risk. These days should also be excluded from future baselines along with activation days. If inactive days are consistently accounted for, then the exclusions of the 8 of 10 baseline become less necessary. However, the exclusions of the 8 of 10 baseline will still protect against the negative effects of an unplanned outage or other extenuating circumstances without moving far from the standard 10 of 10 baseline.

The most challenging aspect of a historical baseline is whether there is a sufficient number of like-days available[11]. A week-ahead notification corresponds to 10-days notification for the days at the end of the

[10] We have also shortlisted historical and nomination baselines for the remaining products, based on our assessment results. We reckon that at this stage it would be beneficial for DNOs to focus on developing a limited number of baseline methodologies which are as consistent as possible across all DNO products. Consistency would facilitate FSPs and market participation. We have further emphasized this message in our final recommendations.

[11] In practice, this would lead to using less days than the method prescribes, which will affect the accuracy of the outcome.

week. The availability of like-days is not only depending on the activation frequency and the distribution of activations (e.g. many consecutive days in the winter period) within the product, but also on the frequency of activation of the asset in other markets (in which case the corresponding days also need to be excluded). If like-days need to be sought too far in the past, then a 5 of 5 baseline should be preferred.

The 10 of 10, mid 8 of 10 or 5 of 5 baselines are designed for weekday events. Such baselines will have a weekend version that parallels the weekday baseline while dealing with the more limited day pool available. It is typical for weekend baselines to be based on the most recent pre-notification like 4 days of data. The smaller number of days makes it less compelling to apply some of the data exclusions such as removing extreme days from among the day pool. On the other hand, a baseline based on only 4 days already gives an extreme day shape a disproportionate weight in the final baseline. A mid 2 of 4 weekend/holiday baseline maintains the basic mid 8 of 10 structure for weekends and holidays. A 5 of 5 baseline would typically correspond to a 2 of 2 baseline for weekends.

As with most historical baselines the suggested baselines should be calculated using a simple mean for each interval across the final set of chosen days. DNOs use metered historical data and FSPs can replicate the baseline using their own data. To support this flexibility, rules for identifying inactive days should be clear and official lists of days which are accessible to both parties.

Nomination

The use of Nomination baselines or Forward Schedules (which is a common term in GB) has not been widely used in GB or in international practices. In GB, the main application of nomination baselines is in the Balancing Mechanism, where the Final Physical Notification (FPN) is used as a baseline and indicates that the Balancing Mechanism (BM) Unit moves away from its initial profile (before Gate Closure) in return for payment. The ESO uses this deviation from the Physical Notification to compensate the BM participant for changing their generation or their load.

Recently, the ESO started using the Physical Notifications as a baseline for the Dynamic Containment, although in this product the baseline is not used for settlement, but for measuring and monitoring performance.

The nomination baselines are less suitable when the main meter is used for settlement (i.e. submetering is not available)[12], since the FSP will have less visibility on the total consumption of the connection. When using the nomination baseline methodology, we recommend the following:

- The nomination baseline should be fixed before the utilisation instruction is sent. The FSP will send to the DNO the planned demand or generation profile of the DER for each settlement period of the service window.
- Timeline of baseline submission. The exact timings should be explained in the service requirements. We recommend that there is at least one agreed time period that the FSP should submit their baseline. As this is a week ahead product, the DNO should have submitted the nomination at least 1 week in advance and well before the utilisation instruction.
- The time resolution of the baseline should be the same as the resolution of product's metering requirements.
- The FSP needs to send an updated nomination when his forecast changes.
- When the DNO submits a utilisation instruction, the most recent nomination is converted into the baseline for that utilisation period.

[12] i.e. the Distributed Energy Resource/flexibility asset shared the same meter with non-flex. assets.

- Once the utilisation instruction has been notified, the FSP cannot change the baseline.
- The DNO will use the planned profile to calculate the deviation of actual sub-metered data, which will be reflective of the activated flexibility, for which the FSP will be compensated.

5.3.3 Category C – Secure Dispatched (real-time)

5.3.3.1 Scoring

Based on our initial analysis, which is described in section 5.2, we scored four baseline methodologies against the assessment criteria and real-time Secure Dispatched parameters: nomination baselines (BLs), historical BLs, regression BLs and historical BL with SDA. Table 11 displays the scoring results.

Table 11: Scoring of baseline methodologies against real-time Secure Dispatched

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression -based**	**Historical baseline with SDA**
Total Scoring		3.7	3.5	3.5	3.7
Simplicity	**DNO imple-mentation costs (13%)**	3	3	1	3
Simplicity	**FSP imple-mentation costs (17.4%)**	3	4	4	4
Replicability	**Replicability (8.7%)**	2	5	2	4
Design fit	**Robustness to data (4.35%)**	4	3	2	3
Accuracy	**Variance (8.7%)**	5	3	4	4
Accuracy	**Bias (8.7%)**	5	3	5	5
Integrity	**Integrity (8.7%)**	2	4	5	3
Inclusivity	**Technology agnostic (17.4%)**	5	3	4	4
Design fit	**Design fit – parameters (4.35%)**	3	3	3	3
Stackability	**Stackability (8.7%)**	4	4	4	3

Table 12 describes our approach when scoring the baselines.

Table 12: Scoring approach for real-time Secure Dispatched

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression -based**	**Historical with SDA**
Simplicity	**DNO implementation costs**	Same as Table 10.			Similar to Historical. The DNO performs the SDA; already before utilisation if meter data is available.
Simplicity	**FSP implementation costs**	Same as Table 10.			Similar to Historical. The main requirement for the FSP is to provide the meter and activation data to DNO.
Replica-bility	**Replica-bility**	Same as Table 10			Similar to Historical. As this BL involves SDA calculations, replicability

