GB scheduling and dispatch – A case for change

A report to ESO

MAY 2024

AFRY
ÅF PÖYRY

AFRY is an international engineering, design and advisory company.

We support our clients to progress in sustainability and digitalisation.

We are 17,000 devoted experts within the fields of infrastructure, industry and energy, operating across the word to create sustainable solution for future generations.

Making Future

CONTACT DETAILS

Stephen Woodhouse
stephen.woodhouse@afry.com
+44 7970 572444

Kostas Theodoropoulos
kostas.theodoropoulos@afry.com
+44 777 25 36 721

Romain Bourdette
romain.bourdette@afry.com
+44 738 4215983

AFRY Management Consulting provides leading-edge consulting and advisory services covering the whole value chain in energy, forest and bio-based industries. Our energy practice is the leading provider of strategic, commercial, regulatory and policy advice to European energy markets. Our energy team of over 250 specialists offers unparalleled expertise in the rapidly changing energy markets across Europe, the Middle East, Asia, Africa and the Americas.

Important

This document contains confidential and commercially sensitive information. Should any requests for disclosure of information contained in this document be received (whether pursuant to; the Freedom of Information Act 2000, the Freedom of Information Act 2003 (Ireland), the Freedom of Information Act 2000 (Northern Ireland), or otherwise), we request that we be notified in writing of the details of such request and that we be consulted and our comments taken into account before any action is taken.

Disclaimer

While AFRY considers that the information and opinions given in this work are sound, all parties must rely upon their own skill and judgement when making use of it. AFRY does not make any representation or warranty, expressed or implied, as to the accuracy or completeness of the information contained in this report and assumes no responsibility for the accuracy or completeness of such information. AFRY will not assume any liability to anyone for any loss or damage arising out of the provision of this report.

TABLE OF CONTENTS

1 Executive summary

1.1 Purpose of the project

ESO is concerned that the current GB Dispatch Mechanism design, including the Balancing Mechanism, is not working as intended.

AFRY was appointed by ESO to support this process, and in particular to help:

- gain a better understanding of current limitations or inefficiencies of the GB scheduling and dispatch design; and
- determine whether there is a case for change to the current market arrangements.

1.2 Key messages

The GB scheduling and dispatch arrangements were designed for a very different power system than the one we have today – in the past generation consisted mostly of large controllable thermal units, there were limited weather-variable RES and embedded units and the demand was broadly inelastic. A lot has changed over the years, and the market design has also undergone various incremental changes to adapt to a very different resource mix.

There is now a case for further, and potentially more radical, change of the status quo. We have identified limitations in the current market design that make market participation more complex and challenge the efficient operation of the system:

- the energy markets do not provide scheduling incentives in line with system needs and operational requirements;
- incomplete ESO visibility of market outcomes and limited access to some resources impacts coherence between wholesale market and balancing; and
- the current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time.

1.3 What has changed since the introduction of NETA?

The GB electricity market has undergone profound change since its introduction in 2001. The intent behind the design was that electricity could be freely traded as a commodity on wholesale markets at portfolio level, with the realities of system operation dealt with by the System Operator outside the core markets. The balancing window was initially the last 3½-4 hours before delivery, quickly afterwards changing to 1-1½ hours. Instructions and transactions that take place in the Balancing Mechanism have a more 'bespoke' nature when compared to trades of electricity in a power exchange.

There was never perfect separation between wholesale markets and system operation – many system instructions have implications beyond the balancing window and the system operator is ultimately responsible for energy balancing – but the intended role of the System Operator could legitimately have been termed 'residual'.

Over time, the scale and impact of System Operator balancing actions has increased, both in terms of the number of instructions and the total MWh of redispatch (compounded by the decreasing national demand over the same time frame, exacerbating the share of redispatch as proportion of national demand). The development of intermittent Renewable Energy Sources (RES), the retirement of coal and nuclear and the changing operating patterns of gas generation as a complement to RES (moving from baseload to more flexible generation patterns) have altered the operation of the system, while increasing the significance of ESO decisions affecting synchronisation or desynchronisation of controllable assets for a range of different system needs, including inertia provision, reactive power and securing sufficient operating reserve. Increasingly variable demand and RES generation patterns need to be complemented by flexible provision. The deployment of storage at scale allows for 'energy shifting' and helps manage frequency. Further interconnection is being added permitting export of RES generation at times of abundance and import of cheaper electricity from neighbouring countries when GB RES production is low.

Given this backdrop, new system needs have emerged, and ESO has developed new ancillary service products to meet them. Thermal network constraints have become prevalent, largely driven by government policy decisions which have allocated access to generation (especially in Scotland) far ahead of network build, and large interconnector projects. Meanwhile, patterns of demand and the share of generation resources at the distribution level are changing.

1.4 Assessment principles for the scheduling and dispatch arrangements

We have used a set of objectives as guiding principles when investigating the performance of the status quo, and these have informed our thinking throughout our analysis. This assessment framework can then also be used as ESO makes progress with the thinking on how the status quo should be adapted.

The overarching objective for the scheduling and dispatch design should be to **facilitate secure operation of a net zero electricity system and drive value for consumers**.

We believe the following aspects are important when assessing the performance of the scheduling and dispatch arrangements:

- **dispatch efficiency**;
  - Are the right resources dispatched to maximise socio-economic welfare?
- **efficient investment incentives**;
  - Does remuneration across the entire set of markets reward the characteristics which are needed to meet system needs in the context of a net zero power system, and does it provide efficient entry and exit signals for different asset types, including those required for flexibility and other system requirements?
- **appropriate risk and cost allocation**;
  - This includes distributional impacts (e.g. between consumers and producers) as well as whether risks are borne by the parties best able to manage them;
- **competition and level playing field**;
  - Does the design promote competition among participants, enabling all resources to contribute based on their capabilities and keep gaming opportunities to a minimum?;
- **adaptive**;
  - Is the design robust to change and can it be adapted in a straightforward and cost-effective way?;
- **transparent, replicable and auditable**;
  - Can participants explain and predict choices made by the ESO?
  - Does ESO have the right type and detail of information and at the right time to fulfil its role?

All the above have informed our thinking whilst looking at the current scheduling and dispatch design.

1.5 What are the limitations of the status quo scheduling and dispatch arrangements?

The underlying causes for ESO to take balancing actions are:

- energy balancing to solve market-level imbalances;
- to manage thermal network constraints;
- to ensure sufficient regulating reserve on a continuous basis; and
- to provide other (ancillary) services which include frequency response, inertia and voltage management.

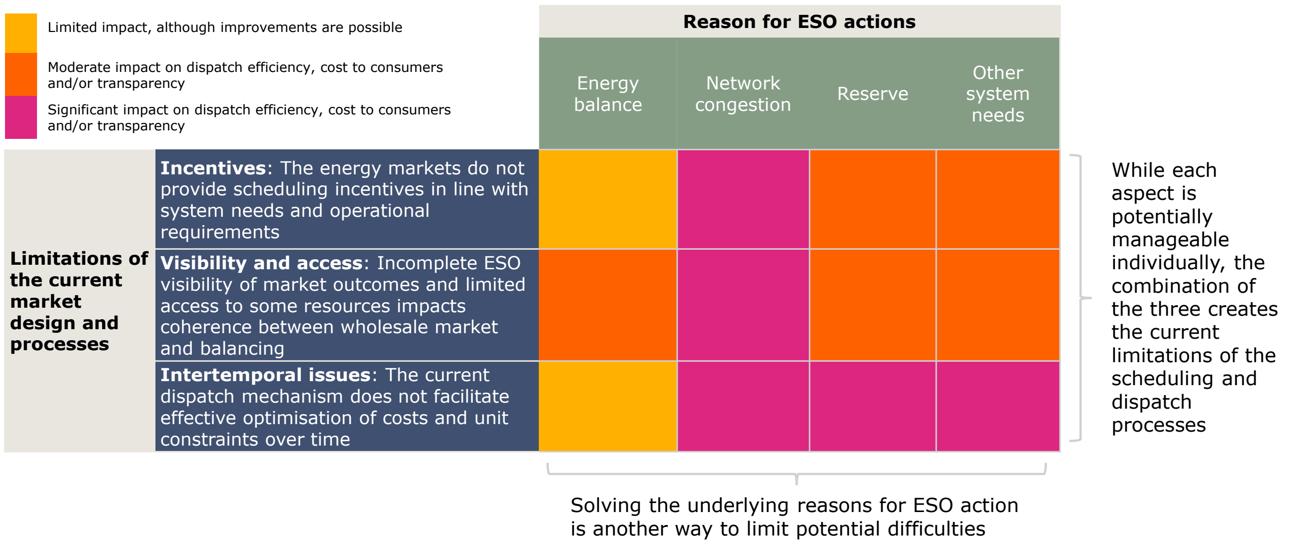
In addition to network capacity challenges, we have identified the following underlying barriers to achieving optimal dispatch outcomes in terms of efficiency:

- **incentives**;
  - the energy markets do not provide scheduling incentives in line with system needs and operational requirements;
- **visibility and access**;
  - incomplete ESO visibility of market outcomes and limited access to some resources impacts coherence between wholesale market and balancing;
- **inter-temporal issues**;
  - the current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time.

These limitations are summarised alongside the underlying reason for ESO actions in Exhibit 1. The colour-coding attempts to highlight the impact of these limitations and the underlying drivers in terms of efficiency of dispatch and cost to consumers. We see some clear trends:

- network congestion is an important source of balancing action needs; and
- the way intertemporal issues are dealt with is also an important source of inefficiency.

Exhibit 1- Key limitations with current scheduling and dispatch market design



Note: in this table 'Reserve' means ensuring sufficient reserve and response capacity; 'Other system needs' cover needs such as inertia and voltage needs

The volume of balancing actions for system constraints and reserve is now significantly greater than the volume of pure balancing energy actions only (~15% of the overall volume of positive balancing actions was for energy balancing in 2022). In isolation, factors requiring a large volume of balancing actions need not be a problem despite the 'residual balancing' philosophy if the balancing activities do not undermine the functioning of the energy markets. However, there is greater potential for inefficient dispatch in the presence of large volumes of balancing actions.

As originally conceived, self-dispatch presupposes a reactive balancing approach with actions taken predominantly within the balancing window and with limited (or no) consideration of inter-temporal costs. The submission of Bids and Offers and PNs from units for later periods are not final, and the underlying cost components of unit commitment and other inter-temporal decisions (e.g. for storage assets) can only be implicitly reflected in the simple bid-offer data structure and supporting technical offer data. ESO, therefore, does not have information or optimisation tools which permit optimal dispatch beyond the balancing window, as scheduling is meant to be performed by the market in a self-dispatch system.

With the growth in flexible resources outside the BM, ESO now faces a growing level of uncertainty and limited level of access to a portion of the controllable units on the system. Demand is becoming more elastic and the need for accurate PNs from suppliers is becoming increasingly important. Most distribution-connected resources do not participate in the BM and are not required to submit PNs. The control room does not 'see' and cannot access these potentially cost-effective sources of flexibility[1]. Operational difficulties also arise from the uncertainty around interconnector schedules due to their large installed capacity and their fast reaction to market prices in both GB and interconnected markets.

Given the changes in the power system – more RES, more storage, more distributed generation and more interconnection – ESO has to increasingly act as a 'central scheduler' in a market environment designed for a 'residual balancer'. While the need for balancing actions grows, ESO faces an increasing level of uncertainty and variability, compounding the difficulty and the potential for inefficient decisions. Additionally, the BM was not designed to accommodate forward-looking decisions in an effective or transparent way.

1.6 What are the impacts of the key limitations on system operation and market functioning?

Below we highlight specific manifestations of the key limitations and the resulting impacts on market participants (and the wider market functioning) and on the operation of the system.

[1] Non-BMUS can participate in the ESO-organised STOR and frequency response markets, but participation is voluntary, and these assets are not accessible if they do not participate in these markets

Exhibit 2 – Manifestations and impacts of the limitations of the current market design and processes

Limitations of the current market design and processes	Specific manifestation of limitation
Incentives	**'Unconstrained' market incentives:** incentive provided by national imbalance price does not align with network constraints and other system needs.
	'National' imbalance price: portfolio level balancing and national imbalance price lead to dispatch/NIV chasing in 'wrong' location
	Potential missing signals for real time reserve procurement: market is not incentivised to provide reserve capacity where and when needed. Market participants may not always be rewarded for the value of the flexibility they provide.
Visibility and access	**Incomplete coverage:** coverage of FPNs is incomplete, particularly for the growing share of flexible non-BM resources, meaning ESO has limited visibility of full market schedules when doing contingency planning
	Inaccurate information: schedules change significantly before gate closure meaning ESO decisions are taken with inaccurate information
	Behaviour: uncertainty on the expected level of system support balancing by flexible non-BM resources (e.g. NIV chasing or response to retail tariffs)
	ESO access to resources: key resources respond to wholesale market signals but are not dispatchable by ESO in balancing timeframes
	Coordination: sequential procurement of balancing services adds uncertainty to decision making for both ESO and market participants
Intertemporal issues	**Timing:** ESO is obliged to take proactive decisions with consequences for future periods beyond Gate Closure, which overlaps with the operation of the intraday market
	Information: ESO takes decisions with inter-temporal consequences based on incomplete data
	Transparency: beyond-the-wall actions and advance commitments cloud transparency and may distort imbalance pricing, making it harder to interpret and forecast by market participants

These limitations challenge both market participation and system operation.

The incentives in the current market design can:

— give rise to greater volumes of balancing actions than could be necessary, increasing costs to consumers;

— result in misallocation of flexible resources (in both investment and operational timeframes); and

— reduce transparency on what is an energy action and what is an action for reserve, limiting the understanding of the underlying value by market participants.

The somewhat limited visibility for ESO and the market and incomplete access to resources, in addition to reducing liquidity, appears to:

- lead to over- and under-procurement of energy and reserve by ESO;
- result in inefficient dispatch decisions with, at times, unnecessary actions for risk mitigation;
- create operational difficulties for ESO and use of expensive actions given the need to react fast to large changes; and
- leave market players facing conflicting incentives and increase the potential for misallocation of resources when bidding in the different markets.

Finally, the lack of effective optimisation of costs and unit constraints over time means that:

- market players can face conflicting incentives with a lack of coordination between ESO actions and market scheduling decisions;
- there is potential for energy-limited and other flexible resources to be underutilised; and
- incentives for market participants to support system energy balance are dampened.

The impacts on market participants, wider market functioning and system operation are further discussed throughout this report.

1.7 Is there a case for change?

Our investigation suggests there is a clear case for change of the status quo as the underlying conditions have changed since NETA was introduced. The current limitations impact both market participants and ESO, and potentially result in inefficient dispatch:

- **market participants do not have appropriate incentives to allocate resources to meet system needs**;
  - the market attempts to balance supply and demand and market participants do not have appropriate incentives to solve network constraints or other system needs;
  - there is no formal recognition of the option value of reserve for ‘using later’ and some providers are – at times – used as ‘free’ reserve; and
  - the BM was not meant to provide forward-looking signals but to reflect the more reactive ‘just in time’ needs for balancing supply and demand; and
  - actions taken for proactive redispatch end up influencing the Imbalance Price, leading to poor interpretability and predictability of the market–.
- **visibility of the market is incomplete and ESO cannot access some resources for balancing purposes;**
  - incomplete ESO visibility of market outcomes impacts coherence between wholesale market and balancing;

  - there are weak incentives for accurate and timely information sharing and this unnecessarily complicates system operation and potentially results in inefficiencies; and
  - ESO cannot use some of the resources in the balancing timeframe- these resources include interconnectors and smaller units[2], which may prove to be more cost-effective sources of flexibility; and
- **the current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time;**
  - many balancing actions – including those to resolve system or energy balance – require ESO to take advance decisions outside the immediate balancing window for which it has limited information, and for which the market framework is inappropriate;
  - with a greater need for ESO-instructed synchronisations and more energy limited units on the system (which in turn arise from the scale of network congestion and the need for improved incentives on participants), it could be argued that the process to optimise intertemporal constraints and costs could be improved – this could also include changes which reduce the extent to which ESO must manage these issues in its scheduling activities .

The market design is not static, and many changes have been made since NETA Go-Live in 2001. There are some changes underway that will help improve scheduling and dispatch. These are summarised in Exhibit 3.

Exhibit 3 – Limitations managed with the "to be" arrangements

	Recent or planned changes potentially mitigating identified issues
Incentives	– **Ongoing network capacity expansion** – **Thermal Constraints Collaboration Project** – **Balancing Reserve will pre-contract some resources to provide reserve availability** – **Half-hourly settlement** – **Ofgem compliance engagement with storage regarding Transmission Constraint Licence Condition (TCLC)**
Visibility and access	– **GC117 proposal to reduce BMU threshold to 10MW** – **Local constraint market (pilot for B6) could allow ESO access to more resources** – **DER/CER Visibility project**
Intertemporal issues	– **Balancing reserve will reduce the need for pro-active scheduling actions in the BM** – **Potential submission of data on energy limited units** – **Ofgem inflexible offers licence condition**

[2] Non-BMUS can participate in the ESO-organised STOR and frequency response markets, but participation is voluntary, and these assets are not accessible if they do not participate in these markets

The ongoing and planned changes presented in Exhibit 3 can help with some of the limitations of the "status quo", but may not be sufficient to ensure that the scheduling and dispatch market design is fully adapted to the realities of the future power system. ESO's role has recently been extended to develop a Strategic Spatial Energy Plan (SSEP) and a Centralised Strategic Network Plan (CSNP). This whole-system strategic planning will help coordinate generation and transmission infrastructure investments in time and location. Given the size of the challenge, we think that strategic planning and market signals should complement each other to provide an efficient outcome for consumers...

What is less clear is what a change to the status quo would look like. We see two high-level approaches:

- giving market participants better incentives and better information to support system operation, which could include some or all of the following:
  - shorter imbalance settlement intervals;
  - smaller zone sizes;
  - improved signals for ancillary services;
  - increase participation in the BM;
  - improved information sharing between market participants and ESO;
- formalising ESO's de facto role by giving greater control at an earlier stage; effectively allowing ESO to coordinate unit commitment and operation of energy-limited units, as well as within-day positions.

In the context of the investment needed to deliver on the decarbonisation agenda, the choice between the two is finely balanced. These two high-level options have complex trade-offs, and need to be considered as part of the wider ongoing Review of Electricity Market Arrangements.

2 Introduction and scene setting

2.1 Introduction

2.1.1 Structure of this report

This report is structured as follows:

- Section 2 describes the scope of this project and sets the scene by introducing some of the key scheduling and dispatch concepts;
- Section 3 provides an overview of the current GB scheduling and dispatch market design and discusses the impact of the energy transition on system operation;
- Section 4 summarises the methodology we have used for our analysis and the assessment principles;
- Section 5 details the limitations of the current scheduling and dispatch market design and drivers behind these; and
- Annex A includes the results of the quantitative assessment of selected historical days used to highlight the limitations with the current arrangements.

2.1.2 Sources

Unless otherwise attributed the source for all tables, figures and charts is AFRY Management Consulting.

2.2 Overview of the project

The objective of this project is to gain a better understanding of current limitations or inefficiencies of the GB scheduling and dispatch arrangements and to determine whether there is a case for change to alternative market arrangements.

This project encompasses several aspects:

- description of the current ESO scheduling and dispatch processes,
- identification and prioritisation of key limitations of the current scheduling and dispatch market design and processes,
- assessment of the ESO actions for specific key study days and events in terms of their efficiency and impact on wider market outcomes. This

includes the review of the potential impact of balancing actions taken on the intraday market (in terms of price formation and liquidity), and;

- quantitative assessment of theoretical alternative scheduling and unit commitment options based on modelling of actual historical days when events of interest have taken place (key study days).

2.3 Scene setting

2.3.1 What is scheduling, unit commitment and dispatch?

'Scheduling and dispatch' describes the process undertaken by System Operators and/or market participants to maintain the balance between supply and demand in a manner that ensures the safe and secure operation of the power system. The process usually spans multiple time periods to ensure secure operation of the system not just in real time but also in the subsequent hours. The outcome of the process is a series of actions taken by, or instructions issued to, resources to meet specific operational setpoints.

There can be confusion with the use of the different terms – unit commitment, scheduling, and dispatch – and these are at times used interchangeably. To aid understanding we make the following distinctions:

- **Unit commitment** is the process used to determine when and which units are 'on' or 'off' (effectively defining start-ups and shutdowns)[3];
- **Scheduling** is the expected profile of MWh production/consumption of each unit at different points in time. It can mean different things depending on the context and market design – in some cases scheduling refers to the 'market schedule', which is the unit commitment and intended production levels as determined by the market, and in other cases scheduling means the System Operator schedule which takes account of the market schedule and amends it to ensure secure operation of the power system planning of flows on the network, effectively encompassing the unit commitment and dispatch as determined by the market and/or the SO.
- **Dispatch** is the decision process of how much each unit should be outputting or withdrawing in real time (usually resulting in issuing dispatch instructions).

It is also important here to clarify one fundamental difference in terms of how a System Operator and a market participant view a 'schedule' and the resulting dispatch. A System Operator has an obligation to maintain system frequency within a certain threshold, and as such is interested in profile of active power output (MW) from generating units at any given time. Market participants, on the other hand, see the market schedule as active energy (MWh) to be delivered in any given settlement period.

[3] The notion of unit commitment is evolving with the change in technology mix. While it traditionally mainly considered start-up costs for thermal units, it must be noted that unit commitment should now cover a wide set of inter-temporal factors (cost and constraints, including storage levels for storage technologies).

2.3.2 What are the options for scheduling?

A usual way to describe scheduling regimes is to differentiate between central dispatch and non-central dispatch systems (also called self-dispatch). Our view is that this general categorisation fails to capture the nuances that can exist between more and less centralised market and dispatch arrangements. In practice, in all markets the System Operator will instruct some or all the units on the system at some point in time. The degree of centralisation of the dispatch regimes can be seen as a spectrum of when and how the System Operator controls units, rather than separate distinct options.

However, understanding the fundamental difference between central and non-central dispatch regimes is still useful to frame the discussion in this report, and also because ESO appears to be increasingly operating as a 'central scheduler'. In broad terms:

- in a central dispatch system, unit commitment, scheduling and dispatch instructions are generally issued by one (or several) central clearing algorithms managed by the System (or Market) Operator. All dispatchable generators must maintain their position (both unit commitment and output) until they receive a dispatch instruction from the SO to move to a different position. In such a centralised system, units may have the possibility to self-commit and self-schedule, but the final dispatch decision for (large) units will lie with the SO.
- in a non-central dispatch system, generators will sync/de-sync and increase or decrease their outputs in accordance with their Physical Notifications unless they receive a SO dispatch instruction to do something else.

While the exact implementation of the dispatch regime varies across markets, in a general sense, 'centralising' means that the SO has greater control over units' position and output, and at an earlier point, compared to non-central, self-dispatch arrangements.

In practice, each unit follows a sequence of availability, commitment, scheduling and dispatch decisions taken over time and which may cover a collection of settlement intervals. These decisions may be taken centrally by the System (or Market) Operator or by the unit operator (e.g. in response to market prices). Generally, approaches are more 'hybrid' with requirements for information sharing and the opportunity for one party to restrict or veto decisions or preferences by the other party.

2.3.3 How is electricity priced in power markets?

The simple microeconomic representation of a well-functioning market includes a conceptual separation into:

- long term decisions for investment and closures; and
- short term operating decisions.

In 'well-behaved' economic problems, marginal costs are assumed to be monotonically increasing so that a single spot price can clear the market. When there is abundant supply of goods and competition, prices should drop to the short-run marginal cost of supply. Over time, however, prices should 'trend' towards the long run marginal cost to ensure fixed cost recovery.