Principles	*Criteria*	Nomination	Historical baseline	Regression -based	Historical with SDA
					is more complex due to the need for on-line meter data.
Design fit	**Robust-ness to data**	Same as Table 10			Similar to Historical.
Accuracy	**Variance**	FSPs have full visibility of their operations and external factors, close to real-time. Nomination has the highest score as it can also include technology-specific forecasting methods.	The high frequency of activation creates challenges as non-event days data should be used. The BL does not take proper advantage of real-time notification, compared to other BLs.	Regression methodologies use same day conditions/weather to re-calculate the baseline.	Through the SDA, the BL is adjusted to same-day's weather conditions and real-time operations.
Accuracy	**Bias**				
Integrity	**Integrity**	Due to high frequency of activation, the FSP can better predict future activations. As such, they can adjust their nomination BL to reflect these expectations.	Same as Table 10		Compared to the Historical BL, it is easier to influence the baseline since only the period used for SDA calculation needs to be manipulated. It is slightly better than Nomination due to the requirement to use historical data as the basis of the BL.
Inclusivity	**Techno-logy agnostic**	Nomination is the most inclusive BL for real time notification as it allows all assets to provide the most updated forecast of the generation/ demand.	Using historic data for long utilisation periods may be challenging for renewable generation, variable loads and weather dependent technologies.	Regression analysis enables weather-dependent technologies such renewable generation.	SDAs are more inclusive than simple historical BL or Regression as they are more suitable for technologies with limited historical data or with variable load (e.g. EVs, RES).
Design fit	**Design fit-para-meters**	Although design-fit parameters is a valid criterion to include in the assessment, we did not use it due to lack of sufficient experience across DNOs operating the products and due to of variants of the key design parameters.			
Stacka-bility	**Stacka-bility**	See section 6.1 (Stackability) for further clarification on our scoring			

All baselines score higher in this product compared to week-ahead Secure Dispatched, due to real-time utilisation instruction which increases accuracy and inclusivity. Regression-based baselines have the lowest scoring due to their complexity and their dependency on data, which both impede replicability. Nomination and historical with SDAs baselines have scored higher than the other two methodologies. As such we have shortlisted these 2 baselines and provided further analysis in the next section.

Although historical baselines without SDA score marginally lower, we have not shortlisted them as SDAs increase both the accuracy and inclusivity. The increased inclusivity is the main argument to prefer this baseline methodology. We acknowledge that accuracy criteria are not prioritised at this stage, however, in

the future and as GB flexibility markets become more mature, we expect accuracy of the baselines to get prioritised by DNOs and FSPs.

5.3.3.2 Shortlisted baselines

Historical baselines – SDA

The principles of the pre-adjustment baseline are the same between real-time and week-ahead Secure Dispatched. The real time Secure Dispatched product should use the same mid 8 of 10 baseline structure as the week-ahead Secure Dispatched. The strengths of this baseline, as discussed in Section 5.3.2.2, remain in place for this product. In addition, a single, consistent baseline approach across products will lessen the administration and education burden of the baseline calculation for all parties. The 8 of 10 baseline will be even more effective for the Secure Dispatched product due to the shorter notification period which will allow the day pool to be the most recent non-activated like days. 8 of 10 baseline allows one minimum and one maximum value do be removed, so that it provides a balanced baseline whilst maintaining a sufficient number of days to smooth variability. The mid 2 of 4 weekend/holiday baseline is also recommended.

The real time Secure Dispatch product should also take advantage of an SDA. The SDA will adjust each interval of the 8 of 10 baseline based on customer load prior to the event notification. SDAs adjustment periods range from 15 minutes to 3 hours in length starting just prior to the notification to calculate the adjustment. Longer adjustment periods are less susceptible to load variability but include load further from the activation period. For the longer activation periods that are possible with this product, a longer adjustment period should provide a more stable adjustment. We recommend a two-hour adjustment period using the last two hours of unaffected load prior to notification.

The nature of the adjustment, additive or multiplicative, may be a more important consideration than the length of the adjustment period. Multiplicative adjustments are intuitively attractive because the adjustment scales with the baseline. In reality, however, it is possible for adjustment error to be large relative to load (especially when adjustment periods fall during overnight or early morning hours), which can introduce extreme errors into the adjusted baseline. Where multiplicative baselines have been used, they are invariably capped at a +/- 20% maximum adjustment. This limits the potential for increased variability but does not change the underlying dynamic.

The additive adjustment, by contrast, seems clumsy but consistently outperforms the multiplicative adjustment in baseline analyses in terms of bias and variation. The only possible caveat to recommending the additive adjustment relates to the maximum length of the product's utilisation periods. Quantitative analysis of different kinds of adjusted baselines has not been performed on simulated events that last over 4 hours. Despite this, in the absence of evidence for longer utilisation periods, ENA should use the additive adjustment in the short term. ENA can perform a baseline analysis to test performance of the two adjustments for the range of activation period lengths actually experienced by customers in the product to support a longer-term decision.

There are fewer reasons to select a 5 of 5 baselines methodology compared to Secure Dispatched (week ahead) The potential impact of using fewer days on the accuracy, is likely to be addressed by the SDA. However, in case that the 5 of 5 is preferred for the Secure Dispatched (week ahead), we recommend using the same methodology for this product to ensure consistency and increase simplicity.

Nomination

The principles of nomination baselines are the same between real-time and week-ahead Secure Dispatched. The main differences lie in the timings and exchange of information between the DNO and the FSP. We recommend the following guidance for the use of nomination baseline in this product:

- The nomination baseline should be fixed before the utilisation instruction is sent. The FSP will send to the DNO the planned demand or generation profile of the DER for each settlement period of the service window.
- The exact timings should be explained in the service requirements. To align with the Balancing Mechanism, the DNOs could ask for at least 1 baseline submission day ahead and allow at least up to 1 hour ahead of each settlement period of the service window. However, feedback from stakeholders has indicated that 1 hour is still not suitable for less predictable assets. DNOs could allow further submission of baselines if the baseline has changed up to the point that the utilisation instruction is sent.
- The resolution of the baseline should be the same as the resolution of products' metering requirements.

5.3.4 Category D – Dynamic & Restore

In this section we first focus on the Dynamic product, which has been more widely tested by different DNOs than the Restore product. For our analysis we have assumed a typical utilization duration of 2-3 hours.

5.3.4.1 Scoring

Based on our initial analysis, which is described in section 5.2, we scored 4 baseline methodologies against the assessment criteria and the Dynamic & Restore Products parameters: Nomination baselines (BLs), historical BLs (without SDA), regression BLs and historical BL with SDA. Table 13 displays the scoring results.