The application of this theory to electricity markets needs embellishment. Supply and demand need to be balanced at all times in power systems. In simple terms, the spot price in each settlement interval could be conceptualised as the marginal incremental cost of production in that interval, plus a scarcity element, assuming there is some scarcity and in the absence of perfect competition. If there was perfect competition (infinite number of generating units all under different ownership) then prices would either be at the level of the marginal incremental cost or the Value of Lost Load, and there would be no scarcity rent.

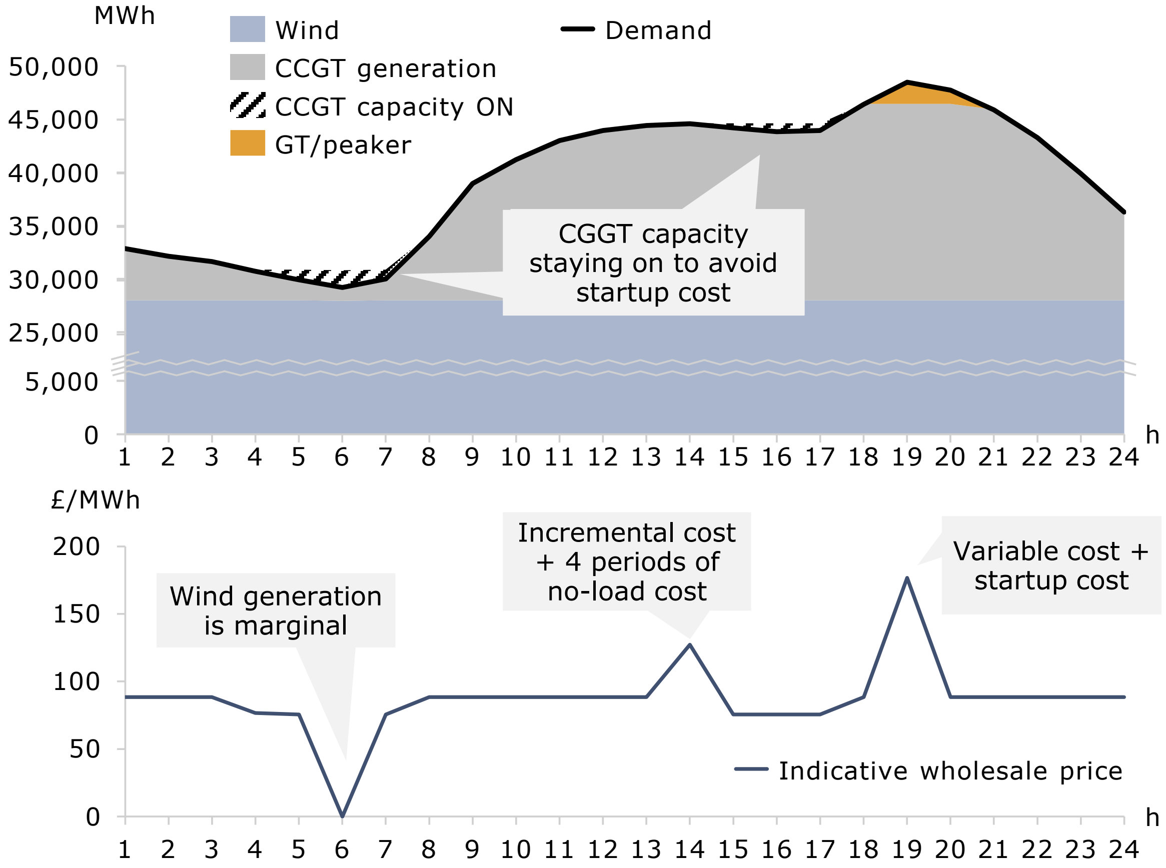
The reality is far more complex. Operational decisions are taken in sequence, with no perfectly clear 'marginal' cost or quantity increment at any point in time for energy delivered in a given settlement interval. Cost functions are not always monotonically increasing, and there are many binary decisions, non-linearities and indivisibilities in cost and availability functions which mean that production in one settlement interval is dependent on production in other settlement intervals. Once certain decisions are taken, they may become 'sunk' (i.e. unchangeable) and are no longer marginal and (in theory) able to influence price.

In any market, the SO has ultimate responsibility for balancing and may take dispatch decisions for (at least some) units closer to real time. By that time, the main binary (unit commitment) decisions may have already been taken. Pricing in such 'balancing markets' may then also include a 'flexibility premium' instead of a purely marginal cost of generation, given than competition may be lower at shorter timeframes. This is because there may be a more limited space of solutions closer to real-time and providers can price in a flexibility scarcity premium.

SOs will need to look beyond the immediate balancing window based on the submitted dynamic data for the relevant units, particularly looking forward to future peaks and sharp ramps. There may be a need to issue warming contracts, to synchronise plants early or to keep them from desynchronising now because of a later need. Different SOs, however, do have different degrees of freedom on what they can and cannot do. The calculation of imbalance prices typically does not allow allocation of costs incurred in one settlement interval to the imbalance price in another interval.

Exhibit 4 – Example of price formation within day

In this illustrative example we assume 28GW of constant wind generation with the remainder of the demand met by CCGT generation. Two CCGTs choose to stay online and operate at their minimum stable in period 6 (and part-loaded slightly above their minimum stable generation in periods 5 and 7) to avoid an additional start in the morning. The price in period 6 is set at the level of wind (£0/MWh assuming no foregone support payments and no variable costs) as the CCGTs offer their output at the minimum load level at a negative price, reflecting the avoided start-up opportunity cost. A peaking unit is needed for the evening peak, incurs a start and internalises this cost in the single period of operation.



Source: AFRY Management Consulting

There are complex interactions between energy delivered in different settlement intervals, both in physical and cost terms. This means that the prices in nearby settlement intervals are strongly co-dependent, and there may be no unique solution to (marginal) pricing.

Further, there are many constraints on system operation which may not always be reflected in the energy market definitions – notably, the profile within a settlement interval, the geographic location within the market price area, and the need to provide other services from the same resources that are providing energy, some of which are complementary (e.g. energy plus X, e.g. reactive power), and some of which are competing (energy or X, e.g. headroom to provide reserve). The interaction between energy and these secondary products is itself multi-layered. For example:

- upward reserve can be provided from synchronous generator units only if they are both synchronised and part-loaded; and
- inertia and reactive power from synchronous generators can only be provided if they are 'on'.

Within the operational timeframe (after investment and closure decisions), the sequence of decisions begins with choices about the timing of scheduled maintenance for network and generation assets. Closer to the operational day, decisions are taken about the timing of availability of capacity. Many larger (thermal) units have significant lead times for synchronising or desynchronising generation to the system, with other restrictions on flexibility such as notice periods, minimum on/off periods, and ramping constraints which in turn vary according the warmth state of the plant (i.e. when it was last operated). There are costs associated with these 'integer' decisions: costs to start, and minimum costs to operate in addition to the per-MWh 'marginal' production costs. The operational schedule of each asset will be the result of this cascade of decisions. Some assets are energy limited: i.e. a pattern of generation (or consumption) at a time has consequences for its generation (or consumption) earlier or later (e.g. demand response to defer or bring forward an action, or a battery or other storage plant).

2.3.3.1 Managing imperfections in self-dispatch power systems, e.g. GB power system

As noted above, the supply curve for the power system (and also the demand curve) does not conform to microeconomic theory of a well-behaved market, and there are many 'imperfections'.

The GB and European market designs deal with these imperfections in an informal way. The spot markets are organised around a day-ahead auction complemented by further auctions, and a continuous traded market which begins before the day-ahead market and closes shortly before gate closure.

Underlying the GB and (most of-) European decentralised market design is a concept of balance responsibility. Each market actor faces responsibility for its metered quantities (demand and generation) and it is generally expected to meet these volumes from within its portfolio or through trading. Each participant may have individual imbalances, but these net out, and the system operator needs to deal with the overall system position not the individual components. The ultimate incentive on each participant to balance is that if it fails to do so, its energy imbalance will be settled at a price which is only known after the event.

There are two different approaches for settling imbalances. This can be done on a portfolio basis or at the unit level. Generally, in Europe, portfolio balancing has prevailed with Balance Responsible Parties (BRPs) attempting to balance the portfolio they are managing. SOs, however, are more interested for in balancing at a unit level. An SO needs to manage the flows on the network and have the right type of and sufficient ancillary services to operate the system securely. Location and knowing which unit is outputting and at what level is therefore critical.

Collectively, this design incentivises the market to act to support the system in terms of balancing supply and demand for a given settlement period (but not within the settlement period). The conservative trading position is to closely match the physical position with the (pre-) contracted position (at known prices). In GB there is no obligation for participants to submit a balanced position (they are entitled to deliberately be long or short provided that they submit accurate Physical Notifications). In other markets there are stronger or weaker rules around the participants deliberately taking imbalance positions.

Day ahead auction

The GB day-ahead EPEX auction closes at 09:20 GMT day-ahead and NordPool day-ahead auction closes at 9:50 GMT (12:00 CET for the EU markets) for hourly contracts for the following calendar day. The market uses the EUPHEMIA algorithm and includes a number of more 'complex' products (e.g. block orders, linked block orders, minimum income condition etc.), which allows some of the inter-temporal complexities and risks for large units to be managed. The algorithm determines both the cleared trades and also the resultant prices. Post Brexit, there is no implicit coupling with the rest of Europe at these two GB day-ahead auctions. In recent years, approximately a third of the physical demand was traded on the day-ahead markets in GB, making it a key starting point for the GB scheduling process.

The definitions of block products in the day-ahead auction represent the most important commercial and physical limitations on operation of large thermal generator units. These order types enable operators to ensure that the full set of operating costs for a unit are covered before the trade can be matched. However, block bids are excluded from pricing, meaning that only 'marginal' incremental decisions on the level of output may set prices. This is a weakness of the pricing algorithm: it might lead to under-pricing in the day-ahead auction, as units or portfolios which use these order types to cover start-up and no-load costs will be excluded from pricing, and that only the pure incremental cost of increasing or decreasing production at the margin (excluding start and no load costs) may set price[4].

EUPHEMIA operates at a portfolio level in GB and most European markets, but at a unit level in some European markets. The risk management offered by the block bid structures is imperfect: it allows some but not all of the characteristics of large thermal units to be represented within the scheduling and pricing optimisation. Portfolio operators which have diversity in their fleet have greater ability to manage inter-temporal issues and uncertainty – e.g. through staging the start and shutdown of units and gaining flexibility through part-loading – compared with operators of single units or units with similar characteristics.

EUPHEMIA is a complex algorithm that allows for advanced bidding products and different network models. While it is robust, and faced only very few

[4] Note that this is a weakness that more centralised pricing and scheduling algorithms share.

incidents, the algorithm limitation has mainly been the run-time. For instance, the auction run time had to be increased by a few minutes in recent years to meet the increased complexity and amount of orders/borders being processed (from a total computation time of less than 30 min). The algorithm is constantly evolving to accommodate more markets and to improve its efficiency.

Intraday auctions

After the GB day-ahead market there is a further day-ahead auction (closing at 15:30) and two ‘intraday’ auctions which are coupled with Ireland (closing at 17:30 day-ahead and 08:00 on the day of delivery), each trading simple 30-minute contracts. These intraday auctions are also cleared with EUPHEMIA algorithm, and they support the same type of complex block orders as the day-ahead market.

Continuous traded markets

For the 30-minute contracts in each calendar day, the continuous market commences at 00:00 day-ahead and runs until shortly before gate closure; i.e. up to 48 hours in advance. This market allows individual settlement intervals (or combinations thereof) to be traded under simple price-quantity offer structures.

Imperfections in the supply curve are dealt with over time through market participants trading and reacting to movements in the market prices in the various spot markets. The spot market time horizon – approximately 2 days – is itself reflective of the need to make unit commitment decisions for large thermal power plants which may take up to 12 hours to start and which may need to operate for a day or more once started (whether for commercial or technical reasons).

Participants typically use the spot markets to fine-tune their contractual and intended production decisions. Before the spot timeframe, forward contracts tend to have very simple contract structures: baseload, 12-hour (07:00-19:00) or other patterns built from 4-hour blocks. As the daily and within spot markets progress, participants may trade for shorter intervals to reach a traded position that suits their intended production (or consumption) pattern, making adjustments as circumstances change. In the continuously traded markets, prices act as the main medium for adjustment of the scheduling decisions. Participants may change their plans depending on the set of prices that they can see and at which they can contract. At any time during the (up to) 48-hour trading window they may change their traded position, locking in new firm trades with known prices. Supply and demand respond to the movements of prices, with an incentive on both buyers and sellers that if they fail to match their physical and traded positions then they will be exposed to (uncertain) energy imbalance pricing.

Example of trading in a decentralised market

Imagine that a large generator with a six-hour minimum run period had not been committed at the day-ahead, but was needed to meet demand in two peak hours in the evening. In this example, assume that demand and supply in adjacent settlement periods had been matched with least cost generation, but that the least cost solution requires the additional large generator to run for the full six hours and that other flexible generation should reduce output for four of these hours to make space.

The market offers and bids for this time window in the continuous market would adjust. For the market to clear, the large generator would need to earn enough money in a period of (at least) six continuous hours to cover its costs, including start and no-load costs. If it sold energy for these six hours, then there would be a market surplus of energy in four of these hours (from the balanced position previously reached in trading) and the prices in the shoulder periods would fall relative to the peak two hours. There might also be some further trading around the shoulder hours before and after the peak as other capacity adjusted its schedule. There is no guarantee that the perfectly optimal outcome would be achieved (and there may be a reliance on ESO to resolve the overall energy balance) but in principle the market actors could seek overall balance through trial and error until Gate Closure.

By allowing the asset operators to control their own dispatch, a wide set of asset-specific considerations (and market imperfections) can be taken into account, provided that there is enough liquidity in the market and that trades can take place before irrevocable decisions are taken. The prices over a period for which a unit (or portfolio of units) is operated would need to cover all of the relevant costs incurred in that period including start-up and no-load costs for large thermal units. (as noted above, this may not be the case in the markets which rely on more complex offer structures, in which complex orders are excluded from market pricing.)

Liquidity and transparency of price formation (including understanding of how the price is formed) is necessary to ensure market efficiency in a decentralised market. It can also be argued that better visibility of what will eventually be dispatched, and introducing incentives for market participants to bridge the gap between 'unconstrained' market positions (which ignore network constraints and other system considerations) and a constrained schedule (accounting for all system needs and constraints) can help with dispatch efficiency..

More flexible assets should be able to take advantage of trading close to real time against more volatile prices, and this might enable them to earn a premium for their flexibility, provided that there is adequate liquidity closer to real time. It can also be noted that is generally easier for portfolio players to understand price formation and to manage risk within a decentralised system, in which they are able to take decisions on the basis of known prices. The GB market design allows market participants to see the prices for series of settlement intervals, as published through the EUPHEMIA algorithm, before committing to start and operate within that time window. No algorithm is needed to allocate the start costs to the price in particular hours (or to pay these costs as side payments): the market interactions themselves determine the price patterns.

2.3.3.2 Managing imperfections in a central dispatch power system

Central dispatch systems typically use market clearing algorithms to determine unit commitment and a 'market schedule', which minimise the production cost (or maximises the social welfare created) over an optimisation horizon, whilst meeting a set of constraints. The market scheduling can reflect the physical realities of electricity production and transportation to differing degrees.

No market scheduling and pricing algorithm can include all constraints and realities of operating an electricity system. The uncertainty SOs face at the time of dispatch is difficult to reflect in the market schedule whenever it is created. Complex market schedules are run at specific points in time and include deterministic assumptions regarding the expectation of demand and plant availability, or with ex-post scheduling have the benefit of hindsight. Differences will always exist even with the most sophisticated algorithms. The constraints used can include:

- technical generating unit parameters (minimum on times, minimum off times, minimum stable generation etc.);
- network constraints (either through zones in the form of NTCs or nodal representation); and
- additional system services constraints/procurement.

Centralised market clearing algorithms also allow for a more detailed representation of the cost structure of generating units, splitting their costs into:

- start-up (or shut-down) costs;
- an incremental offer curve with different prices (heat rate) at different loading levels; and
- no-load costs.

These market clearing algorithms are typically 'run' at the day-ahead stage, and there are then additional unit commitment and/or 'dispatch' runs closer to real-time. In a lot of cases, these within-day runs, however, do not deal with unit commitment variables and solely increase and/or decrease output from already committed units.

Pricing with market clearing algorithms has always been a topic of discussion – there is no unique solution on how to deal with non-linearities. The unit commitment and scheduling is typically solved using Mixed Integer Programming with pricing determined at a second stage and once the integer variables have been 'fixed'. This then means that resulting price is set at the level of the marginal incremental cost (or offer price of the most expensive offer to meet an additional increment of demand). The recovery of start-up and no-load costs can be done in the following ways:

- make-whole payments (termed uplift in the US markets);
  - with this approach generators receive a top-up payment, if and only if the energy market revenues do not ensure full cost recovery;

- targeted cost recovery (this is not so common, but is fairly easy to implement and was used at some point in the past in the Greek market when it was operating a pool);
  - this approach means more 'targeted' payments with each generator recovering exactly its start-up and no-load costs in addition to the revenues realised from capturing the market price;
- uniform uplift adder (used in the SEM pool);
  - this approach allows for all generators to capture inframarginal rent relating to start-up and no-load costs with costs relating to start-up and no-load feeding into the SMP (in addition to the marginal incremental cost).
- implicit recovery by allowing units to bid freely (including above variable cost);
  - in the old Greek pool, units were mandated to bid above the minimum variable costs (and by definition above the incremental cost) – this then meant that no-load (or minimum load) cost recovery was ensured.

Power market clearing algorithms have been designed around system with significant levels of thermal generation, and are quite effective at dealing with scheduling thermal units, accounting for incremental and quasi-fixed costs, alongside other constraints. In the end, the results of a market clearing algorithm are as good as the information it incorporates.

They are well equipped to manage intertemporal constraints and cost optimisation for thermal dominated and (in some cases) hydro-dominated[5] systems. However, the representation of storage is not that advanced, and solutions to manage energy limited units at scale are now being developed.

In theory, the outcomes produced by a central market clearing algorithm should be 'optimal' on the basis of a deterministic availability of non-controllable generation and demand. However, one of the key arguments against the use of such centralised market clearing algorithms is the typical absence of more active continuous trading to respond to changes in availabilities.

---

[5] With the use of Stochastic Dual Dynamic Programming (SDDP), but this is not the case for European markets

3 Impact of energy transformation on power system operation

3.1 The New Electricity Trading Arrangements

The current market design was implemented in 2001 as the New Electricity Trading Arrangements (NETA). NETA's philosophy was based on balance responsibility for market participants, continuous bilateral trading and self-dispatch decisions, replacing the system of centralised procurement through a pool with an optimisation. In its bilateral trading framework, the envisaged role of ESO was of a 'residual balancer'. In other words, market participants were expected to broadly balance supply and demand through trading to mitigate their exposure to imbalance prices. After gate closure (initially 3.5 hours before delivery, later moved to 1 hour) (then-NGC) was expected to adjust generation dispatch to maintain energy balance and ensure system security without significantly interfering with the market.

The Electricity Pool of England and Wales operated from 1990 to 2001 and was created at the time of industry privatisation. The market structure was initially uncompetitive. All non-nuclear generation assets were split between two generation companies, National Power and Powergen, which had control of virtually all of the price-setting generation and were able to exercise significant control over price formation, both in energy and the capacity payment mechanism. The Pool was characterised by the exercise of market power and the perception that the central algorithms for scheduling and pricing (including the capacity payment mechanism) were prone to manipulation.

A significant amount of gas-fired generation was built during the 1990s. Most of this capacity was under contract to the regional electricity companies, which bundled distribution with local monopoly of small-scale retail. For most of that period, gas was significantly cheaper than coal. Much of the gas generation was under take-or-pay fuel contracts (effectively making them a zero marginal cost generation source) and gas ran at baseload. Pool prices were generally set by coal (or oil) which operated more flexibly within the day. National Power and Powergen were obliged to divest some coal generation to Eastern Electricity in the mid-1990s but it was done under an "earn-out" levy of £6/MWh (around 25% of the market price) which effectively boosted the offer price of this capacity into the pool. Pool prices remained high under this effective triopoly. As new gas generation was brought to the system, the dominant generators National Power and

Powergen ceded volume but retained control of prices, which remained at levels significantly above marginal cost.

The early discussion of market reform by the incoming Labour government from 1997 was framed as a brake on the 'dash for gas' which was seen to undermine the British coal industry. The market review project was initiated by government to give more value for 'flexible' coal generators[6], although that issue was less important for Ofgem (initially Offer) which was asked to lead the design with limited government involvement.

Throughout the design process there was an emphasis on reducing the central dispatch role for NGC (which was described as a form of central planning) and on giving participants more control over their own assets. There was an emphasis on simple bid formats to improve transparency and to reduce the ability of generators to manipulate prices[7]. The design was significantly influenced by the Nordic market (introduced in Norway in 1996) and also the GB gas market, which was held up as a liquid, competitive commodity market in which the practicalities of physical delivery were separated from the trading of the commodity itself. Another key change was the removal of the Pool's capacity payment mechanism which had been subject to manipulation especially in the earlier years of the Pool.

Alongside the review of trading arrangements, the government implemented a moratorium on new build of gas generation (which was strongly opposed by Offer/Ofgem at the time). This triggered a frenzy of bidding for older assets and National Power and Powergen finally sold significant quantities of generation capacity at high prices, notably selling Fiddlers Ferry and Ferrybridge to Edison Mission. Edison Mission changed the Pool bidding

---

6 "The Government has become concerned that the fuel mix in generation has the potential of becoming too heavily weighted towards gas. […] Some argue that the present trading arrangements discriminate against particular energy sources (notably coal-fired plant). To the extent that this reflects a concern that flexible coal-fired plant are not adequately remunerated under present arrangements, […]", *Review of Electricity Trading Arrangements: working paper on trading inside and outside the Pool*, Offer, 29 March 1998

7 "It has also been argued that the complexity and opacity of the price setting process has inhibited the development of derivatives markets and reduced liquidity in the contracts market.[…] The greater simplicity of bidding formats in other countries seems to lead to more transparent price setting processes, which can help to promote competition, not least in the associated financial markets", *Review of Electricity Trading Arrangements: working paper on trading inside and outside the Pool*, Offer, 29 March 1998

"The Pool's bidding and price setting mechanisms are complex, reduce transparency and increase the options open to generators to achieve their commercial aims. Bids into the Pool by generators are not reflective of costs and movements in Pool prices have not matched reductions in generation costs. The complex, administered Pool capacity payments do not provide a very effective short-term signal to encourage generation and demand to respond to rapidly changing circumstances and provide a poor long-term signal for the need for capacity. More generally, the present trading arrangements have facilitated the exercise of market power at the expense of customers by enabling all generators to receive a uniform price which in practice has been set by just a few of them", *Rises in Pool Prices in July: A Decision Document*, Ofgem, October 1999 patterns of these plants, bidding much lower to increase their operating hours and triggering a collapse in prices to competitive levels.

The Pool gave an 'unconstrained' day-ahead schedule, excluding transmission capacity or ancillary services needs. The Pool schedule conferred firm access for the scheduled quantities; any redispatch which deviated from the Pool schedule was settled on a pay-as-bid basis. The schedules were determined for an entire 24 hour period, so storage plants with a 4-hour discharge time would typically be given 4 hours of generation schedule and of firm access. With NETA, the principle of firm access was applied to generators' submissions of Final Physical Notifications (FPNs). This gave rise to concerns that the new market increased access rights and consequently the costs to consumers.[8]

When NETA was being designed, a decision was taken to exclude transmission constraints from the market and to concentrate only on the pricing of energy. Constraint balancing actions are isolated from energy imbalance pricing through (manual) 'flagging'. A separate market-based regime was envisaged for transmission access[9] which was linked to the ability to shorten gate closure from 3.5 hours to 1 hour.

Ofgem intended to move to a system of tradeable access rights, initially favouring an entry/exit regime (similar to the gas market)[10]. It later abandoned these proposals, passing responsibility to NGC to define the future access regime[11].

---

[8] "A particular concern is that there is no cap on total generation access rights under NETA other than the total installed capacity on the system", *Transmission Access and Losses under NETA: A Consultation Document*, Ofgem, May 2001.

[9] "Moreover, experience with the new trading arrangements and the intended introduction of new transmission access rights designed to remove, or at least substantially reduce, the extent to which transmission constraints have to be resolved in the Balancing Mechanism, should enable Gate Closure to be shortened"

"Along with many respondents, Ofgem/DTI agree that it would be desirable to remove constraints from energy imbalance prices in the short-term, pending the implementation of a market-based approach to transmission access".

Source: *The New Electricity Trading Arrangements; Ofgem/DTI Conclusions Document*; October 1999

[10] "… Ofgem considers that the implementation of a system based on firm entry and exit access rights has considerable merits. […] Ofgem remains of the view that the auctioning transmission access rights is an efficient means of achieving these aims. Thus, one option would be to employ auctions for both entry and exit rights, with secondary trading being used by participants as well as the System Operator (SO) positions and an access imbalance regime designed to incentivise participants to match their physical positions to their access right holdings", *Transmission Access and Losses under NETA: A Consultation Document*; Ofgem, May 2001

[11] "Under the May proposals, Ofgem discussed the possibility that these rights could be auctioned to generators and suppliers. After listening to representations, Ofgem has decided not to pursue this proposal but, instead, leave it to the industry and NGC to agree the best way to allocate rights", *Transmission access and losses – conclusions*; Ofgem, February 2002.

Until 2005, Scottish generators could only access the England & Wales market by buying available capacity on the Scottish interconnector. In 2005, the market arrangements were extended to Scotland, and firm access (transmission entry capacity (TEC)) was granted to the GB market for all existing Scottish transmission-connected generation in return for paying (high) TNUoS charges. This was the start of National Grid's management of the Scotex constraint, with more Scottish generation trying to access the GB market than the available network capacity.

As renewable support policies were expanded and costs fell, connection applications for wind in Scotland grew sharply. The connection regime until 2010 prevented connections until deep reinforcements could be made to meet security standards. There were significant connection queues, and trading of transmission entry capacity was implemented (with limited success) to allow new connections. The government implemented the connect and manage regime from 2011 which accelerated development of wind in Scotland. This in turn also resulted in greater levels of network congestion.

3.1.1 Markets under NETA

The NETA programme concentrated on the development of the balancing and imbalance settlement process rather than the forward or spot markets, which were deemed to be competitive activities. No fewer than five power exchanges launched following NETA Go-Live although only three ever traded and only two survived (after various take-overs). At the beginning the spot market was a continuous market with no day-ahead auction, launched later in response to concerns about liquidity.

The imbalance settlement arrangements initially used dual pricing (with a separation between system buy price for energy shortages and system sell price for energy surpluses). Dual imbalance prices faced criticisms in GB and European circles because they fail to reward participants supporting system imbalance and over-incentivise parties to balance individually. Single imbalance pricing was introduced in GB in 2015 through a modification (P305) submitted by National Grid under direction from Ofgem.

3.2 General overview of the balancing process in GB

In the current GB market framework, market participants may adjust the traded position of their portfolio through various market routes, either centrally organised or via bilateral trading, and decide themselves on the scheduling of their assets to meet their traded position of the portfolio.

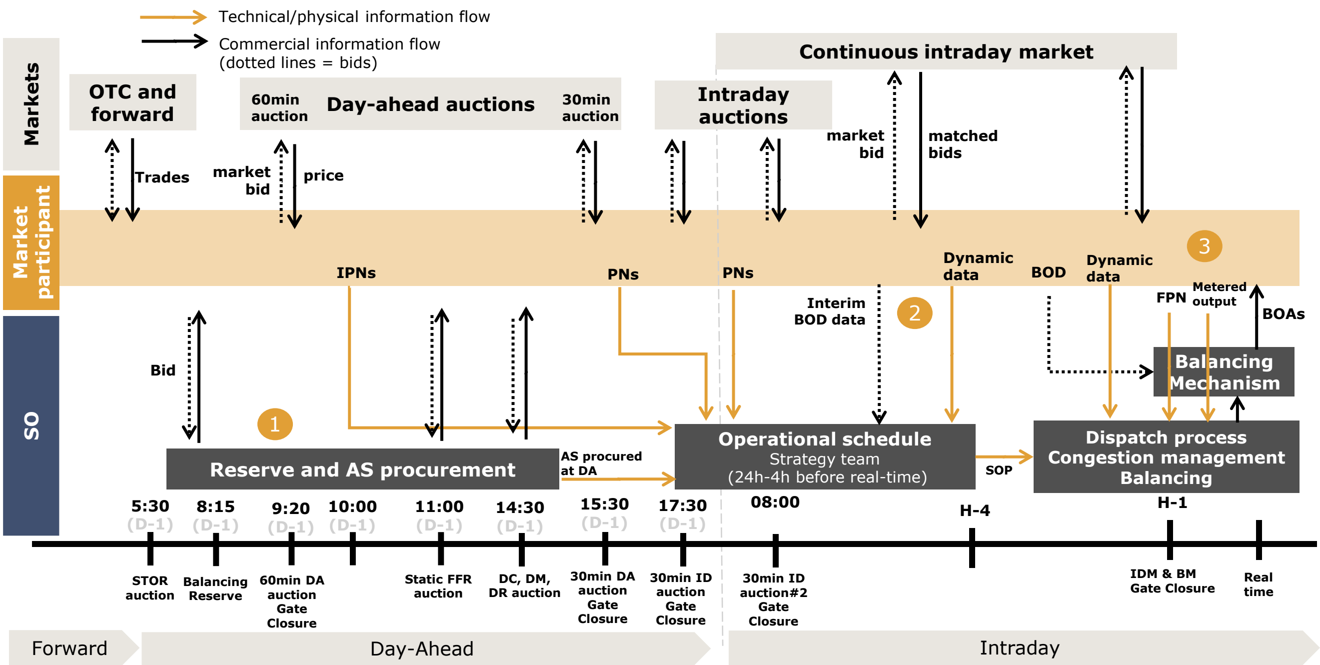
Separately from their trading, players participating in the BM are required by the Grid Code to submit Physical Notifications (PNs) to ESO that should reflect expected level of output or consumption for each unit. At gate closure (1 hour before delivery), the PNs become 'final' – Final Physical Notifications (FPNs) – and these are used for the purposes of taking any required balancing actions in the Balancing Mechanism. PNs can change up to gate closure, but accurate PNs at any given point in time are important from an ESO perspective. ESO needs to have good visibility beyond the next settlement period, to assess and prepare for future margin needs, constraint management and network optimisation.

Generators, large demand users and suppliers notify their PN, which is compared with ESO independent forecast of demand to determine whether the system is long or short. Electricity suppliers submit PNs, but ESO's view of demand in operational timeframes is considered to be more accurate than the aggregated supplier PNs.

In simple terms, the Balancing Mechanism is ESO's primary tool to maintain energy balance, procure sufficient regulating reserve, manage constraints and ensure system stability. Through the Balancing Mechanism, ESO may accept various sets of bids and offers, in effect making payments to (or receiving payments from) units in exchange for them agreeing to change their generation or consumption as compared to their FPNs.

Beyond this general balancing principle, there are several processes at ESO that happen prior to the hour of delivery in order to procure sufficient reserve and various ancillary services products ahead of time, as well as to anticipate the balancing actions that will need to be taken in balancing timeframes. Exhibit 5 provides a simplified overview of the interaction between ESO and market participants and the functions at ESO.

Exhibit 5 – Overview of market participant and system operator actions in Great Britain



Notes: 1. ESO procures various ancillary services at day-ahead stage. Not all of the reserve requirements are necessarily procured DA, ESO studies the counterfactual cost of procuring within day through the BM.
2. BM participants can update Physical Notifications and Bid Offer data up to gate closure; and dynamic data (technical parameters) up to real time.
3. Interim and final PN are not required to reflect traded position.
Source: AFRY

Auctions for the procurement of response and reserves take place at the day-ahead stage. The Balancing Services Optimisation team takes a view on the expected counterfactual cost of procuring reserve and response in real-time through the BM to decide the share of requirement to be procured at day-ahead.

For 24h to 4h ahead of real time, the control room strategy team studies the system as real-time approaches. The strategy team monitors margin levels and constraints, taking into account demand and renewable forecast, interim PN, network limits and balancing bids and offers (at the time). The objective of this process is to provide an indicative plan to the energy team working in balancing timeframes on the actions needed closer to real-time. The System Operating Plan (SOP) is handed over from the strategy to the energy team at 4h ahead of real time. The SOP is a snapshot of key information at key demand peaks and troughs made available to the control room at the time of its creation.

The SOP and the data that it was built upon (PNs, BM prices, etc.), however, are not firm and no unit is committed until a BOA is sent. Prior to the issuance of BOAs, ESO can commit units indirectly through trades and can send warming instructions to units with long 'notice to deviate from zero' times. However, the entire process is structured so that actions are taken as close to real time as possible.

Close to real-time, the energy team operates the Balancing Mechanism. The energy team considers:

- 'firm' information for generating units for the subsequent settlement period (and for which BMUs can no longer trade in the markets nor submit new PNs) – these include FPNs, BOD, dynamic parameters, availability etc.
- 'beyond the time wall' information for generating units, noting that PNs and BOD can change subsequently[12];
- other system-wide information, including:
  - interconnectors flows;
  - demand and renewable forecasts; and
  - network topology.

The scheduling and dispatch arrangements are designed so that the ESO takes actions as late as possible. It can issue instructions within the 'balancing window', which is the current SP and the two subsequent SPs, i.e. up to a horizon that is between 60 and 90 minutes away from real time. Actions can be, and are taken, based on a view of future needs (beyond the time wall), but this is based on indicative PNs and BOD, which can subsequently change. 'Scheduling and dispatch' is effectively a more continuous process with the control room trying to manage the 'now', whilst ensuring there is also capability to manage the same problems (energy

[12] Due to the dynamic parameters, some information 'beyond the time wall' becomes firm once a BOA is issued; in particular the committed energy that is bounded by the Stable Export Limit and the Minimum Non-Zero Time, become a firm parameter.

balance, network congestion and other systems needs) in time periods beyond the next settlement period.

Re-dispatch actions are taken for a variety of reasons:

- managing system frequency (i.e. matching supply and demand);
- ensuring there is sufficient frequency response and reserve to manage any potential sudden frequency deviations, to manage uncertainties in generation and demand forecasts, and to secure the largest infeed and outfeed loss risks;
- synchronising units and/or retaining units already ‘on’ to:
  - have a minimum level of inertia on the power system; and/or
  - have the required reactive power capabilities to then control voltage where and when needed; and
- resolving network congestion.

Since the introduction of NETA, there have been radical changes in the procurement of system services – new products have been introduced, different procurement methods have been used and the procured volumes have increased.

More changes and improvements are underway, including:

- ESO is consolidating its procurement reserve and response services at the Day Ahead stage. In particular, the introduction of Balancing Reserve should result in a reduction of synchronisations in the BM;
- longer-term contracts have been put in place (Pathfinders) to encourage entry from alternative providers, and this should limit the need for synchronising thermal units for inertia and voltage control[13];
- dispatching tools are being overhauled, which will make it easier to issue larger volumes of bid-offer acceptances from smaller assets; and
- the pilot Local Constraint Market (for the B6 constraint) that could allow ESO access to more resources.

3.3 The electricity system has undergone a rapid transformation in the last 20 years

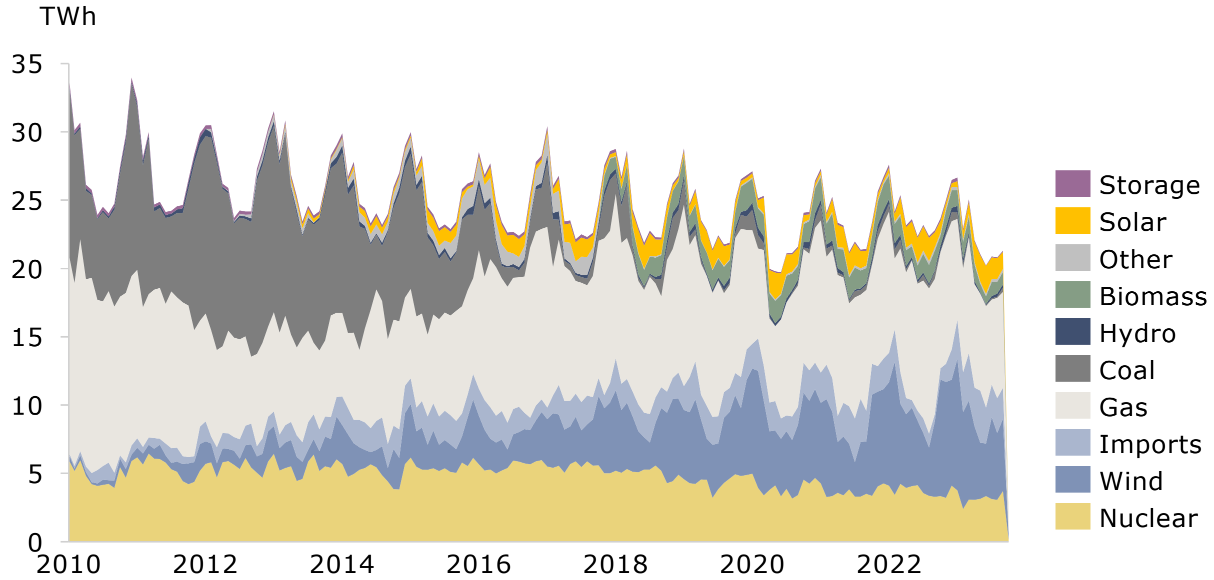
In the early 2000s, the electricity generation fleet in Great Britain consisted mainly of controllable thermal plants such as coal, gas and nuclear power plants. The European electricity sector witnessed a dramatic change in the last 20 years with the rapid uptake of renewable energy sources. Wind generation in Great Britain grew from 3% of the total electricity generation in 2010 to 25% in 2022. Solar PV then followed, reaching ~5% of total generation in 2022.

---

[13] There are however still discussions and considerations with respect to the balance between short-term and long-term procurement of some of these services, and the length of the contracts

Exhibit 6 – Historic electricity generation mix in Great Britain

Until the mid-2010s, the GB electricity system was dominated by dispatchable thermal generation. Non-dispatchable wind generation grew significantly in the last 20 years thanks to government support and decrease in costs.



Source: ESO, AFRY analysis

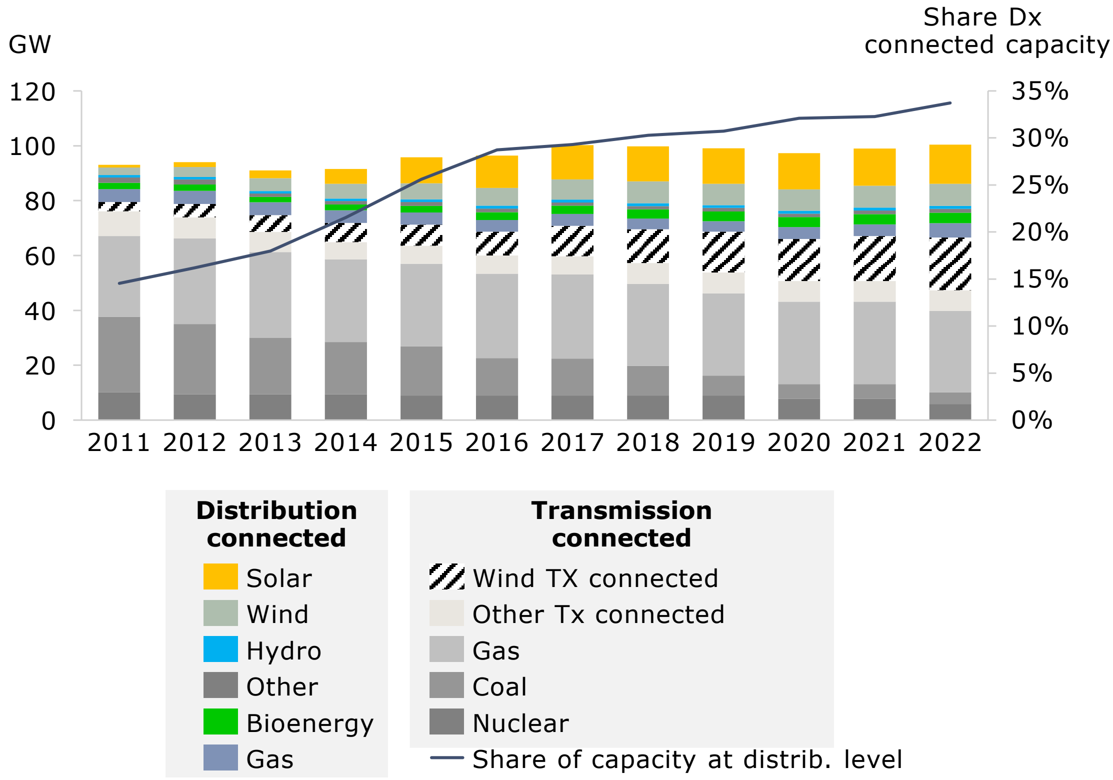
Wind and solar PV are intermittent and their available output is complex to predict. This rapid growth in intermittent RES capacity, which is set to continue in the future, has given rise to:

- greater price volatility, in particular in the intraday timeframe;
- increased 'swings' in residual volume needs from controllable generation;
- more needs for network congestion management; and
- greater challenges for system stability management.

The deployment of renewable energy sources has also led to a significant increase in capacity connected at the distribution level, for which ESO has less visibility. Incentives in the charging regime have also led to more non-renewable flexible capacity (such as engines, battery storage or biomass plant) to connect below transmission level. The exhibit below shows the increase in capacity connected at distribution level in the last 10 years.

Exhibit 7 – Historical installed generation capacity in GB, by connection level

The share of generation capacity connected at distribution level grew from 15% to over 30% in the last ten years



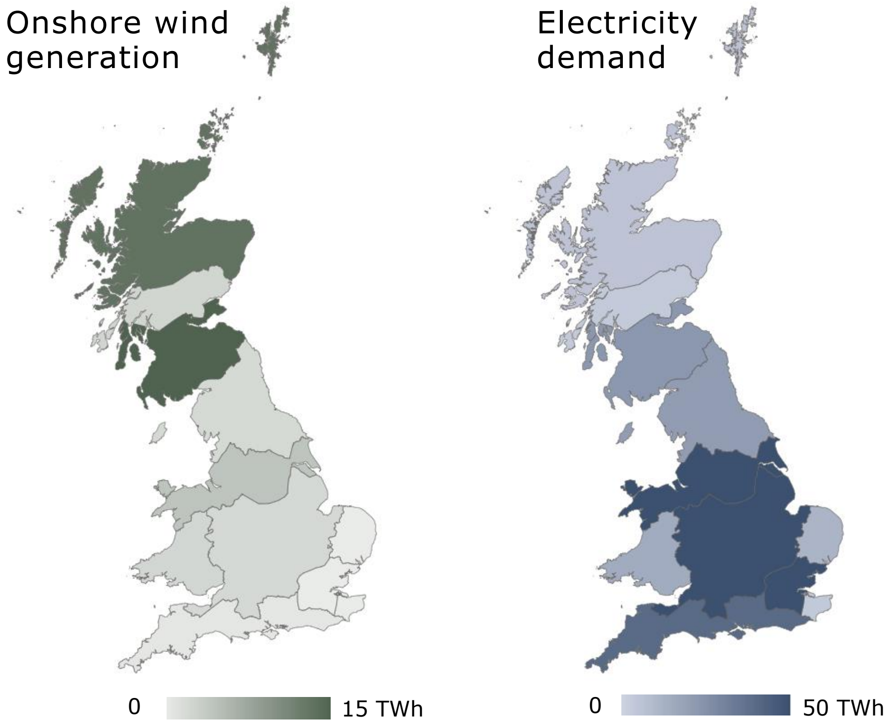
Source: Digest of UK Energy Statistics, AFRY analysis

3.3.1 The GB power system is becoming more congested

The wind resource is more favourable in the north of Great Britain and there are fewer barriers for wind development in Scotland, when compared to England. More wind capacity has therefore been developed in Scotland or the north of England. Given large electricity demand centres are located in the south of England, significant amount of power needs to be transmitted across the network during periods of high wind output levels with the network being constrained more frequently.

Exhibit 8 – Wind generation and demand by region

More wind capacity is connected in Scotland, while demand centres are concentrated in the south of GB



Source: AFRY analysis based on 2023 installed wind capacity and demand forecast for 2023

3.3.2 Inertia and reactive power are becoming scarcer

The increase in renewable energy sources (RES) on the system also has an impact on system stability management. As non-synchronous renewable generation increases, fewer synchronous generators are required to meet energy demand. However, this type of generators traditionally provided inertia or reactive power alongside active power. The ESO instructs transmission connected BMUs, while synchronised, to absorb or inject reactive power to manage the voltage on the transmission system. This activity, together with use of transmission network equipment is used to balance the reactive power injection/absorption from the distribution network and from the active power flows around the transmission network. Reactive power does not travel far, meaning that its management is necessarily localised.

At times with high levels of non-synchronous generation, the supply of inertia and reactive power from more traditional generating units is limited, and there is a need for alternative provision, or, should such solutions be absent, there is a need for committing out-of-merit thermal generation, whilst curtailing RES generation or trading interconnectors in order to maintain an energy balance.

3.3.3 More interconnection helps with more efficient resource allocation, but makes scheduling and dispatch more complex

The importance of interconnection has also grown over the last few years. Interconnectors help with more efficient use of resources and can mitigate the need for additional capacity in GB. Interconnectors provide a range of benefits, for both consumers and industry, including potential for renewable generation exports, greater security of supply and increased competition. However, their significant installed capacity combined with being price-responsive (to several price signals in the interconnected markets) can create operational challenges. They can be a source of rapid swings in net power flow at short notice.

Separately, since the UK departure from EU's internal energy market, trades on the interconnectors between the EU and Great Britain are no longer managed through single market tools, and the UK can no longer participate in cross-border balancing platforms. For example schedules for the North Sea Link to Norway cannot easily be altered after day-ahead (except for emergencies). NSL is connected to the north of a significant transmission boundary and its imports often exacerbate north-south constraints in GB.

3.3.4 The rise of battery storage and price responsive demand

A source of flexibility that has been deployed at scale in recent years is battery storage. Batteries can rapidly change their energy position and, as a result, be used to balance the system. However, in the current framework, ESO may not always be able to access such flexible energy resources, and these may react to price signals, at times, contrary to what is required for system operability.

There have also been developments on the demand side in recent years. Behind-the-meter solar PV and battery systems have grown notably, supplier-driven demand response programs through smart meters and Time of Use tariffs are on the rise. These evolutions from the demand are only starting and expected to become more and more prevalent over time. Moving from traditionally predictable demand patterns to variable and price responsive demand leads to greater challenge for the system operator to forecast demand accurately. The complexity of the demand forecast exercise to compounded by the fact that an increasingly granular (locational) view of demand is required for system operation.

4 Methodology and assessment principles

4.1 Methodology

Our approach for the assessment of the current scheduling and dispatch arrangements includes:

- review of the scheduling and dispatch arrangements in GB and the linkage between the different markets/mechanisms operated by ESO and how the market interfaces with ESO;
- engagement with different ESO teams to gain a better understanding of how scheduling and dispatch works, and to gain a perspective of the challenges experienced by the control room;
- detailed forensic analysis of key historical days including:
  - analysis of balancing actions taken on the day, in particular considering the information available to ESO at time to make decisions in the BM;
  - the review of orders on the continuous intraday market, to study the liquidity of the market and potential interactions with balancing actions; and
  - modelling of historical days under alternative market arrangements to test potential change in dispatch efficiency with different market incentives and/or approaches to scheduling and dispatch[14].