Table 13: Scoring of baseline methodologies against Dynamic

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression-based**	**Historical baseline with SDA**
Total Scoring		3.8	3.6	3.4	3.9
Simplicity	**DNO implementation costs (13%)**	3	3	1	3
Simplicity	**FSP implementation costs (17.4%)**	3	4	4	4
Replicability	**Replicability (8.7%)**	2	5	2	4
Design fit	**Robustness to data (4.35%)**	4	3	2	3
Accuracy	**Variance (8.7%)**	5	3	5	4
Accuracy	**Bias (8.7%)**	5	3	5	5
Integrity	**Integrity (8.7%)**	4	5	5	5
Inclusivity	**Technology agnostic (17.4%)**	5	3	3	4
Design fit	**Design fit – parameters (4.35%)**	3	3	3	3
Stackability	**Stackability (8.7%)**	4	4	4	3

Table 14 describes our approach when scoring the baselines. The scoring of the four baselines against six criteria for Dynamic & Restore is the same with the scoring of real-time Secure Dispatched. This is

because the 3 products (real-time Secure Dispatched, Dynamic & Restore) have real-time utilisation instruction notification. The main differences of the scoring of baselines for Product D category lie in accuracy and integrity. These differences are driven by the post-fault activation and the low frequency of activation of Dynamic & Restore as well as the different utilisation periods which can reach up to days for Dynamic & Restore.

Table 14: Scoring approach for Dynamic

Principles	***Criteria***	**Nomination**	**Historical baseline**	**Regression-based**	**Historical with SDA**
Simplicity	**DNO implementation costs**	Same as Table 11			
Simplicity	**FSP implementation costs**	Same as Table 11			
Replicability	**Replicability**	Same as Table 11			
Design fit	**Robustness to data**	Same as Table 11			
Accuracy	**Variance**	The reasoning for our scoring is similar to the reasoning that we provided for real-time Secure Dispatched (Table 11). Fewer activations further improve the accuracy of the historical baselines. Yet it still does not provide full inclusivity.			
Accuracy	**Bias**				
Integrity	**Integrity**	As the products in the category are post fault products, with very infrequent (and therefore unpredictable) activation, gaming options are minimal. Nomination is scored slightly lower as the FSP can still create a structural bias (although this is easily detectable).			
Inclusivity	**Technology agnostic**	Same as Table 11			
Design fit	**Design fit - parameters**	Same as Table 11			
Stackability	**Stackability**	See section 6.1 (Stackability) for further clarification on our scoring			

At this stage we have shortlisted nomination and historical with SDA baselines for the Dynamic product, as they have the highest scorings in our assessment.

5.3.4.2 Shortlisted baselines

Historical baselines - SDA

The design principles of historical baselines for the DNOs' post-fault products are the same as for real-time Secure Dispatched product. We recommend using a mid 8 of 10 baseline with an additive adjustment. For weekends and holidays, we recommend the mid 2 of 4 baseline with the additive adjustment. We recommend a two-hour adjustment period using the last two hours of unaffected load prior to notification.

There are fewer reasons to select a 5 of 5 baselines methodology for the Dynamic product, although the potential impact of using fewer days on the accuracy, is likely to be countered by the SDA. However, if for the secure dispatched (week ahead) the 5 of 5 is preferred, we recommend using the same methodology for this product to ensure consistency and increase simplicity.

Nomination

The design principles of nomination baselines for the DNOs' post-fault products are the same as for real-time Secure Dispatched product.

5.3.4.3 Restore product

The Restore product has not yet been used by the DNOs which makes it difficult to assess, since the key design parameters may still evolve. We have, however, included a recommendation for a standardised baseline methodology.

The Restore product is very similar to Dynamic product. We differentiate between two categories of utilization:

- A utilization period of several hours (typically 2-3 hours)
  In this case the Restore is almost identical to the Dynamic product with relative short utilization period, discussed in the previous section. Our baseline methodology recommendation therefore follows this one.
- A utilization period of several days (typically 2-4)
  As mentioned before, these extreme long utilization periods are unprecedented in European and North-American markets. At the same time, it is important to realize that only few technologies will be able to participate in such a product. Dispatchable generation is the most likely technology to participate, in which case the nomination baseline would be preferable. Whether other technologies are likely to participate is something that needs to be experienced in the coming years.

Concluding, we recommend the same baseline methodologies for the Restore product as for the Dynamic product. We understand that the accuracy may provide challenges for long utilization periods. In this case the nomination baseline may provide a better solution. The same argument holds for the Dynamic product with long utilization periods.

The two-hour adjustment period may deviate slightly with the Dynamic product, as it may overlap with the outage. Therefore, the last two hours of unaffected load prior to the outage should be applied.

6 IMPLEMENTATION OF BASELINES

6.1 Stackability

The stackability options for flexibility providers do not only depend on the service definition. If a service allows stackability with other services, the choice of a baseline methodology should reflect this principle. For example, if an FSP offers DNO flexibility services and simultaneously participates in a short-term balancing product, what should the baseline reflect as the counterfactual? One could argue that the behaviour without any activation should represent the counterfactual, one other could argue that, from the DNO perspective, the behaviour including the ESO activations reflects the counterfactual. We believe this question has not yet been fully addressed in the GB market, therefore in this section we will focus on the implications on baseline design if stackability needs to be supported to the full extent.

Our analysis will for example show, that more complexity will be introduced when value stacking options are addressed, so a balance between simplicity and stackability needs to be sought.

In this section, we analyse the relation between the choice of baseline methodology and stackability options. This relationship depends on a range of parameters and it is far from straight-forward. Therefore, we rely on examples for our analysis that, for the sake of simplicity, we limit to the stacking of two types of revenues – DNO flexibility services and Wholesale.

First, we differentiate the different timing for stacking services. We employ the terminology that ENA ONP uses in the report on DNO Flexibility Services Revenue Stacking:[13]

- Stacking in adjacent periods of time
- Stacking in the same period of time

We'll analyse both types of stacking looking into 3 variables:

1. The type of baseline:
    - Nomination
    - Historical (without same day adjustment)
    - Historical with same day adjustment
2. Timing of the wholesale trade relative to the DNO flexibility service utilisation instruction notification:
    - Prior to the utilisation instruction notification
    - After the utilisation instruction notification
3. Start of wholesale trade delivery relative to the DNO flexibility service utilisation:
    - Prior to the start of the DNO flexibility service utilisation period
    - After the start of the DNO flexibility service utilisation period

[13] https://www.energynetworks.org/assets/images/Resource%20library/ON20-WS1A-P5%20DSO%20Revenue%20Stacking-PUBLISHED%20300720.pdf

6.1.1 Stacking in adjacent periods of time

Stacking in adjacent periods of time, as defined by ENA, refers to either the previous or next Settlement Period. For services procured in blocks of Settlement Periods, when referring to adjacent time periods, we are referring to the Settlement Period prior to the start or after the end of an availability block.