The electricity market model we have used performs an optimisation of the entire GB power system. This could be thought of as the outcome of a central dispatch optimisation. However, one could also argue that the optimisation results would not be dissimilar to those coming from a competitive self-dispatch market assuming the right incentives are in place. Dispatch efficiency improvements coming from changes in the scheduling process itself are not captured in this modelling exercise (e.g. central

[14] Given the electricity market model performs an optimisation calculation (i.e. akin to a central dispatch), dispatch efficiency improvements coming from changes in the scheduling process itself are not captured by this modelling exercise (e.g. central dispatch vs self-dispatch). The modelling assessment, presented in details in annex A, provides insights in terms of overall system costs and distributional impact assuming alternative market arrangements.

dispatch vs self-dispatch). The modelling assessment, presented in detail in Annex A, provides some insights in terms of overall system costs and distributional impact assuming alternative market arrangements. We have specifically looked at explicit procurement of operating reserve and earlier management of network constraints.

Forward-looking dispatch modelling can be a very powerful tool for assessing some policy decisions, understanding the economics of different generation assets, determining expected price formation and projecting relative generation patterns. But it cannot capture the nuances between the different scheduling and dispatch approaches. The modelling we have performed therefore has an ancillary role, and is not intended to be a full Cost Benefit Analysis of different dispatch arrangements. This would require extended forward-looking modelling of multiple future years, and, as already discussed may still not be appropriate for deciding between different options for scheduling and dispatch arrangements. We should be careful in the interpretation of the results and how these are used to inform the conclusions we draw in our overall assessment.

Although we have looked at some annual historical metrics, such as total balancing volumes and costs, we have chosen to focus on selected days for the following reasons:

— looking in detail at an entire year is a time-consuming exercise and may not have allowed us to focus on the critical issues; and
— it is the days that are outliers today (days with high levels of wind output and binding transmission constraints) that will become more common in the future.

4.2 Assessment framework description

We have developed an assessment framework to help us understand how the current scheduling and dispatch market design performs against some typical assessment criteria. The objectives are presented in Exhibit 9.

Exhibit 9 – Assessment framework

Objectives	Objectives
Efficient operation	- Are the scheduling decision made at the right time and by the right actor? - Are resources allocated appropriately across markets, products and timeframes? - Are units positioned effectively for subsequent dispatch? - Is trading and scheduling on the interconnectors efficient and seamless?
Ensuring operational security	- Are the right information and tools available to the SO to ensure secure operation of the system? - Does the design provide appropriate operational incentives? - Is pricing/remuneration for energy and AS consistent with the needs of the system? - Is there price continuity between the decentralised and centralized decision-making process?
Support investment	- Is pricing/remuneration predictable? - Does pricing support hedging instruments to be developed around it? Does it reflect underlying fundamentals? - Does it provide efficient entry (and exit) signals?
Appropriate risk and cost allocation	- Do the relevant risks fall on the actors best suited to deal with those? Do they have the right tools to deal with such risks? - Are risks and costs associated with production, transportation and consumption of electricity allocated in a fair and reasonable manner?
Supports competition and creates level playing field	- Does the design promote competition between participants? - Does the design enable all resources to contribute appropriately? - Are gaming opportunities kept to a minimum?
Adaptive	- Can the arrangements be developed and modified in a straightforward and cost-effective manner? - Is the design robust to change?
Transparency and replicability	- Can participants explain and predict choices made by the SO? Do participants know why they are being scheduled/dispatched and does this knowledge allow them to make effective decision going forward? - Does the SO have the right type and detail of information needed to fulfill its role? Does it have access to this information at the right time?

This assessment framework has been guiding our thinking when investigating the potential limitations with the current scheduling and dispatch design.

5 Limitations of the current scheduling and dispatch arrangements

This section provides context around how the ESO role has evolved since the introduction of the current market arrangements, and gives a description of the limitations of the current scheduling and dispatch arrangements:

- chapter 5.1 presents the shift in ESO's role due the change in the electricity system; and
- chapter 5.2 details the identified limitations of the scheduling and dispatch across three key themes.

5.1 ESO is increasingly acting as a central scheduler in a market environment designed for a 'residual balancer' SO

As already discussed in Section 3, the GB market was designed around a principle of self-scheduling and self-balancing. Buyers and sellers were expected to trade to balance their positions, leaving the TSO to deal with any residual issues relating to within-settlement period balancing and other network and system constraints (see section 3.1).

At the advent of New Electricity Trading Arrangements (NETA) a connection was made between the timing of gate closure and the extent to which NGC, the SO at the time, would need to deal with transmission constraints in the Balancing Mechanism. The intent was that a set of tradeable access rights would be introduced to manage congestion, although this was not taken forward[15].

These market arrangements were envisioned and implemented at a time when the market was dominated by controllable thermal generation, with generally less uncertainty to account for. The output of different generating units was almost independent from weather conditions, and there was

---

[15] Proposals were made for "a market-based approach to the allocation of transmission access rights" to be introduced post-NETA go-live (reference "The New Electricity Trading Arrangements and Related Transmission Issues – proposals on Licence Changes – a Consultation Document", Offer, December 1999)

significantly less congestion on the network. In the early days of NETA the total volume of redispatch actions by the TSO was low (in 2006, annual positive redispatch volumes were <1% of the annual national demand, compared to over ~4.5% in 2022).

Since then, the electricity system has undergone a major transformation. NETA was introduced for England and Wales (excluding Scotland), with access for Scottish participants via (limited) access rights. Congestion management today is far more prevalent than envisaged in the original design of the trading arrangements, compounded by the allocation of firm access to Scottish generation at the advent of BETTA (British Electricity Trading and Transmission Arrangements), the subsequent connect-and-manage[16] decision and the difficulties of building onshore wind in England. The average generation mix has moved from 95% of controllable generation[17] in 2009 to ~65% in 2022. Furthermore, a large share of generation is now small scale distribution connected over which the TSO has very little visibility or control : in 2022, ~10% of electricity generation came from embedded wind and solar PV.

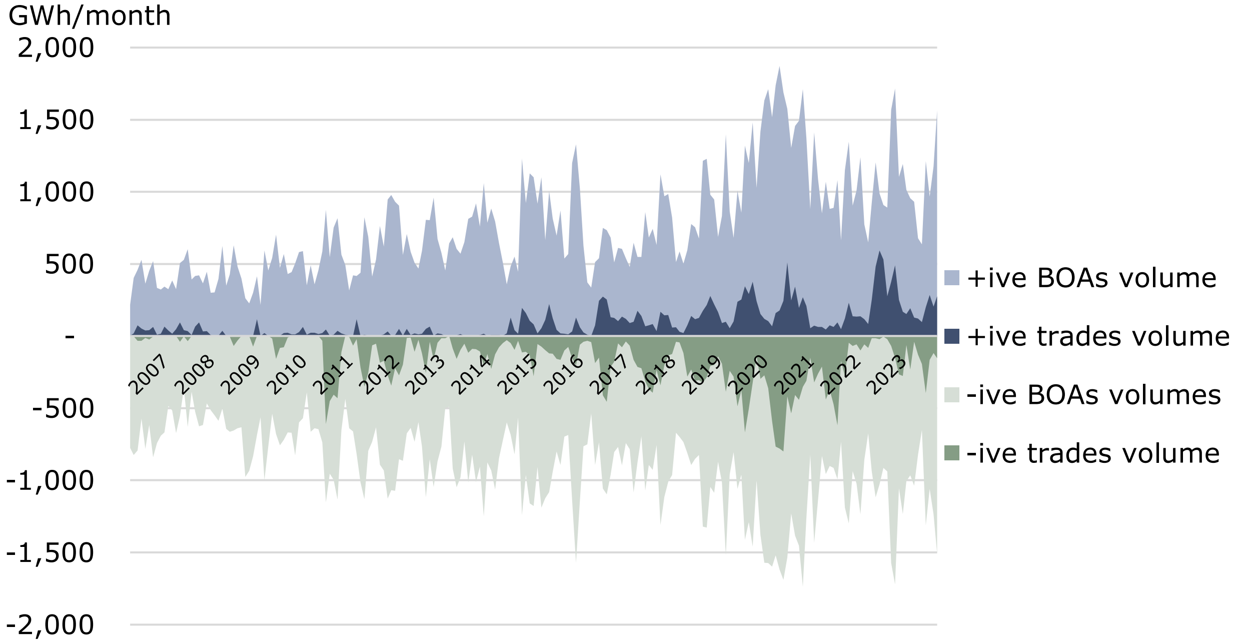
Exhibit 10 shows the volume of Bid Offer Acceptance (BOAs) and trades, both for positive and negative balancing actions in recent years. Volumes of activated balancing energy are clearly on the rise – this is to some extent expected given the greater unpredictability and imbalances created by more intermittent renewables on the system. In 2022, the average volume of positive balancing energy activated represented ~4.5% of GB national electricity demand. Although these are all classified as 'balancing energy' volumes, most of these are redispatch for reasons other than matching supply and demand.

---

[16] Ofgem introduced the enduring 'Connect and Manage' regime in 2010, which allowed National Grid to give earlier grid access to new and exiting generation projects without the need for transmission system reinforcement prior to connection of further capacity
[17] Dispatchable generation means here power generation that can adjust its output on demand, as opposed to intermittent and non-dispatchable renewable energy sources such as wind power or solar power.

Exhibit 10 – Monthly balancing volumes (BOAs and trades), 2006-2023

The volume of balancing actions, both Bid-Offer Acceptance and trades, consistently increased in the last 20 years. 2020 provides a glimpse into the future – electricity demand was very low because of the impact of COVID-19, resulting in an increased relative share of intermittent renewable generation



Source: ESO, AFRY analysis

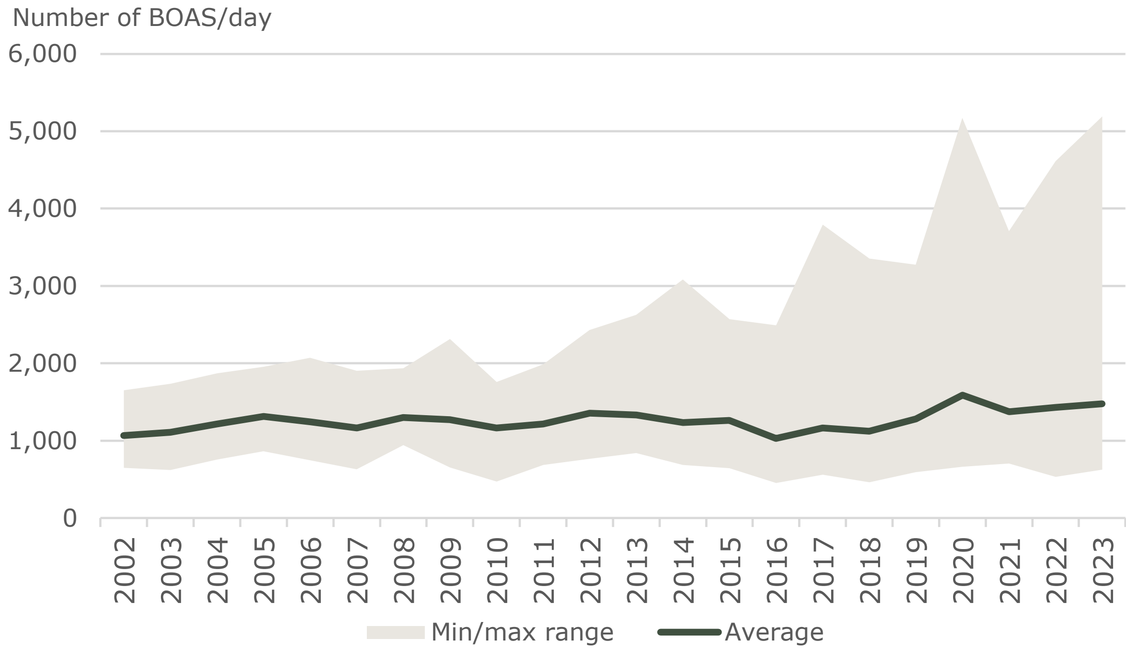
The annual average may not be that high, and on most days ESO's role is much more residual. However, on individual days the volume of redispatch can be rather large:

- ESO redispatched 20% and 30% of the market (through the BM and trades) on 12 April 2023 and 2 July 2023 respectively – these were days with significant levels of network congestion[18]; and
- on 1 January 2023 the need for inertia and voltage control meant a need of around 14% volume redispatch.

The overall volume is one side of the problem. The number of individual BOAs is also important and reflects the complexity of the dispatch process. The minimum daily BOAs have remained broadly unchanged. The average number of daily BOAs has been increasing, in particular over the last five years. The sharpest increase is in the maximum number of BOAs. This suggests there are days with a very strong need for ESO intervention.

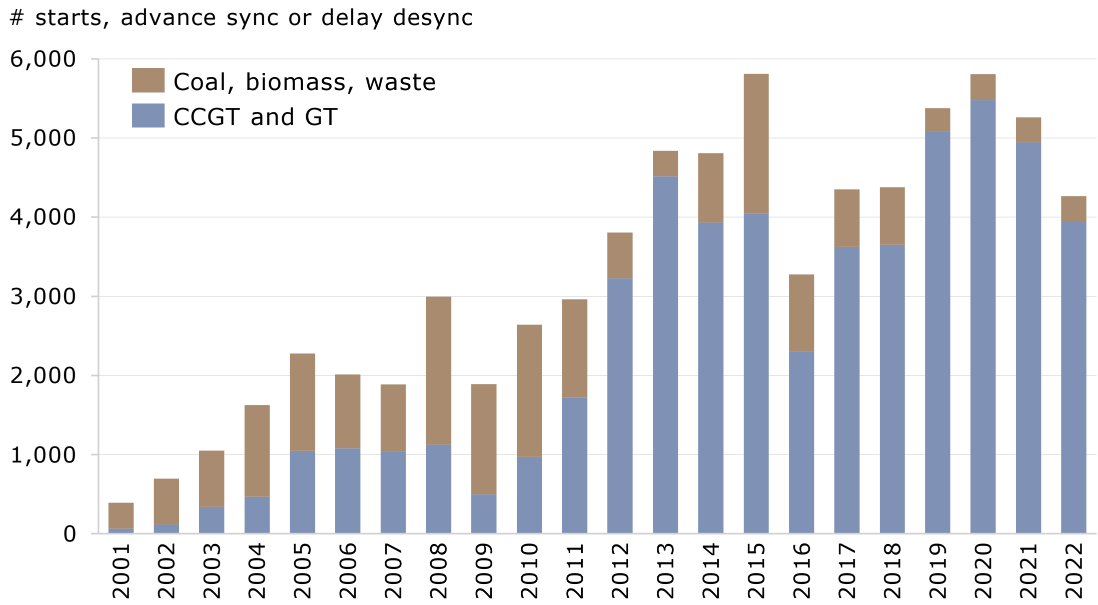
[18] The 12 April 2023 represented 0.35% of the redispatch volumes of the year; however, it accounted for 0.7% of the balancing costs of the year. It illustrates that the relationship between balancing costs and volumes is not linear, and that days with high redispatch volumes are also the most significant contributors of balancing costs.

Exhibit 11 – Minimum, maximum and average number of daily BOAs



Beyond the volume of balancing, and the number of individual actions, further evidence that ESO's role has changed over time is the number of unit commitment decisions made through the BM. Aside from adjusting the dispatch of already synchronised units, ESO needs to either synchronise additional units, advance a synchronisation or delay a scheduled desynchronisation. Exhibit 12 shows the trend in unit commitment actions via the BM over the last 20 years. ESO synchronises more units today when compared to the early days of NETA. However, we do need to note that unit synchronisations have been dropping over the last two years (mainly driven by changes in system requirements, e.g reduction of the minimum system inertia requirement, and the introduction of the new Dynamic Services products).

Exhibit 12 – Number of unit commitment decision through the BM, 2001-2022



Note: 2001 from April only
Source: Elexon, AFRY analysis

ESO increasingly takes actions on some days to balance the system that could be compared to a central dispatch regime, but without the appropriate processes and tools. This does not happen on all days of the year, but when looking at specific days it is clear that as intermittent RES increases, the frequency of days when ESO will take significant redispatch actions will also rise in the 'status quo'.

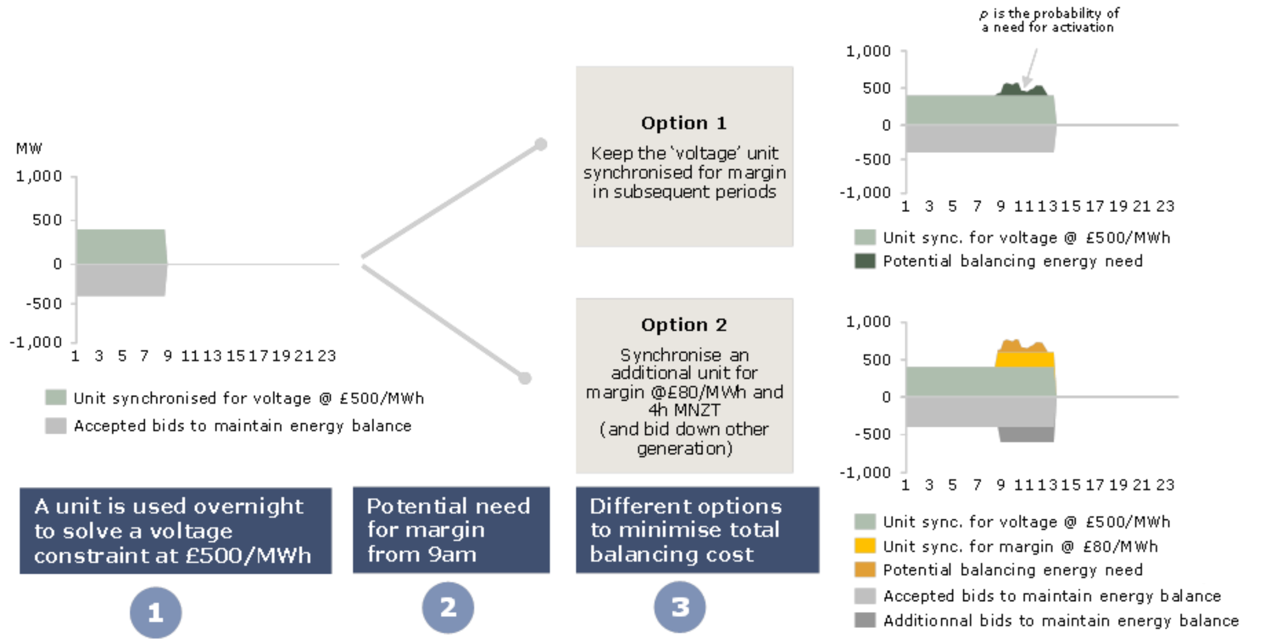
In real time, constraints, reserve and system operability needs have to be considered together and cannot be dealt with independently. ESO is effectively trying to perform a total cost optimisation. To achieve this, multiple decisions are made well ahead of real-time in an environment of continually evolving forecasts and information as the solution space becomes more limited towards real-time. Assessing all the various options is a challenging task.

In Exhibit 13 we provide a simple example of a decision ESO may face. ESO foresees the need to secure additional regulating reserve during the morning periods, and can choose between:

- keeping an already synchronised unit online with a potentially high activation price in the case where the 'headroom' ends up having to be activated; or
- synchronising an additional unit (and incurring the associated cost of synchronisation), but having the option to have access to a lower activation price.

The decision will obviously depend on the expected probability of needing to activate balancing energy and the relative costs of the different bids and offers.

Exhibit 13 – Example of ESO assessing different options for total cost optimisation



This is obviously, a very simple example. In reality, there will be a wide range of choices that need to be assessed continuously and under uncertainty.

5.2 Limitations of the current scheduling and dispatch market design and processes

ESO takes actions through trades and the BM to:

- ensure supply and demand are balanced (on a second by second basis) and to have enough regulating reserve (as well as frequency response);
- manage thermal limits on the network congestion (network congestion); and
- ensure that other system requirements are met (sufficient inertia, reactive power etc.).

Some aspects of the current market design and the scheduling and dispatch arrangements are creating issues for ESO and market participants, with an impact on level of balancing actions needed and potential inefficiencies in dispatch and procurement. This may in turn translate in higher cost to consumers and higher overall system costs.

Market participants are also affected. The current scheduling and dispatch processes make it difficult for them to understand and expect some ESO decisions and there is potential for underutilisation of some resources.

Together with ESO, we have identified some of these limitations in the current scheduling and dispatch market design given the current resource mix and network topology. These can be grouped into three themes:

- **incentives**
  - the energy markets do not provide scheduling incentives in line with system needs and operational requirements;
- **incomplete visibility and access**
  - incomplete ESO visibility of market outcomes and limited access to some resources impacts coherence between wholesale market and balancing;
- **intertemporal issues**
  - the current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time.

Exhibit 14 – Key limitations with current scheduling and dispatch market design

Limitations of the current market design and processes	Specific manifestation of limitation
Incentives	**'Unconstrained' market incentives:** incentive provided by national imbalance price does not align with network constraints and other system needs. Market participants may not always be rewarded for the value of the flexibility they provide.
	'National' imbalance price: portfolio level balancing and national imbalance price lead to dispatch/NIV chasing in 'wrong' location
	Potential missing signals for real time reserve procurement: market is not incentivised to provide reserve capacity where and when needed
Visibility and access	**Incomplete coverage:** coverage of FPNs is incomplete, particularly for the growing share of flexible non-BM resources, meaning ESO has limited visibility of full market schedules when doing contingency planning
	Inaccurate information: schedules change significantly before gate closure meaning ESO decisions are taken with inaccurate information
	Behaviour: uncertainty on the expected level of system support balancing by flexible non-BM resources (e.g. NIV chasing or response to retail tariffs)
	ESO access to resources: key resources respond to wholesale market signals but are not dispatchable by ESO in balancing timeframes
	Coordination: sequential procurement of balancing services adds uncertainty to decision making for both ESO and market participants
Intertemporal issues	**Timing:** ESO is obliged to take proactive decisions with consequences for future periods beyond Gate Closure, which overlaps with the operation of the intraday market
	Information: ESO takes decisions with inter-temporal consequences based on incomplete data
	Transparency: beyond-the-wall actions and advance commitments cloud transparency and may distort imbalance pricing, making it harder to forecast and predict by market participants.

In the following sections we provide additional details on each of the individual limitations we have identified.