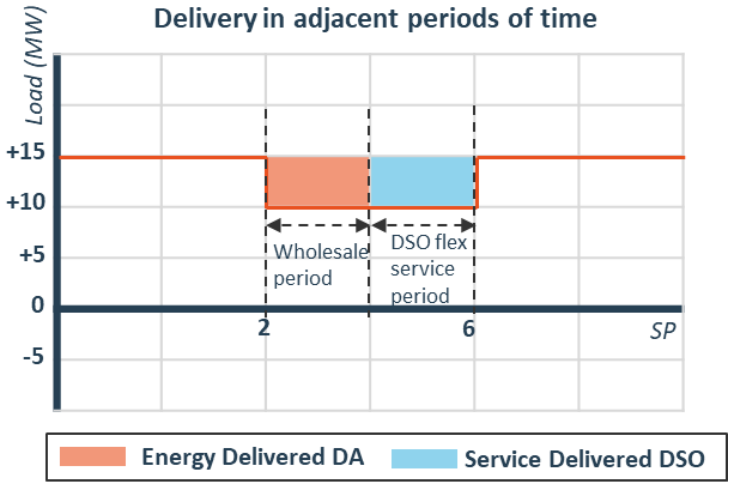


Figure 6: Delivery in adjacent periods

Figure 6 shows an example of stacking wholesale and DSO flexibility services in adjacent periods of time. From period 2 to 4, the FSP activates a trade in the wholesale market through a load reduction of 5 MW, which translates as 5 MWh in energy. In the consecutive settlement period, from 4 to 6, the same FSP with the same pool of assets, uses the 5 MW load reduction for a DSO flexibility service, in energy terms this is equivalent to 5 MWh.

Since, in this example, the wholesale trade / utilisation period ends after period 4, it is clear that the counterfactual of the DSO utilisation from period 4 to 6 is at +15MW. A baseline that allows this type of value stacking, should not be affected by the wholesale activation in the previous period.

Figure 7 provides two examples of two different approaches in the use of baseline for stackability purposes:

- ✓ Example 1: The baseline is set at 15 MW, not taking into account the previous wholesale delivery. This results in an accurate DSO flexibility delivery calculation.
- ✗ Example 2: The baseline is set at 10 MW, assuming the wholesale delivery as the normal behaviour of the load. In this case, the calculation falsely indicates that the flexibility was not delivered.

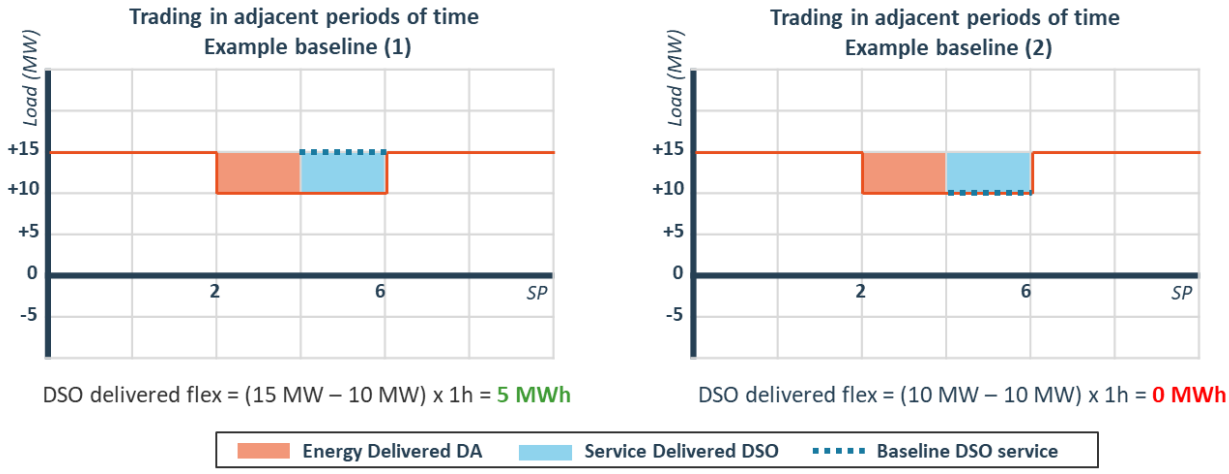


Figure 7: Baseline examples for delivery in adjacent periods

Considering the three different types of baseline methodologies:

1. **Historical baselines** do not take the same day into account, and are not affected by the adjacent wholesale activation.
2. In the case of **historical baseline with SDA**, there is a risk that the phenomenon seen in example 2 happens when the adjustment period coincides with the wholesale delivery period. Therefore, this type of baseline is compatible with stackability in adjacent periods of time when a) the wholesale delivery happens after the DSO service activation or b) when the wholesale trade delivery does not affect the adjustment.
3. Nomination baselines would forecast the expected load at +15MW and would not be affected.

If we broaden the definition of adjacent windows, taking into account wholesale trades on previous days:

1. Historical baselines, both with and without SDA, will be affected by wholesale trading on previous days, when the wholesale activations are part of data that is used for the calculation.
2. Nomination baselines are not likely to be affected, as the FSP is aware of these activations and will exclude it from its own calculations.

We recommend that, if ENA wishes to facilitate value stacking with wholesale markets,

- Prior-days wholesale flexibility activations are excluded from historical baseline calculations, for this purpose the FSP needs to inform the DNO about these activations
- Same-day wholesale flexibility activations are excluded from SDA calculation, for this purpose the FSP needs to inform the DNO about these activations.

If this recommendation is followed, then we can conclude that **nomination and historical baseline methodologies with and w/o SDA** are fully compatible with stackability in adjacent periods of time (for wholesale and DNO products).

One could also argue that, the more a portfolio or asset is active in other markets, the more these activations become part of the "normal behaviour" of that portfolio and asset, which should be reflected by the baseline. The ENA should decide if stackability is the most important criterion, or whether this argument, combines with a simpler solution, should be preferred. Also, as indicated before, a balance needs to be sought between stackability and the complexity that this concept introduces.

6.1.2 Stacking in the same period of time

Stacking in the same period of time refers to the simultaneous delivery of two different services. For services procured in blocks of Settlement Periods, when referring to same time period, we refer to the delivery of the second service coinciding or overlapping with the delivery of the first service.

This type of stacking is only possible when both services are activated 'in the same direction'. In other words, both services should request either upward or downward power.

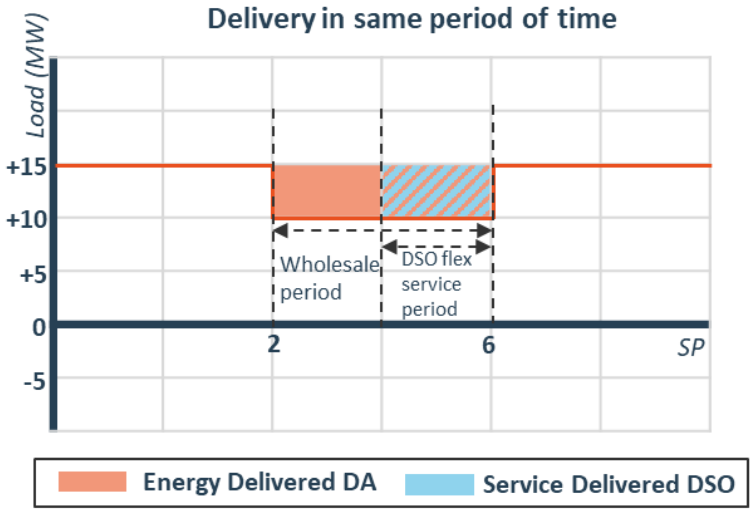


Figure 8: Delivery in same period of time

Figure 8 shows an example of stacking wholesale and DSO flexibility services in the same period of time. From settlement period 2 to 6, the FSP activates a trade in the wholesale market through a load reduction of 5 MW, which corresponds to 10 MWh in energy. Simultaneously, starting from the 4th settlement period to the 6th, the same FSP with the same pool of assets, uses the 5 MW load reduction for a DSO flexibility service, equivalent to 5 MWh in energy.

Stacking in the same period is more complex than stacking in adjacent periods; in this case the timing variables (timing of the wholesale trade and time of the delivery) come into place. We analyse the example of different baselines below:

- **Nomination baseline**: the FSP needs to send the nomination previous to the DSO utilisation instruction. Therefore, in this case, the relevant aspect is whether the wholesale trade happens prior or after the DSO utilisation instruction. **Figure 9** shows an example where the DSO sends a utilisation instruction at the 48th Settlement Period. The wholesale trade can take place either prior or after the instruction. In our example:

  ✗ Nomination 1 - Wholesale trade prior to the utilisation instruction: the FSP trades and plans 5 MW load reduction for wholesale from SPs 2 to 6, therefore the FSP updates its nomination to reflect the forecasted activation during those periods. **Figure 9** (nomination 1) shows the shape of the resulting nomination. Since the nomination already indicates 5 MW of load reduction for the wholesale delivery as the counterfactual, the DSO service delivery calculation would indicate that the service was not delivered, in other words, value stacking would not be possible.

  ✓ Nomination 2 - Wholesale trade after to the utilisation instruction: the FSP does not update the baseline with the planned wholesale trade. **Figure 9** (nomination 2) shows that in

this case the nomination is flat and set at 15 MW. When validating the DSO service delivery, the calculation will show the 5MW load reduction, thus enabling value stacking.

Conclusion: For nomination baselines, same period stacking is possible when wholesale trading occurs after DSO notification. In case the wholesale trading occurs before the DSO notification, value stacking is only possible when the wholesale trade is NOT included in the nomination.

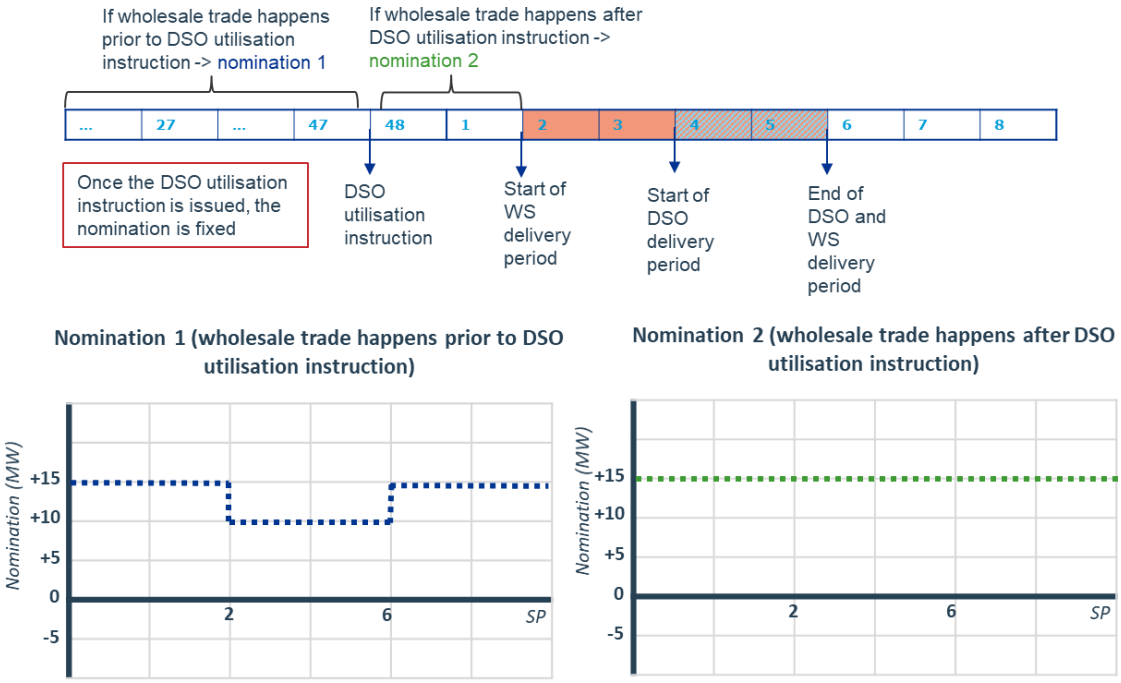


Figure 9: Baseline examples for delivery in same periods

- **Historical baseline w/o SDA**: The baseline will be the same regardless of the timing of the wholesale trade and activation, therefore it is fully compatible for stacking in the same period of time.
- **Historical baseline with SDA**: The conclusion is very similar to the one drawn in the section 6.1.1, this type of baseline is compatible with stackability in the same period of time provided the wholesale trading period does not overlap the adjustment period, e.g. if the wholesale delivery starts after the DSO utilisation instruction.