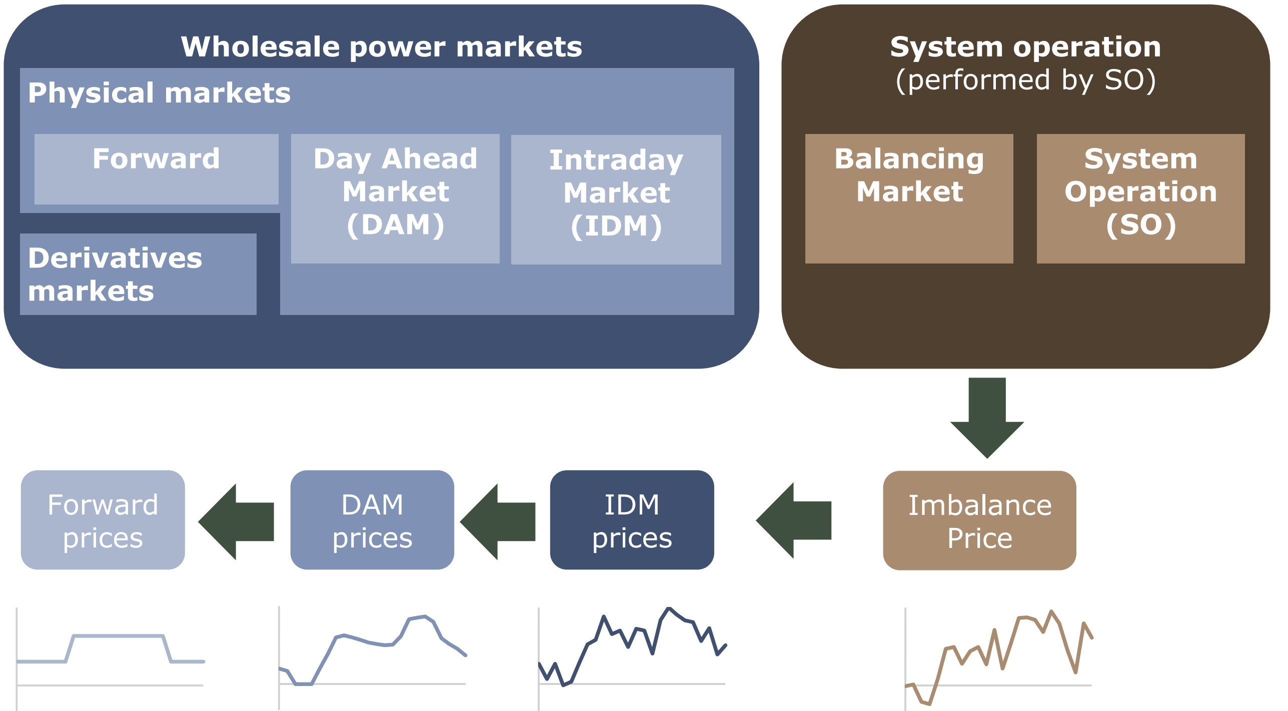
5.2.1 Incentives

Energy markets do not provide scheduling incentives in line with system needs and operational requirements

The wholesale market has been set up to trade a standard, GB-wide energy product. The Imbalance Price is the price that all participants ultimately face for any electricity volumes injected to the grid, and all ex-ante trading is for hedging against this price.

Market actors have balance responsibility and manage this through market trading and portfolio balancing. Collectively, the market is incentivised to support national supply and demand balance through exposure to the Imbalance Price. There are no obligations for individual participants to balance their own positions, and participants may continue to use non-BMU resources after GC for portfolio balancing or NIV chasing. Exhibit 15 shows schematically how the Imbalance Price backpropagates to all other ex-ante markets.

Exhibit 15 – Imbalance Price informing price formation in ex-ante markets

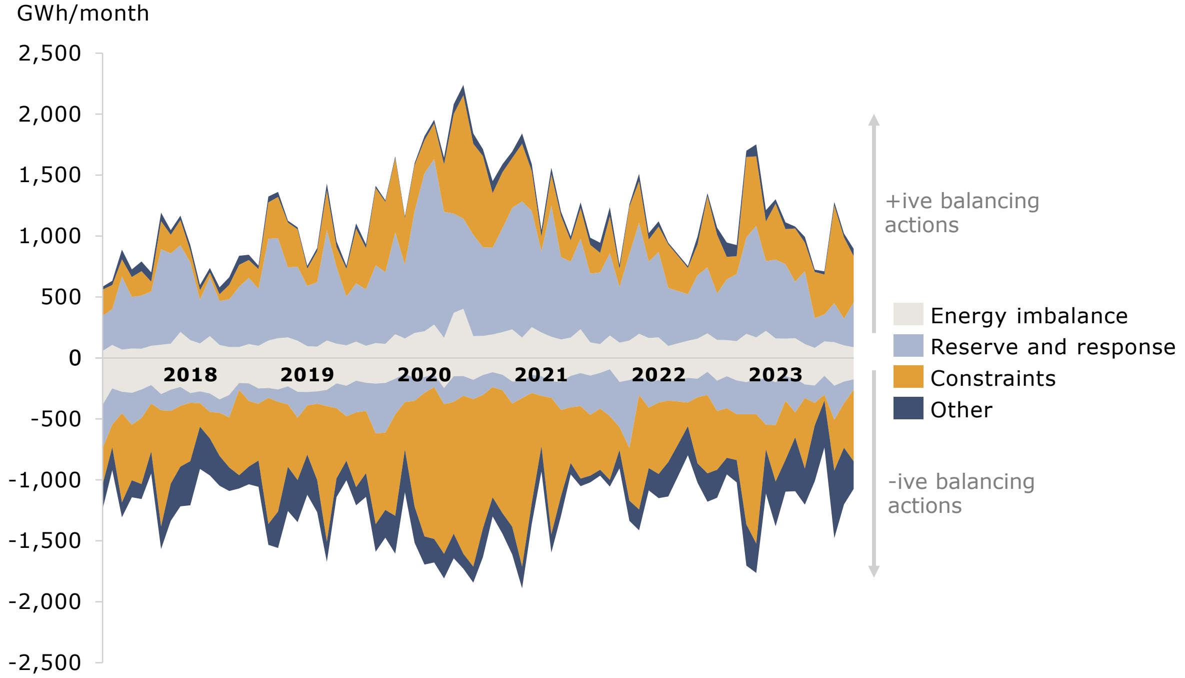


The absence of locational incentives before gate closure has been discussed thoroughly in the previous phase of REMA. The intention of this section is not to highlight the cost savings/ transfers that could be achieved via improved locational signals, but to set out how the single national price incentivises particular scheduling behaviours.

ESO starts from the 'unconstrained' PNs submitted by market participants, and redispatches units to manage system constraints and ensure sufficient operating reserves. While this type of redispatch is clearly part of the SO role, the complexity for ESO comes from the large volume of actions and the difficulty to effectively optimise costs and unit constraints over time.

Our analysis of historical information suggests that the volume of balancing actions for system constraints and reserve is now significantly greater than the volume of pure balancing energy actions. Exhibit 16 shows the monthly volume of balancing actions in recent years.

Exhibit 16 – Monthly balancing volumes by type of action



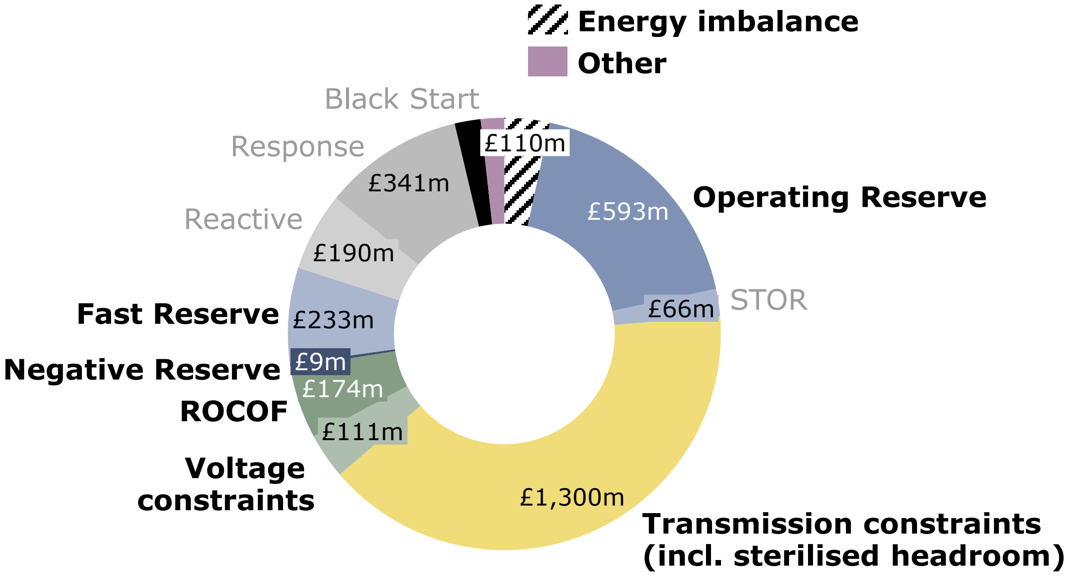
Note: 'Constraints' in this chart include transmission constraints and other system needs (e.g. inertia and voltage). These actions are system tagged. 'Reserve and response' on this chart represent the volume of balancing actions to position units to provide response and reserve, and not the activation of these reserves.
Source: Daily BSUoS volume Data, AFRY analysis

It is not only the volume of BM actions for reserve and constraints that is high, but also the associated costs, as shown in Exhibit 17. Over time, the procurement of system services has evolved with:

- the move from long term tenders to day-ahead co-optimised auctions for response, via the Enduring Auction Capability platform;
- the introduction of commercial approaches to test competition with TOs solutions for non-energy needs via Pathfinders, and;
- the introduction of new services such as Balancing Reserve.

However, the Balancing Mechanism remains ESO's primary tool to maintain energy balance, procure sufficient operating reserve, manage transmission constraints, and ensure system stability. Transmission constraints and operating reserve account for almost 2/3 of the overall balancing cost.

Exhibit 17 – Total balancing and ancillary services cost for FY 2021/22



5.2.1.1 'Unconstrained' market incentives

The incentives provided by the national imbalance price do not align with network constraints and other system needs

Market participants do not have any incentive to provide energy or reserve in the right location (relative to thermal constraints or voltage needs). Providers behind an export limit may therefore end up selling electricity in the wholesale market even though this may subsequently have to be curtailed. On the other hand, providers behind an import limit constraint may not trade their output even though they may eventually be needed to generate in real-time. This then means ESO reduces output from units behind export limit constraints and increases output from units behind import limit constraints though BM actions. The resulting dispatch may be similar to what it would have been assuming network constraints formed part of the market. This may not always be the case though – the later such information feeds into the dispatch process, the more likely it is that some options may not longer be available. As we approach real time, some options are inaccessible and only more flexible solutions can respond.

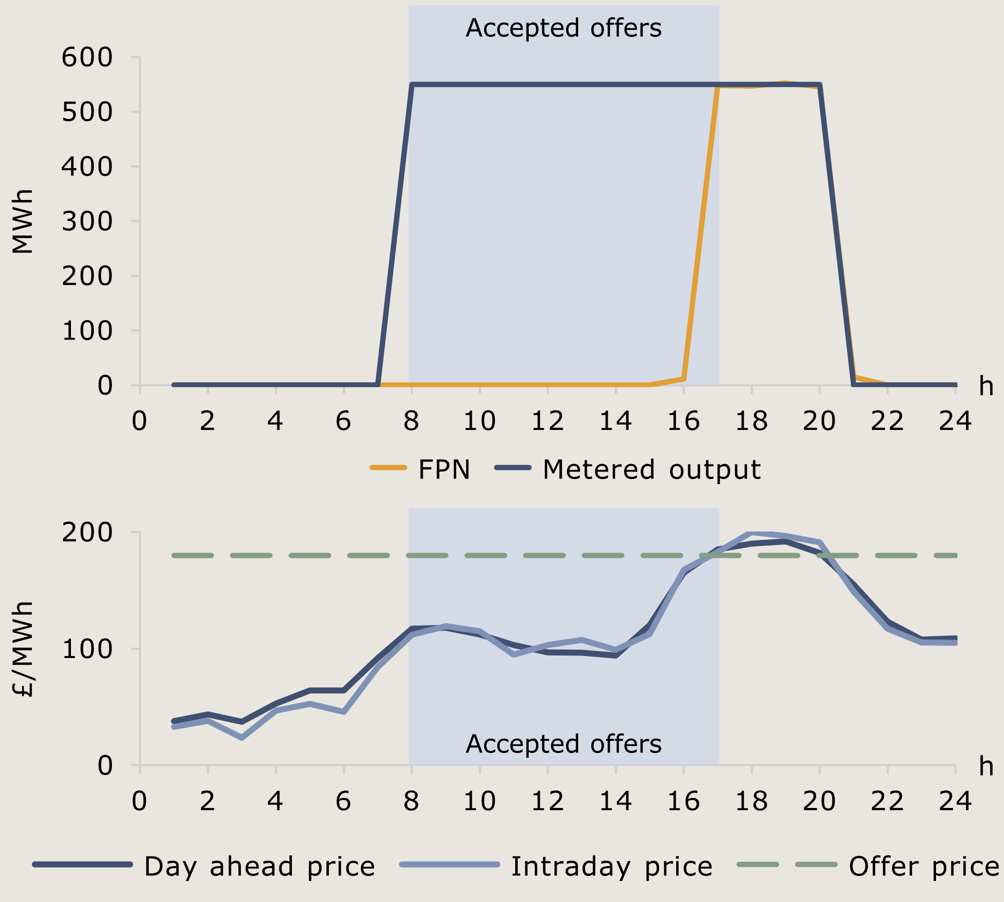
This can also be more easily explained through some examples.

Generator in an import-constrained location

Exhibit 18 shows the FPN, BOAs and offer prices for a thermal unit in an import-constrained location. This unit trades volumes in the ex-ante 'unconstrained' markets and submits a positive FPN over the evening peak periods. Market prices are, however, below its short-run cost of operation in the morning and in the afternoon, and the unit is not scheduled to generate.

ESO issues BOAs to synchronise the unit earlier to relieve the import constraint. The national Imbalance Price does not provide a signal for the unit to synchronise in the morning.

Exhibit 18 – Illustrative FPN and BOAs for thermal generator in import-constrained location

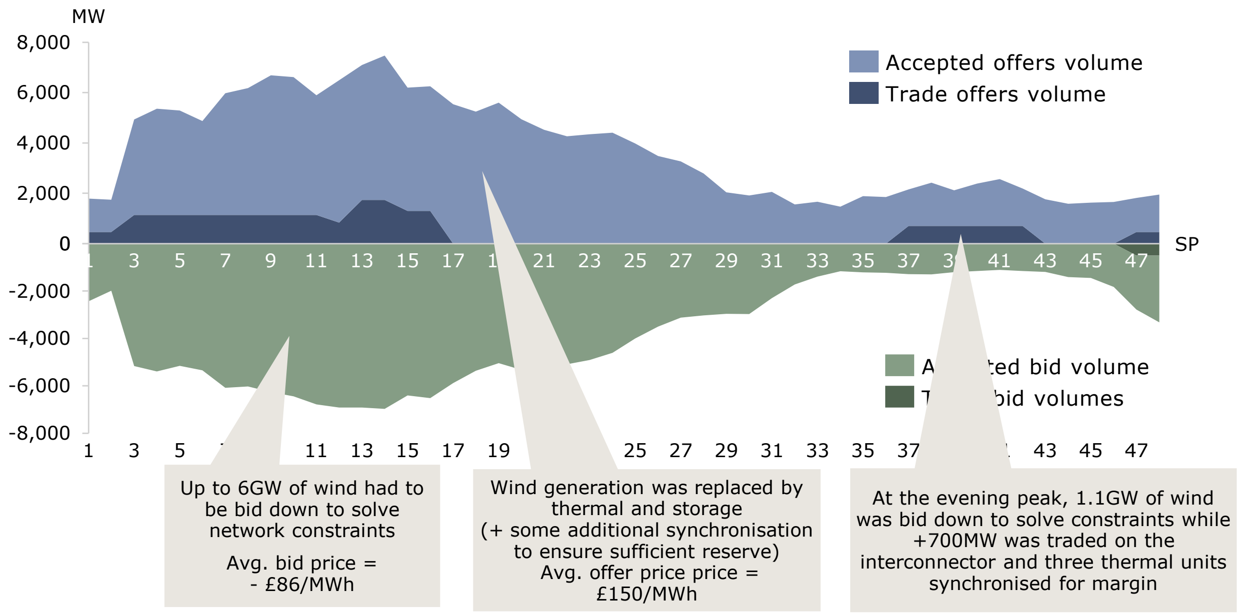


One of the most typical examples for the need to re-dispatch to manage network constraints is at the boundaries in the north.

On 12 April 2023, the north-south export constraints were active (from around 8am, B4, B6 and B7 constraints were active, driven by very high levels of wind output in Scotland). Exhibit 19 shows the accepted bids and offers on that day. Wind and pumped storage behind constraints had bids accepted and their scheduled output was replaced by thermal generation in the south. This happened for most settlement periods of the day.

Exhibit 19 – Scale of re-dispatch on 12 April 2023

In the morning hours, more that 6GW of redispatch was required, for a national demand of ~20GW

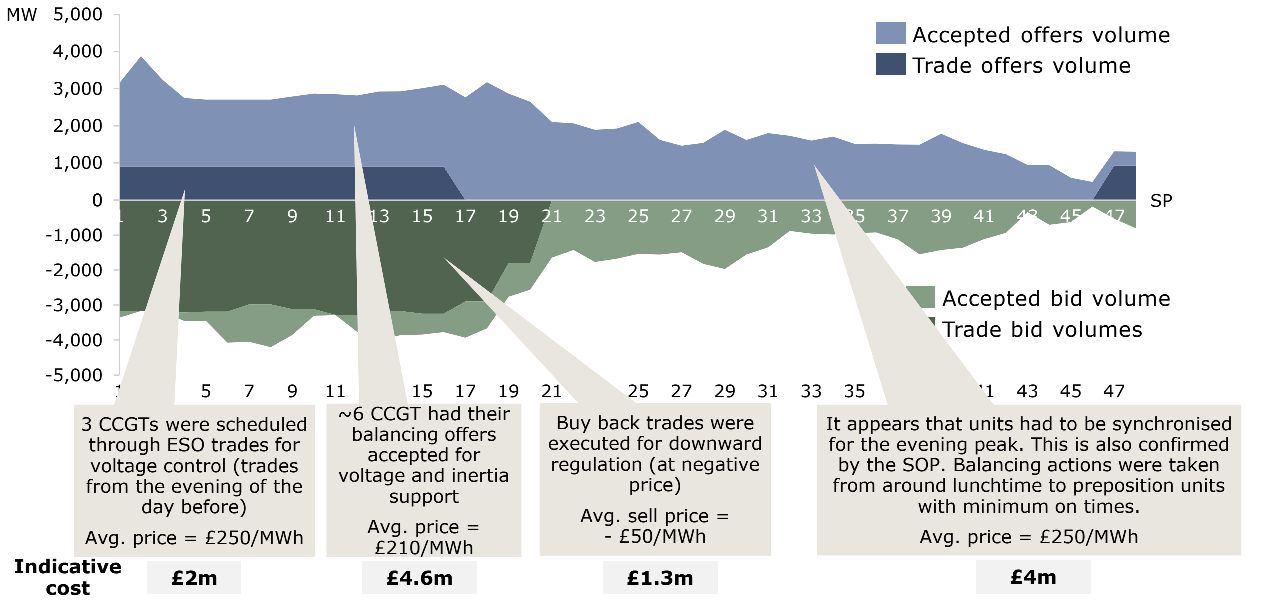


Indicative cost	**£5m until noon**	**£8m until noon**	**~£1m between 6 to 9pm**

The example of 12 April 2023 highlights the scale of re-dispatch when network constraints are binding. In that case there was disproportionately more wind output in Scotland and all redispatch was for managing the thermal limit constraints. Another example of a ‘poor’ starting point from the market for subsequent dispatch is when there is too much non-synchronous generation on a nationwide basis with the market ignoring the need for a minimum level of inertia, and capability for reactive power absorption/injection in specific locations.

On 1 January 2023, demand overnight was fairly low and there is relatively high wind output. The market delivers very little synchronous generation and ESO synchronised units to manage inertia levels and for voltage control. This is shown in Exhibit 20.

Exhibit 20 – Scale of re-dispatch on 1 January 2023



We have also modelled two days (12 April and 2 July 2023) in 2023 when transmission constraints were binding to determine the potential impact on dispatch efficiency. In the counterfactual modelling we have performed for these days, we have identified that the absence of locational signals in the ex-ante markets can result in less efficient results (higher system costs), but also in an increase in cost to consumers.

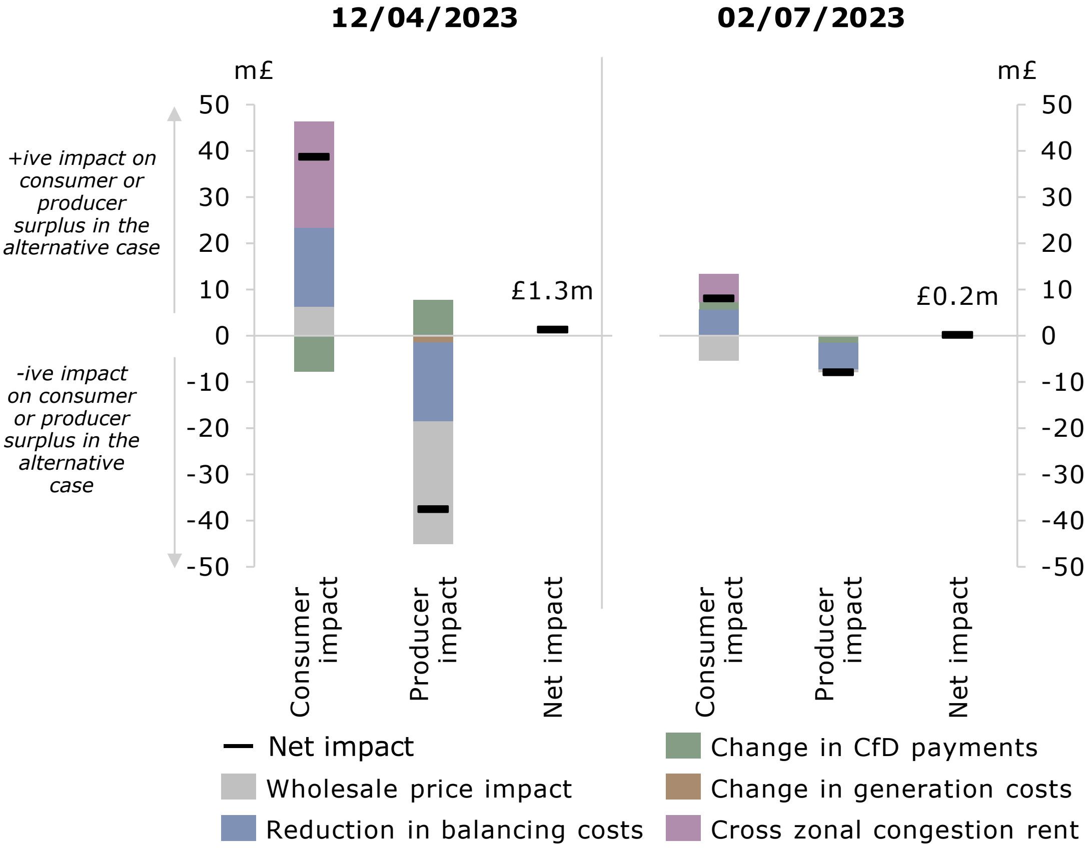
This is shown in Exhibit 21. On 12 April 2023 in particular, the estimated increase in consumer surplus is around £38m in the presence of market incentives to reflect some key network limitations. This is because of:

— a sharp drop in balancing costs; and

— a reduction in the demand-weighted average wholesale price.

The modelled increase in consumer surplus is partially offset by an increased need for CfD payments to supported RES in Scotland. However the below numbers do not account for any potential change in the TNUoS charges, which would take place assuming some form of market zoning. This would then also mean the benefit to consumers would be significantly smaller than what is presented here. This would, however, not affect the modelled reduction in system costs, which comes as a result of improved scheduling of some of the storage and thermal units.