We recommend that, if ENA wishes to facilitate value stacking with wholesale markets, FSPs using nomination baselines should not include the effects of flexibility trading on other markets in their nomination.

6.1.3 Conclusions on stackability

As shown in the previous sections, baseline design can have a large impact on stackability options. However, the stacking options are not primarily driven by the baseline design; both market design principles and product design principles (including TSO-DSO coordination on product design) are key when shaping stackability. Baseline design should follow these principles, the previous sections have shown this can be accomplished in most situations.

However, at this point in time, no final answer can be given if / how the recommended baselines should facilitate value stacking, for the following reasons:

- **Lack of clarity about stackability options:** Although there is a strong preference to always allow value stacking, in practice we observe quite some limitations:
    - there is also a common understanding that, in case of TSO/DSO congestion management, value stacking is only allowed in the same direction (which is not the case when wholesale services are stacked);
    - it is not always clear that the DNO should pay for something that will happen anyway (take the example of section 6.1.2 between ISP 4 and 6);
    - the ENA report on DNO Flexibility Services Revenue Stacking only mentions wholesale trading and capacity market as possible markets to combine with DSO congestion management;
    - markets where energy is sold cannot be stacked since energy cannot be sold twice.
- **Impact of joint ESO-DSO procurement**
  If the ESO and DSOs decide for joint procurement of flexibility, it is questionable if ESO and DSO products can still be stacked.
- **Complexity**
  Only stacking wholesale trading and DSO products already leads to some complexity. Adding more markets and products will further increase the complexity of the validation and settlement processes, as the baselines of the different products will need to take the utilisation of the other products/markets into account. We believe a central validation and settlement of all markets and products will be unavoidable in this case.

We recommend further study is needed determining the extent to which stackability should be facilitated, how the validation and settlement should be organised, and finding the right balance between stackability and simplicity.

6.2 Information exchange

6.2.1 Use of historical baselines without SDA

Amongst other characteristics, historical baselines show:

- High replicability: In general, all stakeholders that have understanding of the details of the baseline methodology, and have access to the relevant historical data, should be able to calculate the baseline for any given period and asset/connection.
- Known in advance of the utilisation period: Since the FSP will be validated and remunerated based on the calculated baseline, knowing the baseline in advance is an important feature of any methodology.

We recommend that DNOs facilitate the FSPs by calculating and publishing all relevant baselines after the notification has been submitted, (well) in advance of the utilisation period.

Although the FSP should be able to perform this calculation itself, we recommend that DNOs calculate the baseline for the following reasons:

1. Cost efficiency: The actual implementation of a historical baseline has some complexities (e.g. exclusion of holidays and prior-event days). A central calculation facility could avoid the need for the FSPs to implement this, further reducing barriers to participate.
2. Unambiguous view: Different calculations could still lead to different outcomes, for example when data quality issues arise with input data. Providing the baseline in advance clearly indicates the required service delivery by the FSP, as the same baseline will be used for validation and settlement.

Since only previous dates prior to the notification will be used, the calculated baseline can be provided at the time of notification.

In case sub-metering is used, the FSP needs to provide the sub-meter data to the DNO (or allow access to the sub-meter), allowing the DNO to use this data for future calculations.

6.2.2 Use of historical baselines with SDA

In general, our recommendation for this type of baseline is similar to our recommendation for the historical baseline without SDA. However, complexities may arise with respect to the (timely) availability of the meter data to calculate the same-day adjustment. This is depending on the type of meter being used.

- When the main meter is used, the FSP may not have access to this data. The DNO will have access to this data, but may only receive it the next day (after the utilisation period).
- When the sub-meter is used, the DNO may receive the data only the next day (see previous section).

We recommend that DNOs facilitate the FSPs by calculating and publishing all relevant baselines after the notification has been submitted, (well) in advance of the utilisation period, in case on-line meter data is available (either from the main meter or made available by the FSP for the sub-meter). If the meter data is not available to the DNO, the baselines without SDA shall be published.

To reduce complexity for the FSP, without creating new barriers, we further recommend:

In case of sub-metering, when technically and economically feasible, the FSP should allow the DNO on-line access to the sub-meter.

6.2.3 Use of nomination baselines

Since the baseline needs to be fixated before the utilisation instruction notification is submitted, the FSP needs to submit its nomination on a regular basis when using a nomination baseline methodology. Only requesting a nomination when the flexibility will be utilised will largely increase gaming opportunities. This is similar to the physical notification in the balancing mechanism, yet a lower submission frequency ca be applied to reduce administrative burden.

We recommend that FSPs agree with the DNO to use a nomination baseline, send their nominations on a regular basis to the DNO. It shall be sent for the full week (Secure dispatched with week-ahead notification) or for one or more full days (Secure dispatched with intraday/real-time notification, Dynamic and Restore), at least covering the entire service window and taking into account the maximum duration of the utilization period. FSPs need to send updated nominations when relevant, until the notification has been received.

6.3 Monitoring and Validation

There are two main reasons to implement monitoring and validation of baselines during the operational phase of the service window:

- To assess whether the applied baselines are sufficiently accurate (and remain sufficiently accurate during the contract period). Inaccurate baselines lead to inefficiencies (paying for flexibility that is not provided) or ineffectiveness (paying for flexibility that does not solve the congestion).
- To identify gaming or strategic behaviour
  An FSP that is able to manipulate the baseline can also render the congestion management product inefficient or ineffective. It should be noted that strategic behaviour may be difficult to demonstrate, as normal behaviour could also influence the baseline in a very similar manner (e.g. preheating before the utilisation period, when an SDA is applied).
- We recommend that DNOs check the baseline accuracy regularly (or continuously) on non-event days, comparing either the calculated historical baseline or the nomination with the measurements. Both bias and variance should be checked against pre-defined thresholds.