Exhibit 21- Change in system costs and welfare distribution with earlier management of boundary constraints



Note: AFRY modelled historical days to explore the change in balancing costs and overall system costs (representing dispatch efficiency) under alternative market arrangements. This chart presents the difference in consumer and producer surplus between the '*Status-quo'* (counterfactual model run reproducing the scheduling and dispatch decisions taken on the key study day) and
*Early management of boundary constraints (*representing either a) a zonal market; or b) the presence of a scheduling process where ESO redispatches the market at an early stage to solve for transmission constraints with access to cost-reflective three-part bids)
"Cross zonal congestion rent" is, in a zonal market, the congestion rent between the zones captured by the system operator and transferred to consumers (e.g. through a reduction is network charges)

Conversely, on that same day there is a reduction in producer surplus. Producers in Scotland face a much lower wholesale price and some generating units miss out on additional income from BM actions. This is partially offset by greater CfD payments to some RES supported generation. As already mentioned, this reduction would be much smaller if we accounted for a change in the TNUoS charging in the presence of more locational wholesale pricing.

5.2.1.2 'National' imbalance price

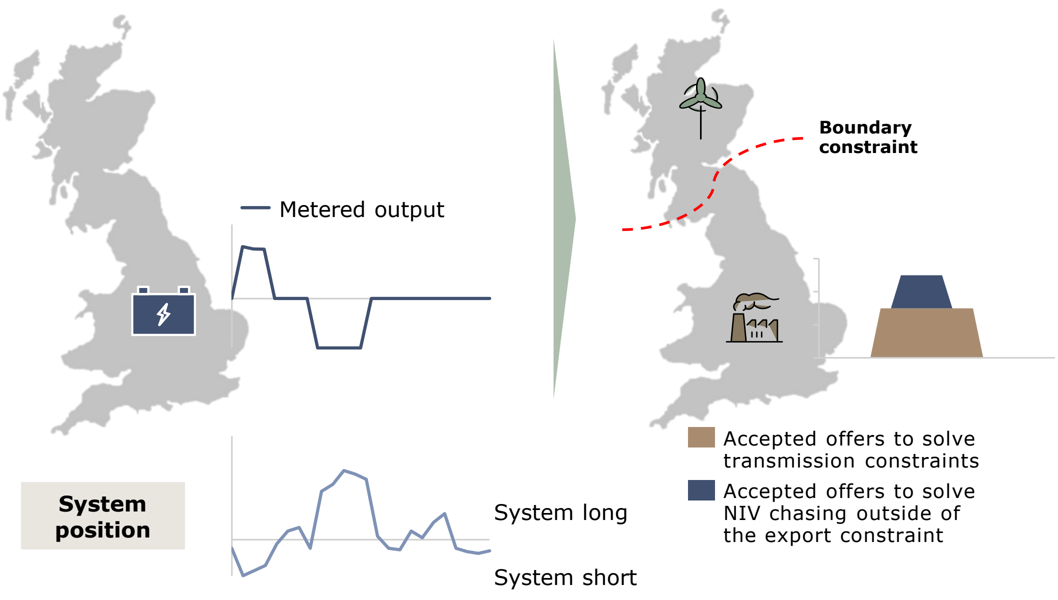
Portfolio level balancing and national Imbalance Price lead to dispatch/NIV chasing in 'wrong' location

NIV chasing

The Net Imbalance Volume (NIV) is the difference between outturn and contracted volumes in each settlement period (i.e. how much more or less electrical output is needed to balance supply and demand). 'Net Imbalance Volume (NIV) chasing' is a practice where market participants try to anticipate the overall net system position and adjust their own positions when they expect the imbalance price to be favourable when compared to their short run cost of operation. In other words, market participants may choose to be imbalanced in the opposite direction of the system to be paid more or to pay less than they would have otherwise paid or been paid if they traded their position in the ex-ante markets.

NIV chasing can, in theory, be helpful to the system, resulting in overall lower NIV. However, in practice, it only supports system energy balance at the settlement period level (30min), does not ensure frequency control (done continuously), and may also create additional challenges for ESO. In particular, the single national imbalance price to which market participants respond ignores locational factors. In case of transmission constraints, NIV chasing may be in the 'wrong' location, exacerbating constraints instead of supporting system operation. ESO reacts to both resolve congestion, and ensure energy balance (effectively undoing the NIV chasing position), resulting in an increase in balancing actions. This dynamic between two uncoordinated processes, ESO actions and NIV chasing, is complex to quantify, as one would need to know BSC parties' contract positions, and the operation of their underlying assets.

Exhibit 22 – Illustration of the impact of NIV chasing in case of transmission constraints

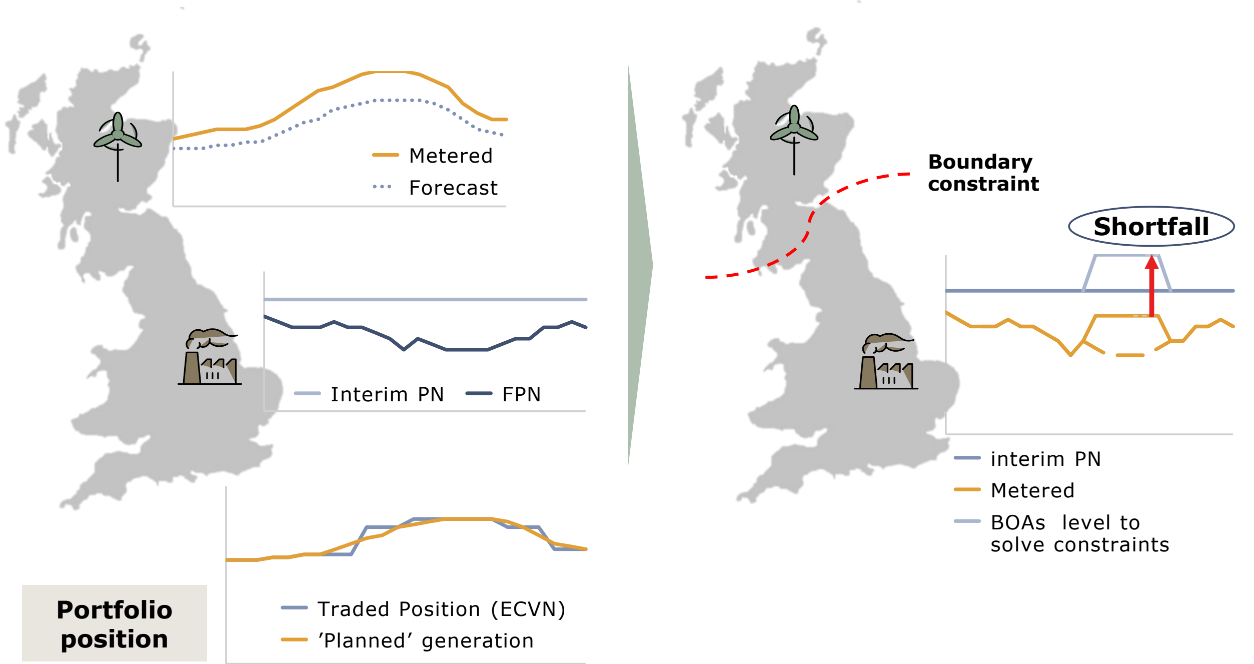


Portfolio balancing

Imbalances are calculated at market participant portfolio level (rather than at unit level), and are settled based on the difference between the portfolio contracted position and the aggregated metered output of units in the portfolio. With portfolio level balancing and a single price zone, the resulting position may end up giving rise to network congestion.

For instance, in case of a wind generation outturn exceeding the forecast, a portfolio manager who owns wind assets and thermal units located on opposite side of a transmission constraint may decide to lower the output of thermal units to balance its portfolio. It may further lower the output to have an overall short position, expecting that the system will be long (NIV chasing). In this high wind generation situation where ESO may need to bid down wind and to accept offers from thermal generation to alleviate network constraints, the lower PN from the thermal unit increases the balancing action needed from ESO. This type of portfolio balancing has been reported during engagement with the control room team; it is however difficult to evidence and quantify without knowing the decisions and operations of individual market players.

Exhibit 23 – Illustration of the impact of portfolio level balance responsibility in case of transmission constraints



Source: BMRS, AFRY analysis

5.2.1.3 Potential missing signals for real-time reserve procurement

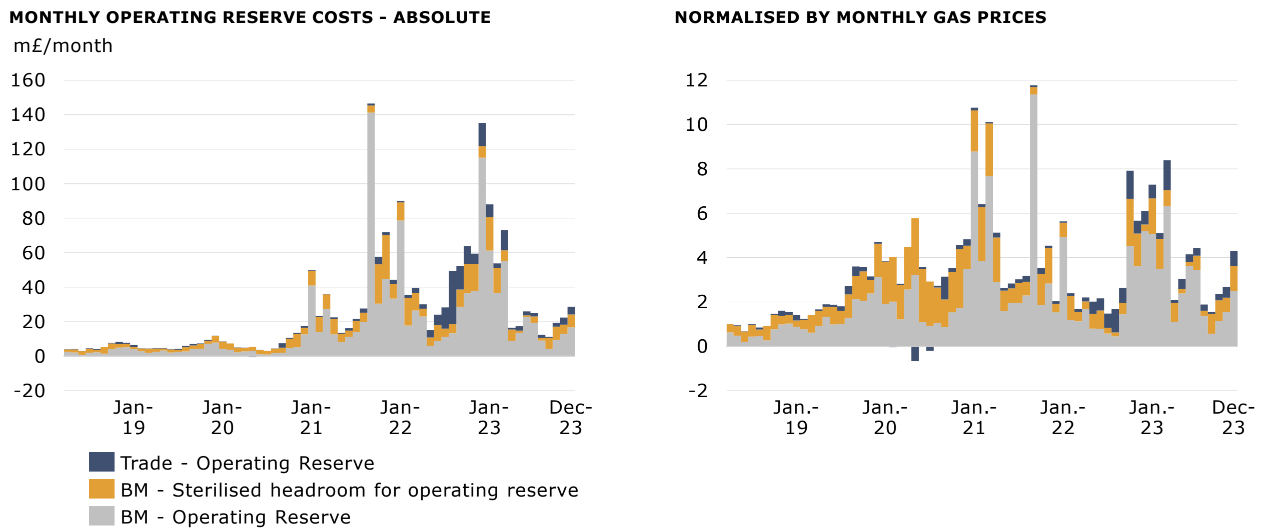
The market is not incentivised to provide reserve capacity where and when needed

Frequency response and most of the reserve products (such as STOR) are procured through auctions at the day-ahead stage. Until recently, Regulating Reserve was procured indirectly through voluntary bids and offers in the BM and through trades. A new day-ahead auction for Regulating Reserve ('Balancing Reserve'), and this should limit, to some extent, the actions taken in the BM to have sufficient 'headroom'.

Regulating reserve is important as it provides flexibility for managing differences between generation and supply. Demand and RES generation forecast uncertainty makes it more challenging to accurately define the actual regulating reserve requirement. At the same time, ESO monitors the 'headroom' from different units (and effective potential for regulating reserve) that is delivered from the market to understand how much additional regulating reserve it would need to make available through BM actions and trades.

The cost of procuring operating reserve has grown markedly in recent years, and beyond the impact of the rise in commodity prices, as shown in Exhibit 24.

Exhibit 24 – Total operating reserve costs



Economic theory suggests that the market should deliver some 'headroom' given the potential for additional income from activation of balancing energy in the BM or as imbalance insurance in anticipation of outages or reduced availability. A market participant can assess the market conditions, take a view on the probability of system imbalance and forego some 'firm' ex-ante market energy trade in anticipation of a potential BM activation with a potentially greater reward. In practice, however, this is not always the case, and there is potential for under-delivery of regulating reserve from the market:

— some 'headroom' may be sterilised because of network constraints as market participants do not have an incentive or the information to provide regulating reserve in the desired locations;

— ESO and market views in terms of regulating reserve needs may differ, meaning that the system wide regulating reserve requirement doesn't necessarily align with the sum of headroom individual market players want to hold to manage their own imbalance or in expectation of activation in the BM;

- ESO has a stronger incentive to avoid demand disconnections and prices rising at VOLL (Value of Lost Load) than market players;
- there are weak incentives on market participants to share information about MEL and PNs at an early stage (e.g. there are obligations to share interim PNs from day-ahead stage, but the accuracy of interim PNs is not enforceable, and PNs don't need to reflect traded positions).

The introduction of the Balancing Reserve product should reduce the need for ESO to synchronise units through the BM or trades to ensure there is sufficient operating reserve continuously. However, it is likely that the BM will continue to be used for some reserve actions. Because Balancing Reserve is procured nationally, some of the headroom may end up not being usable ('sterilised headroom'), and the need for some unit synchronisation through the BM would persist.

The lack of sufficient incentives for the market to provide reserve capacity where and when needed means that the BM is used to synchronise units for headroom. This can lead to inefficient dispatch decisions, as explored in the next sections. The understanding of the underlying value by market participants is also impacted, as they do not have visibility on what is an energy and what is a reserve action in the BM.

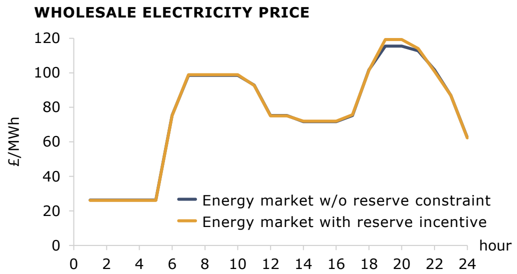
As part of the modelling we have performed, we have also looked at:

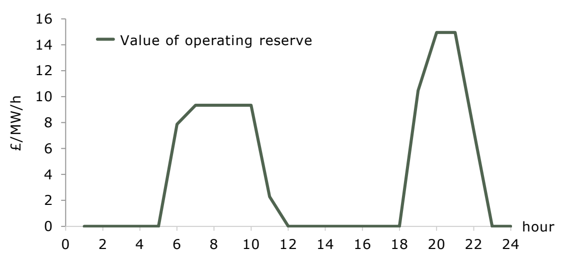
- an ex-ante market assuming no reserve requirement; and
- an ex-ante market assuming a signal for real-time operating reserve provision.

On a selected modelled day we see the following:

- ex-ante wholesale prices would have been higher in some periods if the market was incentivised to deliver the required reserve; and
- there is a value in 'reserving' capacity during the morning ramp and the 'peak' – in all other periods reserve is practically 'free'

Exhibit 25 – Modelled price formation with and without a real-time reserve requirement





5.2.2 Visibility and access

Incomplete ESO visibility of market outcomes and limited access to some resources impacts coherence between wholesale market and balancing

Balancing Market Units (BMUs) submit Physical Notifications (PNs) to ESO that should reflect expected level of output or consumption for each unit. At gate closure (1 hour before delivery), the PNs become 'final' – Final Physical Notifications (FPNs). These are then used for the purposes of taking any required balancing actions in the Balancing Mechanism. PNs can change up to gate closure, but accurate PNs at any given point in time are important from an ESO perspective since they inform whether any advance balancing actions are cost effective. ESO needs to have good visibility beyond the next settlement period (for instance, synchronising a unit with a Notice to Deviate from Zero of 90min and a Minimum Non Zero Time of 6h will have an impact on the next 7.5h).

Generators, large demand user and suppliers notify their PN, which is compared with ESO independent forecast of demand to determine whether the system is long or short. The underlying assumption in the BM is effectively that demand is fixed and that generation is flexible, even though electricity suppliers submit PNs.

The theory is (or at least was) that ESO has visibility of the majority of the generation resources on the system with the PNs acting as a good indication of intended production and through the BM having access to resources for manging energy balance and all other system constraints. As we will see in more detail below, this has been changing over time – there is greater variability and there are increasingly more resources that ESO does not have access to.

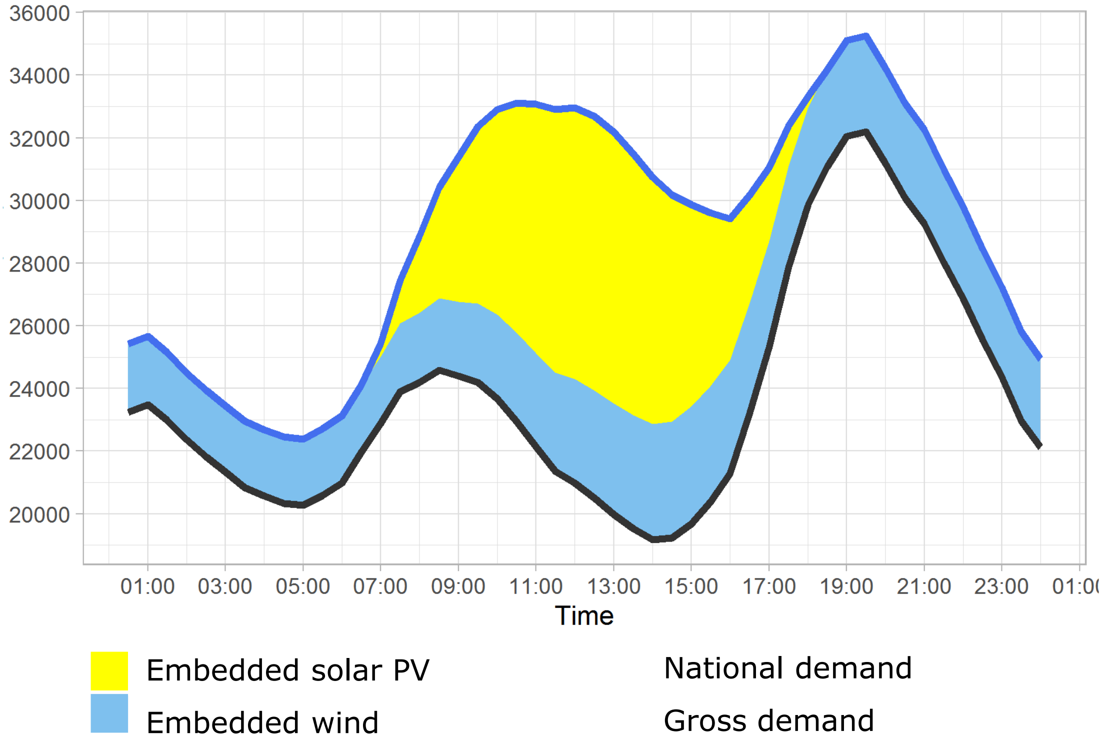
5.2.2.1 Incomplete coverage

Coverage of FPNs is incomplete, particularly for the growing share of flexible non-BM resources, meaning ESO has limited visibility of full market schedules when doing contingency planning

As part of scheduling and dispatch, ESO compares 'national demand' against the sum of FPNs to form a view of the overall system position. National demand is the total demand net of embedded generation. In practice, ESO forecasts total 'gross' demand, and then subtracts embedded RES generation forecasts to obtain the national demand forecast, as shown in Exhibit 26.

Non-renewable embedded generation (for assets such as biomass, engines, or batteries) is not forecasted in the same way as embedded wind and solar PV generation. The impact of non-renewable embedded generation appears implicitly as a reduction in gross demand. The gross demand is also becoming more difficult to forecast due to the increase in demand flexibility initiatives (for which ESO does nothave direct visibility).

Exhibit 26 – National demand forecast on 19/03/2022 (MW)



According to the Grid Code, reaction to market prices by controllable embedded generation and demand response cannot be considered in the national demand forecast prepared and published by ESO. Market parties may then use the published ESO demand forecast (which does not consider price responsive embedded generation), and this can, in turn, have an impact on the market expectations and price formation.

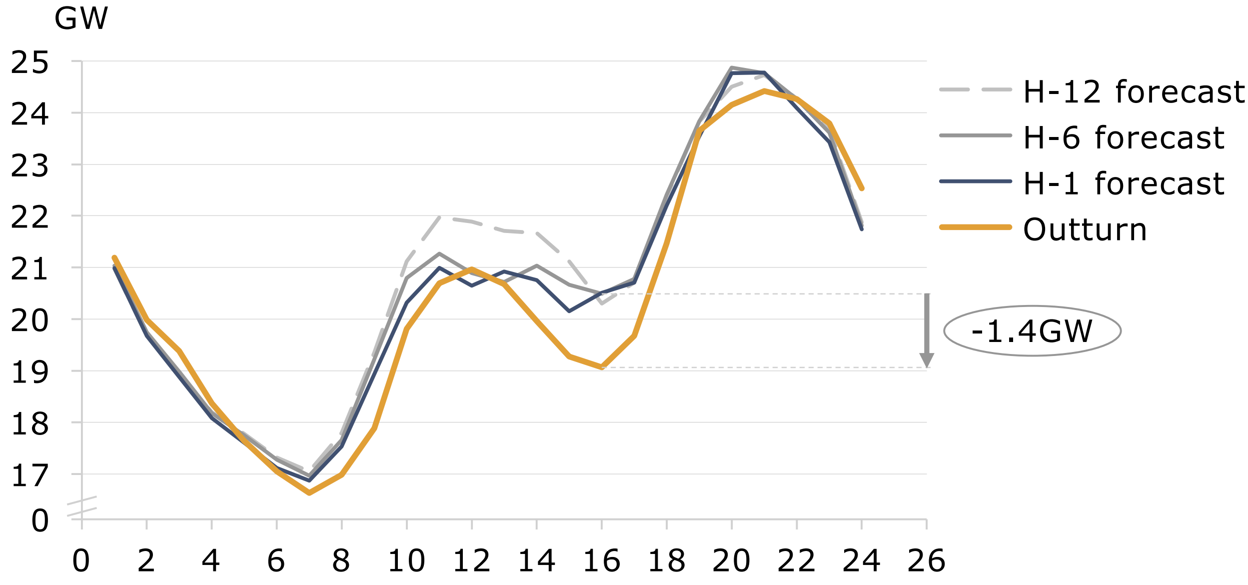
In operational timeframes, the actual national demand can vary significantly compared to the forecast. This is mainly due to the variable output of price-responsive embedded generation and demand. When the market was set up, aggregate PNs were a good indication of the overall market position[19]. However, nowadays this is not always the case because of the growing share of embedded capacity, in particular price-responsive embedded capacity.

Exhibit 27 illustrates a day with a significant difference in outturn national demand compared to the initial demand forecast. In such a situation, ESO

[19] When NETA was introduced the assumption was that demand was inflexible and any unit that is controllable (or price responsive) would submit a PN

has limited visibility and manages change in the expected position of the system close to real time.

Exhibit 27 – National demand forecast and outturn on 09/07/2023



Source: ESO, AFRY analysis

The limited coverage of the system by FPNs leads to uncertainty for the system operator on the demand and supply balance – the share of generation not required to submit PNs is seen indirectly and late by ESO, as a change in national demand. This results in potential for over- and under-procurement of energy and reserve through the day, and ultimately can lead to inefficient dispatch decisions.