This comparison can be applied both on portfolio level and on asset/connection level. We suggest performing both, as the performance on technology level is also relevant to monitor during these early stages of the products.

In the previous section, we already acknowledged the need to receive sub-meter data in case of historical baselines. For monitoring, this is needed for all types of baselines methodology:

Where sub-metering is used, the FSP needs to provide the sub-meter data to the DNO (or allow access to the sub-meter), allowing the DNO to use this data for baseline monitoring. Data can be sent on a daily basis, covering at least the service window.

Prequalification

Similar checks can be performed during the prequalification phase, either before the start of the contract when the baseline accuracy of an FSP's portfolio needs to be checked (either on portfolio or asset level), or during the contract period when new assets or customers are added to the FSP's portfolio.

We recommend that DNOs check the baseline accuracy before the contract phase of an FSP's portfolio and/or individual assets/connections. Both bias and variance should be checked against pre-defined thresholds. Failure to meet these thresholds could lead to the selection of another baseline methodology for specific assets / connections, or ultimately the exclusion of specific assets / connections. This also applies to newly added assets / connections.

When a nomination baseline is used, the FSP needs to send nominations during a certain period prior the contract period / service window, to allow the DNO to check the accuracy. If new sub-meters are installed for service validation, the data needs to be collected and verified for a similar period prior to the contract period.

To increase inclusivity, we suggest that the thresholds for variance and bias should not be very demanding. These can be further sharpened in the coming years, when more experience with the products, the baseline methodologies, the FSPs and the different technologies has been obtained.

7 RECOMMENDATIONS & NEXT STEPS

This section provides our recommendations for a baseline methodology for each ENA DNO product, based on our market assessment and subsequent analysis, as well as proposing next steps for the ENA.

7.1 Recommendations

We acknowledge that GB DNO flexibility markets are still at their infancy. Although all DNOs have now procured flexibility, they still need to gain more experience in procuring and operating flexibility. There is still progress to be made for further alignment and standardisation of DNO product parameters.

We formed our recommendations taking into account the current status of GB DSO flexibility markets and the current priorities of DNOs and FSPs, as per the consultation responses and the bilateral stakeholder engagements. Simplicity and inclusivity of flexibility markets are key priorities at this stage, so that wider market participation in DSO services is facilitated. We also acknowledge that stakeholders as well as Ofgem and BEIS ask for consistency across DNO products, ESO products and the wider industry.

On this basis, our recommendations focus on three types of baselining methodologies that are relatively simple, are known in GB markets, and which are currently in use by DNOs and/or in ESO balancing services and/or in the Balancing Mechanism. To facilitate technology-agnostic solutions and allow FSPs to agree a baseline with the DNO based on their needs and types of asset, we recommend an agile baseline methodology solution:

- Use a **historical baseline** methodology for all products as the **default option**:
    - Only for Secure Dispatched – real time, Dynamic and Restore we recommend performing a **same-day-adjustment**, to increase the accuracy.
    - In case of historical baselines, we recommend that DNOs calculate and share the baseline prior to the utilisation period to facilitate the FSPs and increase transparency and visibility of their processes.
- **Allow** for all technologies to use a **nomination-type baseline**, when the historical baseline is not sufficiently accurate (e.g. for future technologies such as EV charging and future customer segments such as residential):
    - The use of nomination baselines can be particularly suitable for generation assets, or connections with dominant generation, in all products. In addition, ESO products move also to nomination baselines (e.g. Dynamic Containment) so that there is alignment with the Balancing Mechanism. As such, the use of nomination baselines aligns with the consistency considerations across the GB industry.
    - FSPs choosing this option need to submit nominations regularly, prior to the utilisation instruction notification.

These generic recommendations do not apply to the Sustain and Secure Scheduled products, as we explain below.

Table 15 summarises our final recommendations for the ENA DNO flexibility products. Our engagement with the GB DNOs has shown that the same product is often used in a different way and under different circumstances. As such, we would like to highlight that these recommendations reflect the product parameters and use as displayed in Table 6.

Table 15: Final Baseline Recommendations for ENA DNO Products

Product	Recommendation	Reasoning
Sustain	More experience needs to be gained by all DNOs before moving to the standardisation of the validation process (including baselines, if applicable). We recommend testing technology-specific validation mechanisms, as further detailed in section 5.3.1.	Long utilisation instruction notification periods and long utilisation periods allow limited options for this product. A drop-to mechanism could be also suitable if no utilisation payment was included.
Secure Scheduled	More experience needs to be gained by all DNOs before moving to the standardisation of the validation process (including baselines, if applicable). We recommend testing technology-specific validation mechanisms, as further detailed in section 5.3.1.	Long utilisation instruction notification periods and long utilisation periods allow limited options for this product.
Secure Dispatched (week-ahead)	Historical baseline without SDA Mid 8 of 10 for weekdays, mid 2 of 4 for weekends. Nomination to be used • dispatchable generation • connections with dominant dispatchable generation • if accuracy levels of historical baselines are (too) low	Historical and nomination baselines are more suitable for longer utilisation periods and for long instruction notifications, where MBMA and WBA cannot be applied. Historical and nomination baselines are simpler compared to regression. Nomination is recommended for predictable and non-variable assets, esp. when sub-metering is available. Nomination can also be a solution for new assets, when there is lack of historic data. In addition, nomination can be an option in case of frequent activations, where the use of recent historic data can be challenging. All baselines score relatively low against the parameters of this product. In the future and as DNOs get more experience in operating flexibility services, DNOs could re-assess the suitability of these baselines and particularly their accuracy. If all baselines are not sufficiently accurate, then DNOs may decide to move to day-ahead utilisation instruction, or may consider regression-type baselines.