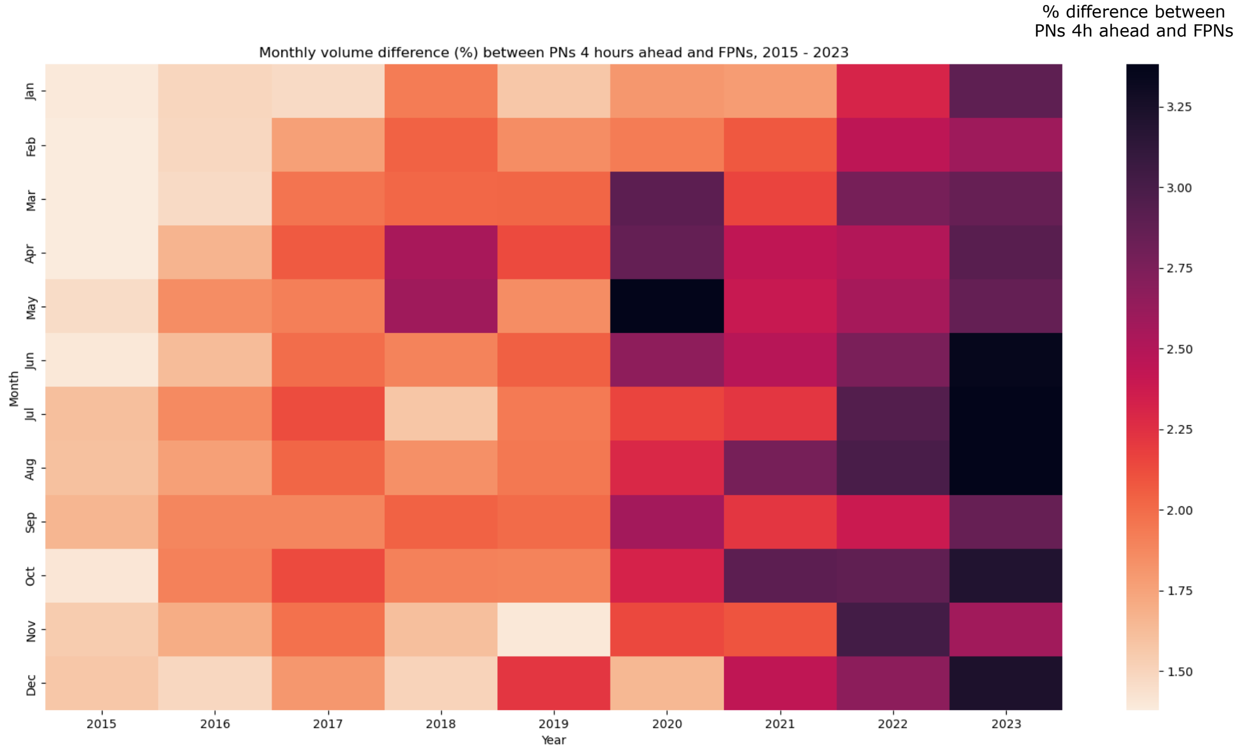
5.2.2.2 Inaccurate information

Schedules change significantly before gate closure meaning ESO decisions are taken with inaccurate information

Ahead of gate closure, the ESO strategy team monitors margin levels and constraints, taking into account demand and renewable forecast, interim PNs, network limits and balancing bids and offers (at the time). The objective of this operational schedule process is to provide an indicative plan for the needs closer to real-time, as well as long notice scheduling actions.

PNs are important for understanding the expected market positions. However, schedules can change significantly before gate closure. Over time, we see a clear trend of increasing deviations between PNs and FPNs. Exhibit 28 shows the evolution of average monthly difference between PNs 4 hours ahead of time and FPNs for generators and interconnection.

Exhibit 28 – Monthly volume difference between PNs 4 hours ahead and FPNs (%)



Note: physical notifications presented in the chart exclude suppliers physical notifications.
Source: ESO

Uncertainty and changes in the period leading to delivery is unavoidable, especially with the growing share of variable and flexible technology on the system. The changes in physical notifications approaching real time have been increasing in recent years, due to:

— the growing underlying variability on the system; and

— increase in volumes traded closer to gate closure.

If ESO foresees a margin shortage a few hours ahead of real time, it chooses between:

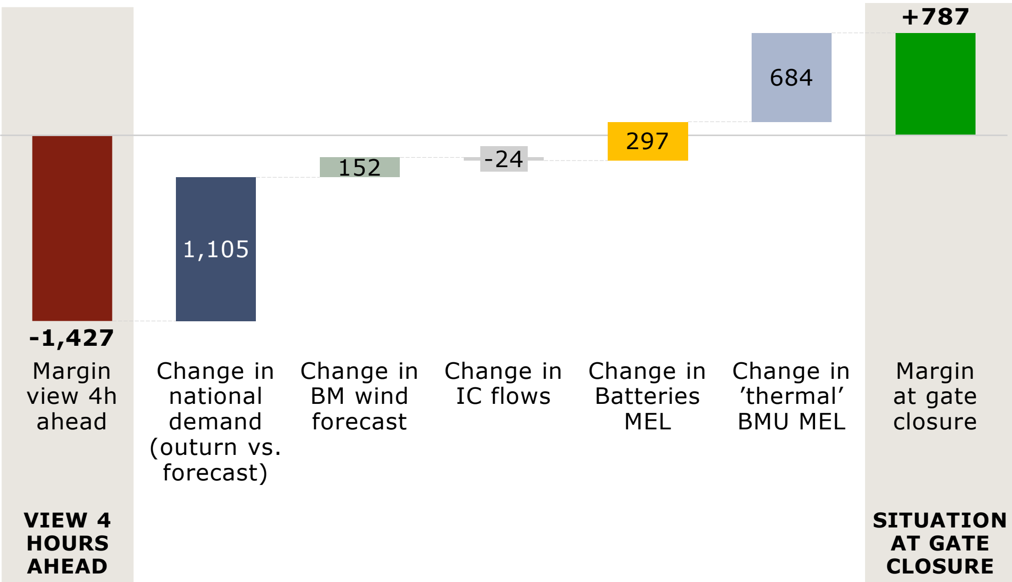
— issuing synchronisation instructions and/or keep units synchronised; or

— waiting for the market to react and deliver the required synchronisations and ‘headroom’.

The latter carries risk. While ESO may prefer to take actions as close to real time as possible to give the market the opportunity to improve their position and for uncertainty to reduce, its obligations for system security incentivise ESO to be more proactive and procure the required regulating reserve ahead of real-time.

Such a situation happened on 1 January 2023. At the time, when the strategy team produced the System Operating Plan (SOP) for the darkness peak 4 hours ahead, it expected a shortfall of around 1.5GW based on the information available (MELs, interconnector flows and RES and demand forecasts). This meant ~10 BMUs needed to be synchronised during the afternoon to ensure sufficient regulating reserve.

Exhibit 29 –Overview of the margin for darkness peak at 5:40 pm on 01 January 2023



Compared to the initial view four hours ahead, at gate closure :

- national demand did not reach the forecast level;
- wind generation was slightly higher than forecast.
- outturn battery contribution at the peak was higher than the operating plan estimate; and
- several thermal BMUs with an interim PN=0 at the peak self-scheduled in the afternoon, resulting in an increase in the overall headroom;

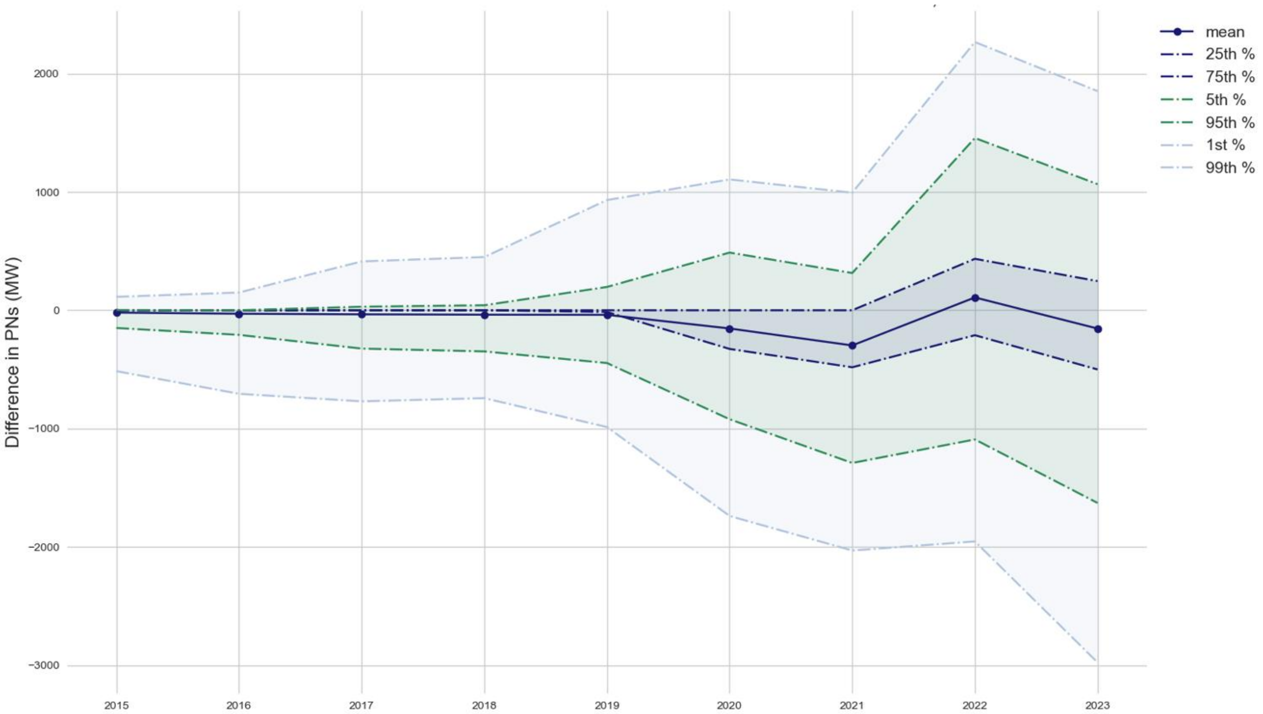
Overall, it appears that, in this example, the market could have solved the anticipated margin need based on the PNs submitted closer to real-time and actual storage contribution. However, this was after ESO issued synchronisation instructions. It is therefore hard to conclude whether the market would have eventually provided the required headroom if ESO had not taken 'early' actions (i.e. it isn't possible to know whether the units that self-dispatched in the afternoon would have done so if ESO hadn't synchronised them through the BM).

Large change in interconnector schedules close to real time

As for other market participants, trading entities on interconnectors submit nomination schedules that can evolve over time until gate closure. While the uncertainty around PNs exists for all type of BM units, the large installed capacity of interconnection, the very high ramp rates, the absence of intertemporal constraints (e.g. no minimum non zero time) and the fact that their schedule reflects evolution of market prices in two markets, change in interconnectors schedules can have more significant impacts.

As presented in Exhibit 30, interconnectors become the single largest source of change in schedules close to real-time.

Exhibit 30 – Distribution of difference in PNs 4 hours ahead and FPNs for interconnector schedules



Note: Dotted lines indicate the % of time in the year the change in interconnector PNs is less than the value on the X-axis, i.e. the green dotted lines show that for 10% of the time the change in interconnector PNs to FPNs was larger than ~1000MW or less than ~1’500MW.
Trades on the interconnectors executed by ESO lead to changes in interconnector PNs; however the effect of these ESO trades are minimal on the chart above given a) they are relatively infrequent over a full year, and b) they tend to occur more than 4hours ahead. In short, this chart mainly shows changes in the interconnectors PNs coming from the market, rather than driven by ESO trades.
Source: ESO, AFRY analysis

This level of change in expected interconnector flows close to real time can cause operational challenges for ESO to manage as it materially changes the overall position of the system.

In theory, assuming the market is balanced beforehand, any change in interconnector schedule is typically accompanied by a corresponding change in generation and therefore would not lead to a large energy imbalance. However:

- the commensurate changes are often on embedded resources, which ESO cannot see, and as a result ESO cannot plan accordingly;
  - in this situation where the source of energy changes between interconnectors and embedded generation, ESO doesn't have proper visibility of the change and ESO will only detect it through rapid swings in the frequency. It practice, it means that additional reserve and response capacity is held to mitigate the potential impact of significant interconnector changes;

- the timing and profile of the MW changes across a settlement period do not align;
  - the right number of MWh are delivered, but the fast ramping of the interconnectors means that ESO still manages the within settlement period imbalances in the first few and last few minutes of the period – this can cause significant frequency deviations if ESO does not act to either schedule other units against the ramp, and/or pre-position the frequency;
- irrespective of the net change in flow to/from GB, the flow into, out of and across different areas can be significant , resulting in potential network reconfiguration to solve (switching circuits / substations, QB tapping etc.) and
- the interconnector flow may be facilitated by capacity ESO was expecting to be providing 'headroom', leaving ESO with a gap in its regulating reserve.

5.2.2.3 Behaviour

Uncertainty on the expected level of system support for balancing by flexible non-BM resources (e.g. NIV chasing or response to retail tariffs)

'Net Imbalance Volume (NIV) chasing' is a practice where market participants try to anticipate the overall net system position and adjust their own positions when they expect the imbalance price to be favourable when compared to their short run cost of operation. In other words, market participants may choose to be imbalanced in the opposite direction of the system to be paid more or to pay less than they would have otherwise paid or been paid if they traded their position in the ex-ante markets (some examples shown in Exhibit 31).

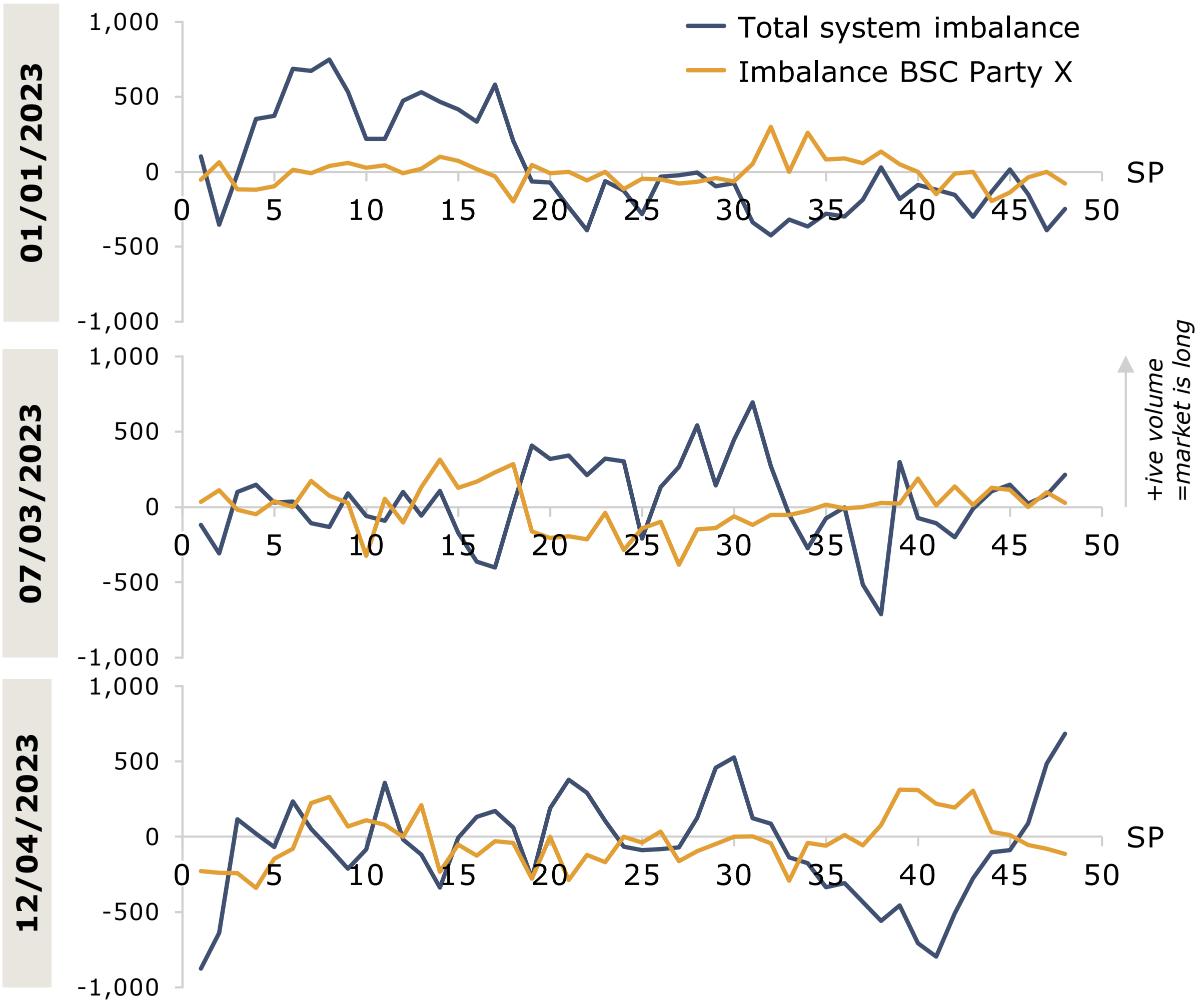
Non-BM unit are not required to submit PNs, and can therefore adjust their output after gate closure. It is easier for market participants to anticipate the length of the system closer to real time (e.g. by monitoring which actions ESO takes in the BM), and therefore makes NIV chasing less challenging and less risky.

In certain circumstances, NIV chasing can be helpful to the system, resulting in overall lower NIV. However, in practice, NIV chasing does not guarantee frequency control and may create additional challenges for ESO:

- NIV chasing is done on a settlement period basis, whereas frequency needs to be managed on a second by second basis;
  - there is a temporal and locational misalignment in terms of the price signals for NIV chasing and the need for ESO to maintain the frequency within a given range at all times;

  - note that other markets are moving towards smaller settlement intervals: the EU design has settled on 15 minutes and some other markets use 5 minute intervals;
- market participant behaviour can be hard to predict and ESO cannot formally rely on NIV chasing for ensuring supply and demand are balanced, even at a settlement period level;
  - ESO may have to take actions as an 'insurance' and subsequently have to 'undo' such balancing actions.

Exhibit 31 – BSC party imbalance vs. system position (MWh)



5.2.2.4 ESO access to resources

Key resources respond to wholesale market signals but are not dispatched by ESO in balancing timeframes

ESO cannot currently access non-BMUs (including embedded generation and demand side response)[20]. Being able to bring more flexible resources into the same set of balancing arrangements (e.g. in the BM) would create more competition, increase liquidity and provide for more efficient dispatch solutions.

Interconnectors are a key source of GB's future flexible resources, but are not fully 'dispatchable' given the link to another SO (which may have conflicting needs) and given the current regulatory arrangements following the UK exit from the EU's internal energy market.

Ahead of balancing timeframes, ESO uses trades to change flows on the interconnectors. Trades are voluntary and take place before gate closure. While redispatch of interconnectors is theoretically possible via trades, difficulties exist in practice. For instance, the flows for NSL are fixed at day ahead and cannot be altered. Additionally, interconnectors with Ireland can be difficult to redispatch given Ireland tends to be subject to similar market evolution as GB (e.g. in case of change in wind patterns), meaning the Ireland may not be able to accommodate GB trades.

In balancing timeframes, relying on interconnectors for cross border balancing could lead to more efficient dispatch solutions. However, following the UK exit from the EU's internal energy market, significant regulatory developments and improved SO-SO cooperation would be required to access interconnectors flexibility in balancing timeframes.

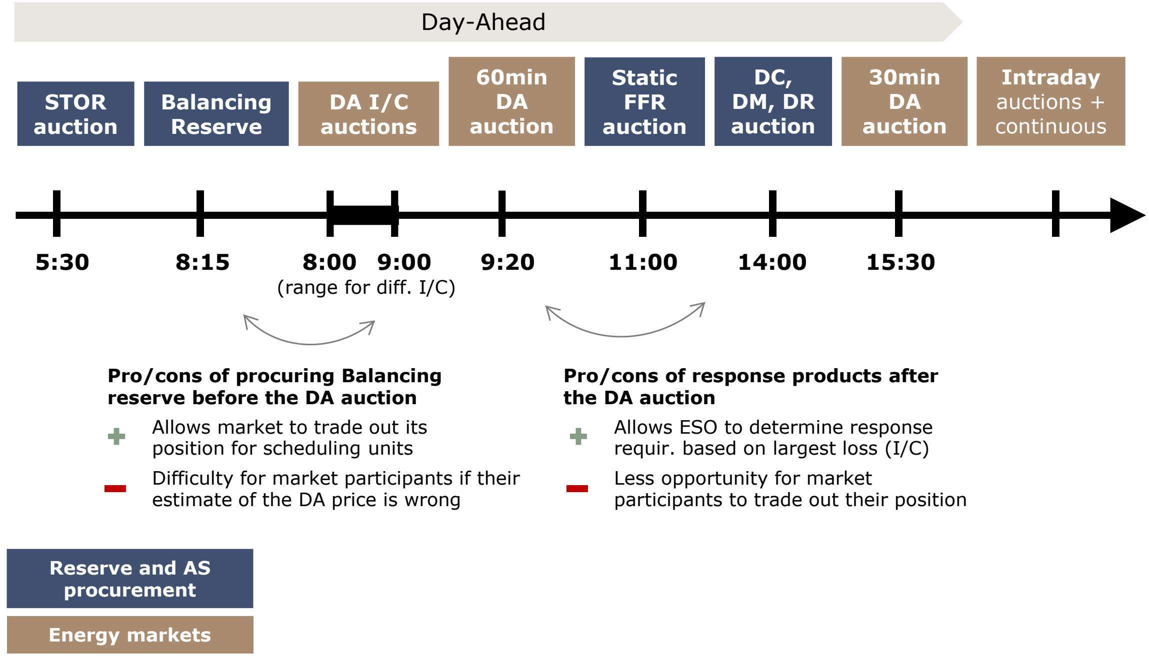
5.2.2.5 Coordination

Sequential procurement of balancing services adds uncertainty to decision making for both ESO and market participants

In recent years, ESO has split out services from the BM into separate pay-as-cleared markets, and most balancing services are procured at different times through several auctions at the day-ahead stage, as illustrated in Exhibit 32.

[20] GC0117 is however aimed at improving transparency and access arrangements across GB.

Exhibit 32 – Overview of energy and ancillary services auctions at day-ahead stage



This has delivered significant benefits to date, but given the number of markets also creates complexities.. The timing of the different auctions for balancing products is a compromise between the opportunity for market participants to trade out their position, and the visibility of the market results for both ESO and market participants. With the sequential procurement of energy and ancillary services at day-ahead, market players need to take decisions in different timeframes against a moving intraday target. The market players can face conflicting incentives from the different auctions and they risk forecast errors when bidding, as they need to anticipate the outcome of subsequent auctions. Provision of some services may be mutually exclusive and there may also be other interdependencies.

The sequential procurement of balancing services also creates uncertainty for ESO as requirements for the various products depend on the outcome of the market.

5.2.3 Intertemporal issues

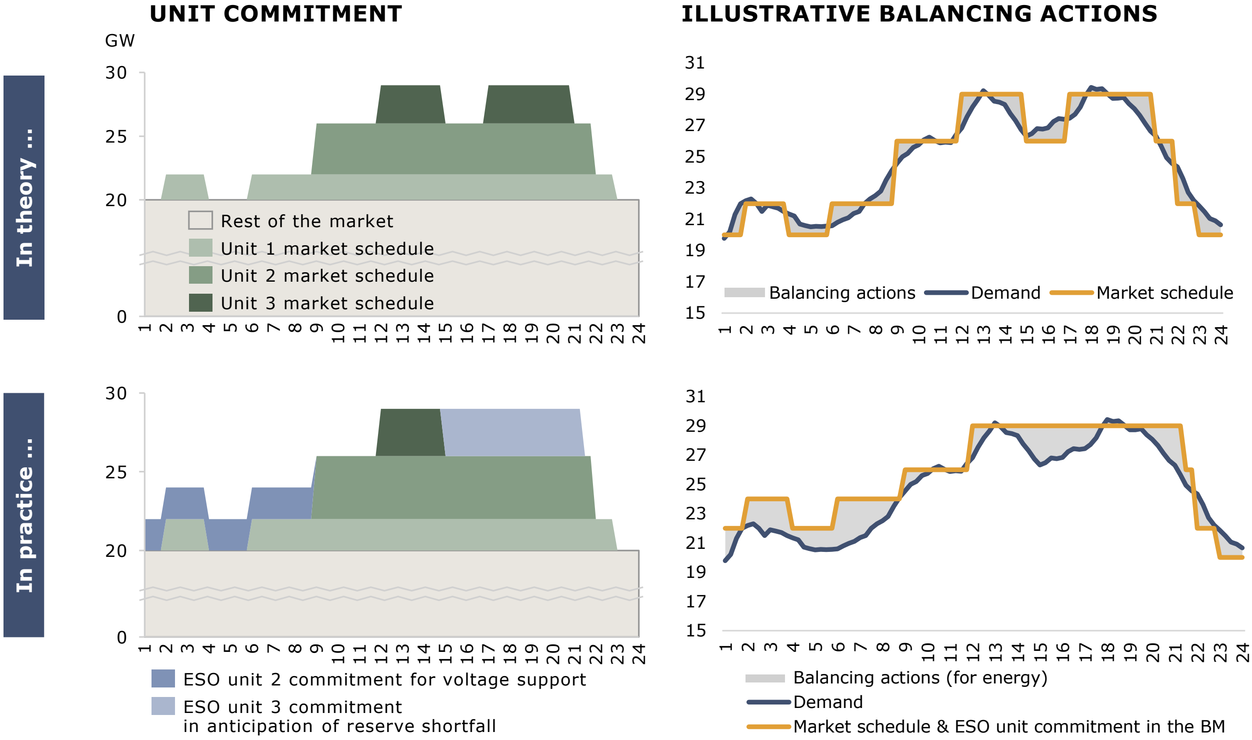
The current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time

The self-scheduling approach, which has been used since the introduction of NETA, has also been important for the overall EU IEM scheduling philosophy. The starting point for all Member States is the use of a self-scheduling model. Under the EBGL, the TSO role is focused on procuring balancing services to ensure secure system operation. Although not explicitly mentioned in the EBGL, TSOs are meant to have a much more residual role, and predominantly to take balancing actions after gate closure.