Product	Recommendation	Reasoning
Secure Dispatched (real time)	Historical baseline with SDA Mid 8 of 10 for weekdays, mid 2 of 4 for weekends. Additive SDA using the last two hours of unaffected load prior to notification. Nomination to be used • dispatchable generation • connections with dominant dispatchable generation • if accuracy levels of historical baselines are (too) low	Historical baselines with SDAs increase the accuracy compared to historical baselines without SDA. Both historical with SDAs and nomination baselines are simpler than regressions. Historical with SDAs and nomination are future-proofed solutions in case of short utilisation periods in the future, as they can still be used and provide similar results to MBMA and WBAs without the need to update the processes. Nomination is the most inclusive baseline for real time notification as it allows all assets to provide the most updated forecast. Historical with SDAs are also better suitable for technologies with limited historical data or with variable load (e.g. EVs, RES).
Dynamic and Restore	Historical baseline with SDA Mid 8 of 10 for weekdays, mid 2 of 4 for weekends. Additive SDA using the last two hours of unaffected load prior to notification (or prior to outage). Nomination to be used • dispatchable generation • connections with dominant dispatchable generation • if accuracy levels of historical baselines are (too) low	As this is a real-time post fault product with low frequency of activations, all baselines scored higher compared to the other products, driven by accuracy and integrity scoring criteria. We recommend historical baselines with SDA and nomination as they are both inclusive of different technologies, and to remain consistent with the Secure Dispatch product. Longer utilization frequencies can have a negative impact on the accuracy, in this case the FSP can decide to apply a nomination baseline, esp. when operating dispatchable generation.

Summarising our recommendations for different technology types we have the following observations:

Table 16: Recommendations for technology types

Technology	Recommendations
• **Generation**: Both dispatchable and non-dispatchable, fossil-based and renewables	We recommend nomination baselines for both dispatchable and non-dispatchable generation. These recommendations do not apply to Sustain and Secure Scheduled products.
• **Demand (load)**: Electricity consumption	

Technology	Recommendations
• **Storage**: Mainly batteries • **Any combination of the above**	We recommend historical baselines if this provides sufficient accuracy. Otherwise nomination baselines can be used.

7.1.1 Recommendations for implementation

On stackability:

We recommend that, if ENA wishes to facilitate value stacking with wholesale markets,

- Prior-days wholesale flexibility activations are excluded from historical baseline calculations, for this purpose the FSP needs to inform the DNO about these activations
- Same-day wholesale flexibility activations are excluded from SDA calculation, for this purpose the FSP needs to inform the DNO about these activations.
- FSPs using nomination baselines should not include the effects of flexibility trading on other markets in their nomination.

We recommend further study is needed determining the extent to which stackability should be facilitated, how the validation and settlement should be organised, and finding the right balance between stackability and simplicity.

On information exchange:

We recommend that

- For historical baselines, both with and without SDA, DNOs facilitate FSPs by calculating and publishing all relevant baselines after the notification has been submitted, (well) in advance of the utilisation period, when on-line meter data is available (either from the main meter or made available by the FSP for the sub-meter). Where meter data is not available to the DNO, the baselines without SDA shall be published.
- Where sub-metering is used, the FSP must provide the sub-meter data to the DNO (or allow access to the sub-meter), allowing the DNO to use this data for future calculations and for validation purposes. Data can be sent on a daily basis, covering at least the service window.
- Where sub-metering is used, the FSP should allow the DNO on-line access to the sub-meter, provided this is technically and economically feasible.
- FSPs that agree with the DNO to use a nomination baseline, must send their nominations on a regular basis to the DNO. Nominations should cover the full week (Secure dispatched with week-ahead notification) or one or more full days (Secure dispatched with intraday/real-time notification, Dynamic, Restore), at least covering the entire service window and taking into account the maximum duration of the utilization period. FSPs must send updated nominations when relevant, until the notification has been received.

On monitoring and validation:

We recommend that:

- DNOs check the baseline accuracy regularly (or continuously) on non-event days, comparing either the calculated historical baseline or the nomination with the measurements. Both bias and variance should be checked against pre-defined thresholds.
- DNOs check the baseline accuracy before the contract phase of a FSP's portfolio and/or individual assets/connections. Both bias and variance should be checked against pre-defined thresholds. Failure to meet these thresholds could lead to the selection of another baseline methodology for specific assets / connections, or ultimately the exclusion of specific assets / connections. This also applies to newly added assets / connections.

7.2 Next steps

We believe that standardisation of baselines methodologies across the DNOs is both desirable and feasible. A strong prerequisite is product alignment across the DNOs, which goes beyond the alignment of terminology, but also relates to the same use of the products, especially with relation to notification period, utilisation frequency, utilisation duration, type of remuneration.

➔ For the Sustain and Secure Scheduled products, we recommend to further gain experience and standardise these products.

Also, the other products and their use may evolve over time. This may also impact the performance of certain baseline methodologies. We therefore recommend to

➔ Finetune the baseline methodology against the product design, if the latter evolves over time. Also, product design may take the impact of certain parameters on the baseline effectiveness into account, based on the principles highlighted in this document (e.g. move the notification closer to real-time).

Our assessment of the baseline methodologies has been purely qualitatively.

➔ We recommend augmenting this study by calculating the variance and bias for the proposed historical baselines for a large, representative set of assets.

This quantification can further justify:

- The choice between the historical without SDA and historical with SDA for the products Secure Dispatched (real-time), Dynamic and Restore;
- The choice within the historical baseline methodology: using 5 or 10 prior days, excluding outliers or not (e.g. 8 out of 10 vs. 10 out of 10), creating intentional bias or not (e.g. high 3 out of 5 or average 3 out of 5);
- The minimum requirements that should be set on bias and variance, either on asset or portfolio level; and
- Demonstrating that recommended/selected baseline methodologies show inclusivity for all current technologies, or adding another baseline methodology (e.g. regression) to obtain this inclusivity.

In the coming years, a greater number of new technologies is expected to seek to participate in DSO products, not all of which may be suitable under the recommended baseline methodologies. We recommend assessing the suitability of the chosen baseline methodologies for a new technology when a provider wishes to include this technology in its portfolio. Should the current baselines methodologies prove not sufficiently inclusive, then a technology-specific baseline methodology can be trialled for this technology and added to the list of available/allowed baselines methodologies.

APPENDIX 1 – OUR APPROACH

The following figure illustrates our approach in developing our recommendations.

Figure: DNV GL's approach to the assessment of baselining methodologies

APPENDIX 2 – OVERVIEW OF BASELINE PRACTICES

The overview of baseline practices will be provided as a separate spreadsheet.