This expected behaviour in terms of unit commitment and scheduling is presented in Exhibit 33. The unit commitment decisions are taken by the market participants in response to market prices. The role of the SO is 'residual' ensuring energy balance within the settlement period.

In practice, however, ESO ends up taking unit commitment decision for other system needs, but also for securing reserve margin and balancing supply and demand. This is done with the use of the BM on a continuous basis and is a form of a 'rolling optimisation', rather than a discrete solution for a single period.

Exhibit 33 – Expected behaviour with self-dispatch and



With the current market design market participants are expected to manage inter-temporal costs and constraints implicitly, when trading in the different markets or making scheduling decisions (as described in section 2.3.3.1).

The ESO, however, also accounts for these inter-temporal issues in its decisions. It was always expected that some ESO decisions would be taken that would have implications beyond the immediate settlement period, but this was not expected to be significant. These actions would be predominantly to manage thermal limits or voltage needs. It is also possible that the expectation was that such actions would reduce over time. As already discussed, these actions have actually increased and will continue to increase given the evolution of the underlying generation mix and system needs. With more variable generation and demand patterns within day, combined with a lower share of synchronous generation, the importance of intertemporal constraints when making unit commitment decisions is growing.

With the current scheduling and dispatch design, ESO does not directly consider intertemporal constraints with the more intuitive use of complex offers, including technical parameters, over an extended optimisation horizon. Instead, ESO indirectly considers intertemporal issues by accounting for plant dynamic parameters (e.g. minimum zero and non-zero times) when accepting offers for a given settlement period.

5.2.3.1 Timing

ESO is obliged to take proactive decisions with consequences for future periods beyond Gate Closure, which overlaps with the operation of the intraday market

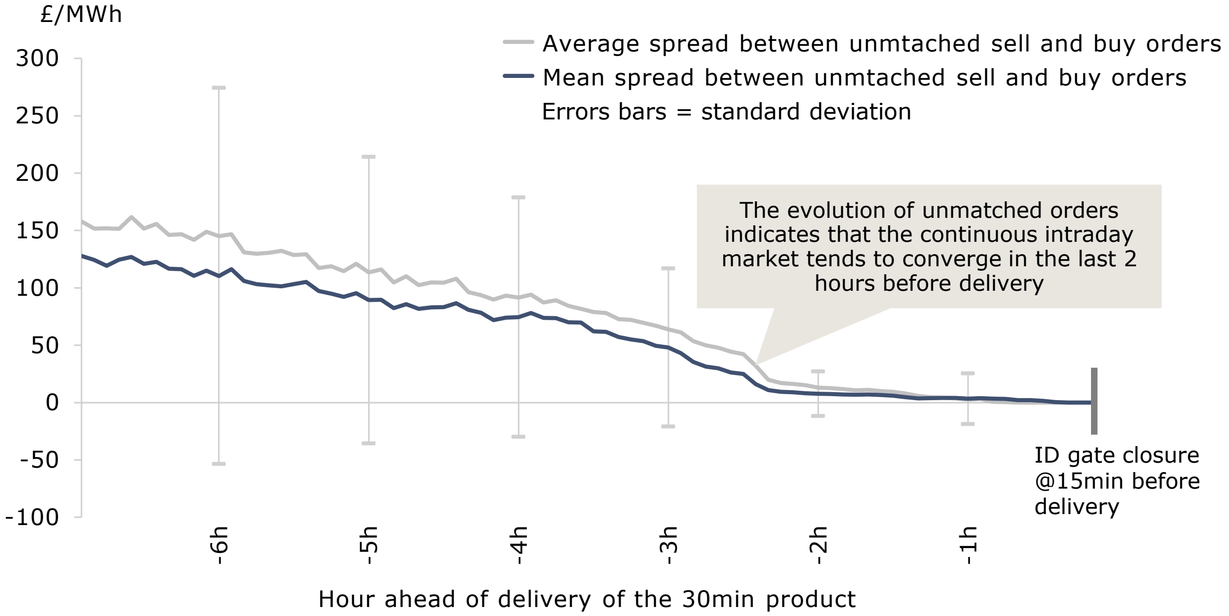
Due to dynamic parameters of thermal units (e.g. minimum on and off times, notice to deviate from zero), ESO, based on its own forecasting, takes unit commitment decisions ahead of the anticipated need. Balancing actions to synchronise thermal units to meet system needs or to ensure sufficient reserve margin are typically taken 3 to 4 hours before the time of the need. This is because of a 'notice to deviate from zero' typically between 1h and 1.5, a ramping time to full load at least 1.5 hours (for a gas-fired unit), and because units are instructed so they can get to full load ~1h before peak for contingency.

We have done a review of the continuous intraday market to understand the interaction between these balancing actions with intertemporal consequences and the intraday market. This review considered the amount of trades and the evolution of unmatched buy and sell orders in the period leading to the delivery of the product. Prices for 30min products appear to converge in the last 2 to 3 hours before delivery, with most of the volumes traded between 2h ahead and gate closure. This finding seems consistent with the assumption that market participants would finalise their positions at the last possible moment to avoid committing to the wrong position.

This finding is important as it demonstrates that the liquidity in the intraday market is poor several hours before delivery, when ESO starts taking unit commitment actions. The system may look short at a point in time when ESO needs to decide whether it requires to secure additional reserve margin.

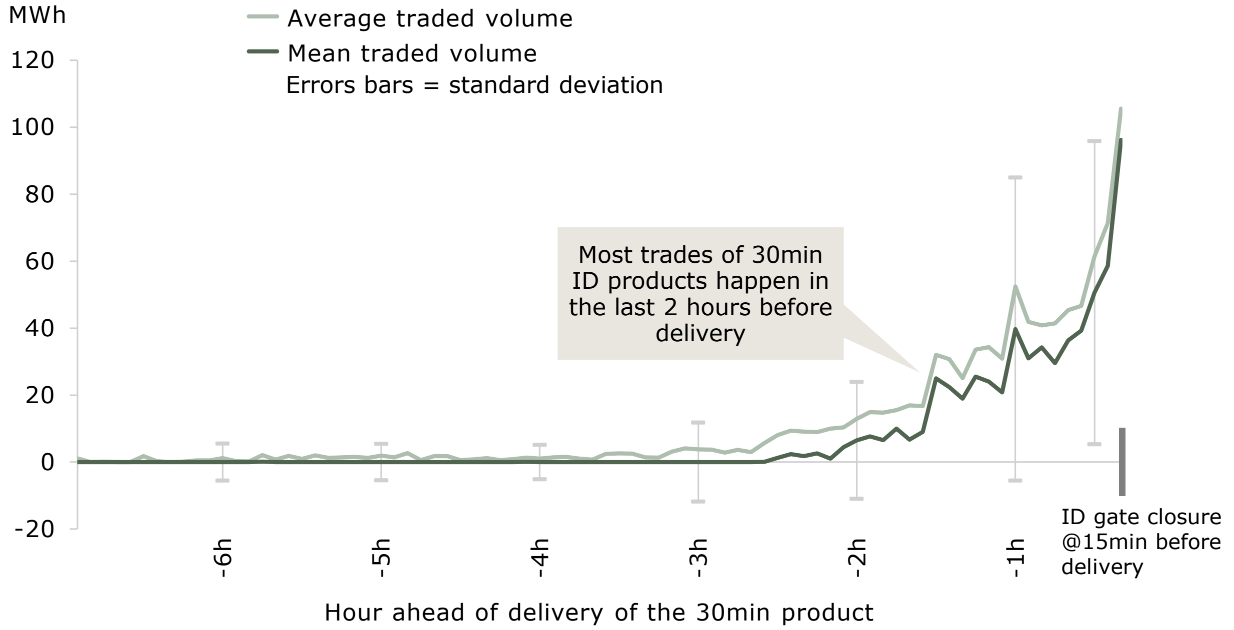
It is however difficult to conclude whether ESO actions drive poor liquidity in the intraday market or whether the intraday market is not facilitating effective repositioning. In any case, market players face conflicting incentives, with a lack of coordination between ESO actions and market scheduling decisions

Exhibit 34 – Spread between unmatched buy and sell orders for 30min product on the GB continuous intraday market



Notes: Analysis based on 18 days in 2023, based on key study days presented in Annex A.
Source: EPEX, AFRY analysis

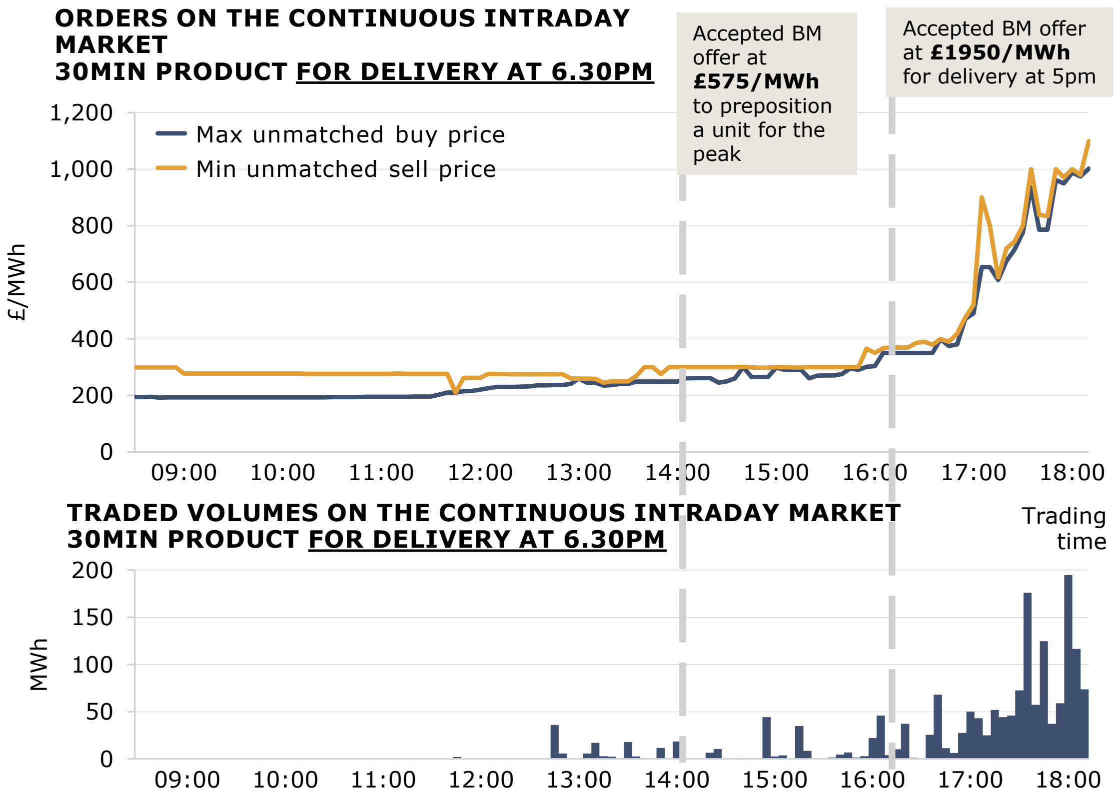
Exhibit 35 – Traded volumes for 30min product on the GB continuous intraday market



Notes: Analysis based on 18 days in 2023, based on key study days presented in Annex A.
Source: EPEX, AFRY analysis

As presented in Exhibit 36 below, on the 03/07/2023, the intraday market did not lead to effective repositioning of units to provide sufficient margin. While ESO accepted expensive BM offers through the afternoon in anticipation of the peak, the intraday order data for delivery at 18.30 show limited market activity through the afternoon.

Exhibit 36 – Unmatched orders and traded volumes for 30min product on the GB continuous intraday market for delivery at 6.30pm on the 07/03/2023



Source: EPEX, AFRY analysis

On this day, the margin for the evening peak was tight. The continuous intraday price reached high peak levels, but not as high as the imbalance price: the intraday price for delivery at 7pm reached £540/MWh, while the system imbalance price for the same period was £1950/MWh. Additionally, the intraday prices converged late, close to delivery time. On this example day, the Balancing Mechanism appears to be supplanting the effective positioning of units in the intraday market.

5.2.3.2 Information

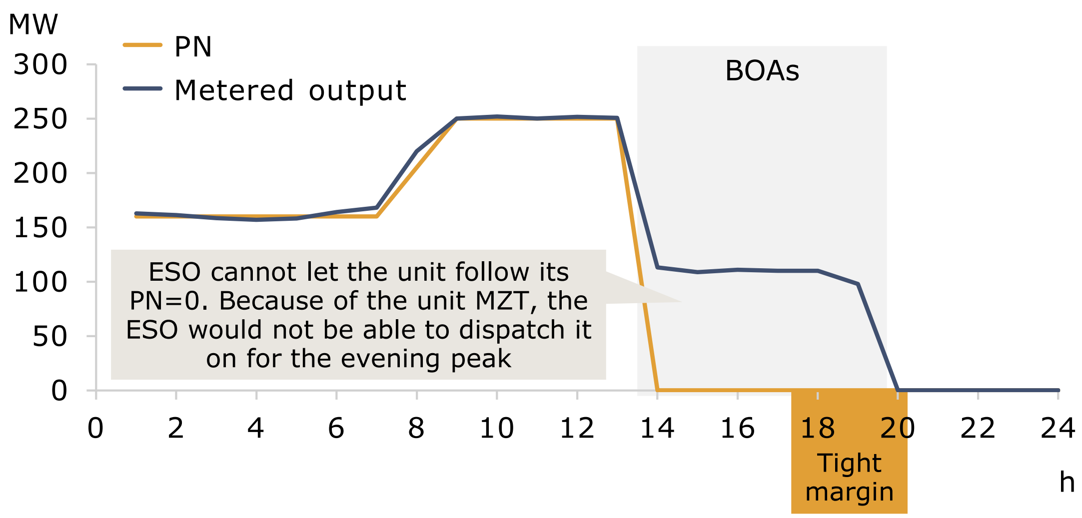
ESO takes decisions with inter-temporal consequences based on imperfect and incomplete forward-looking data

The availability of resources that can respond in short timeframes is limited – some units will be limited by their ramping capabilities and energy-limited units may be in inappropriate state of charge. ESO, as do most SOs, takes long notice scheduling decisions to manage the system. In the GB system, such scheduling decisions are taken indirectly, by accepting offers in the

Balancing Mechanism in some settlement periods to cover for needs in subsequent periods.

For example, the ESO may anticipate that a thermal unit, which is planning to desynchronise early afternoon (according to its indicative PNs) will be needed to meet margin needs in the tight evening hours. Given its min-off time of 6h, the control room needs to take balancing actions ahead of gate closure to keep the unit synchronised all afternoon, as illustrated in Exhibit 37.

Exhibit 37 – Delay de-sync example with min-off time of 6h



Note: MZT= Minimum Zero Time
Source: AFRY Management Consulting

Such extended balancing actions can lead to higher balancing costs than would have been necessary, especially if the unit can anticipate the dispatch instructions from the control room and submits expensive offers.

Market participants are entitled to submit any physical notification in line with their expected output. However, there have been concerns as to whether market participants purposedly submitted zero PN for the evening peak more frequently, together with dynamic parameters than would force ESO to keep them synchronised for an extended period. In a letter from September 2020[21], Ofgem reiterated that under the Grid Code, generators must ensure that their dynamic parameters "reasonably reflect the true current operating characteristics of the BM Unit", and that generators should not use dynamic parameters as a commercial tool to influence balancing payments from the ESO.

This example illustrates a case where:

[21] Ofgem, 29 September 2020, Dynamic parameters and other information submitted by generators in the Balancing Mechanism

- the ESO is restricted in terms of options and is forced to rely on potentially expensive and inflexible BM offers in anticipation of future settlement periods; and
- market arrangements and Grid Code potentially give an opportunity to market participants to extract rents beyond what could have been achieved with different rules and scheduling arrangements.

This issue has been addressed by the introduction a new permanent licence condition to the generation licence called the 'Inflexible Offer Licence Condition' on the 26 October 2023. The new condition seeks to prevent generators to obtain excessive benefits by revising their PNs from a positive MW value to zero MW within the operational day, with a Minimum Zero time higher than 60min and a high BM offer.

Beyond the potential for high accepted BM offers, the fundamental limitation of long notice scheduling action via the BM is the fact that forward-looking data available to ESO is incomplete and non-firm when making such decisions. Unit commitment decisions are taken at a time when other BMUs PNs are not firm, and Bid Offer Data have not been finalised. The structure of unit cost and technical submission data in the Balancing Mechanism does not provide a complete representation of capabilities and cost of resources for future periods. More cost-effective solutions could potentially emerge closer to delivery, but, given its risk management role, ESO is, in most cases, required to take long notice scheduling decisions.

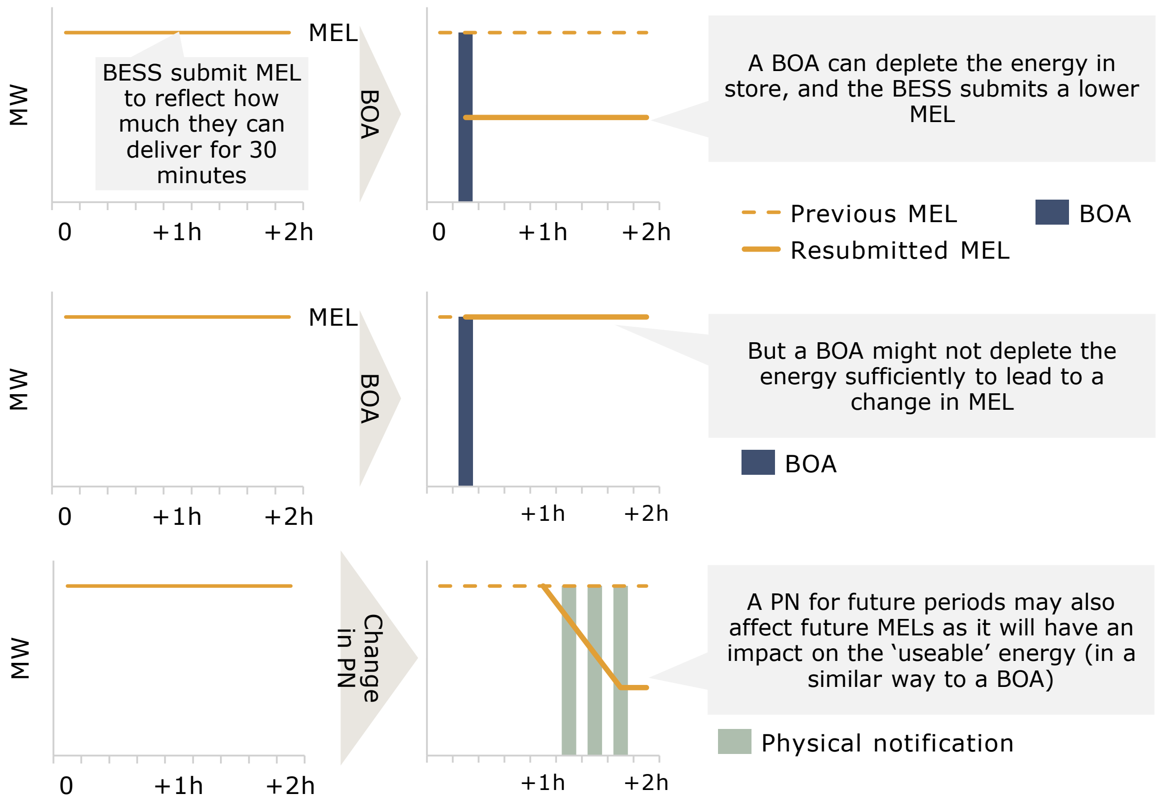
The control room has limited tools in its disposal for 'early' scheduling actions, other than accepting BM bid and offers for extended periods. The control room can agree energy trades with generators or over interconnectors ahead of the Balancing Mechanism timescales through the trading team. However, participation is voluntary, and there may be limited liquidity available.

Impact of energy-limited nature of storage units

Contrary to thermal assets that can be synchronised to ensure their availability in later periods, the energy-limited nature of storage units leads to uncertainty when ESO is making 'advance' scheduling decisions. Given the lack of information with respect to the State of Charge (SoC) of energy-limited units and the inability to formally commit units over a longer horizon, the capability of energy-limited assets for future periods cannot be known with certainty by ESO.

The current approach to manage energy-limited storage assets in the BM relies on the submission of MEL (Maximum Export Limit) to ESO to reflect the capability of the asset in the next 30min, as presented in Exhibit 38.

Exhibit 38 – Current approach for the management of batteries in the BM



Note: MEL=Maximum Export Limit
Source: AFRY Management Consulting

An alternative approach for improving the use of energy-limited units in the BM is the use of real-time signals for available import/export energy. This could be used to estimate the energy state after a BM action (and considering other commitments based on PNs). This would not mean an explicit multi-period optimisation, but it would provide ESO dispatchers with better information, and could result in a relaxation of the '30-minute MEL' protocol.

The main limitation of the proposal is that the energy-limited assets can still change their PN up to gate closure. An 'energy available' (or 'state of charge') signal would only provide some visibility of the available energy to ESO within the hour, but not beyond that timeframe[22]. Even if ESO had clear visibility of the State of Charge of energy-limited assets, it cannot be certain about the 'usable' energy for future settlement periods.

[22] With continuous trading open until 15 minutes ahead of delivery, even the available energy for the next hour is uncerain

5.2.3.3 Transparency

Beyond-the-wall protocols and advance commitments cloud transparency and may distort imbalance pricing

The beyond-the-wall protocol set out in the Balancing Principles Statement states that, when ESO issues a BOA that is expected to be extended by other BOAs after the end of the current BM window ('beyond the wall'), the intention is to base the action on the submitted dynamic and price data for all subsequent anticipated BOA timescales. In other words, for a 'beyond the wall' action (e.g. a unit synchronisation), the principle is that the accepted offer price will stay the same for the whole period of continuous BOAs[23].

A consequence of this beyond-the-wall protocol is that advance commitment actions can distort the system imbalance price on settlement periods around the actual need. Because of the simple bid offer data structure, market participants have to embed their start-up costs in their offer price. Under the current arrangements, there is no means of allocating costs to the imbalance settlement intervals other than those when the energy was purchased/sold. It can result in cross-subsidisation between settlement periods in the resulting system price.

Exhibit 38 shows actual market prices, actual system imbalance price and a theoretical illustrative system imbalance price on the 01/01/2023 to illustrate this situation.

On this day, balancing actions were taken:

- in the morning for inertia and voltage (run-through of units[24]); and
- in the early afternoon to cover for the evening peak

Comparing the indicative theoretical system price and the actual historical system prices shows that part of the cost of the 'early' actions is allocated to those early periods when the need is actually for the evening peak period.

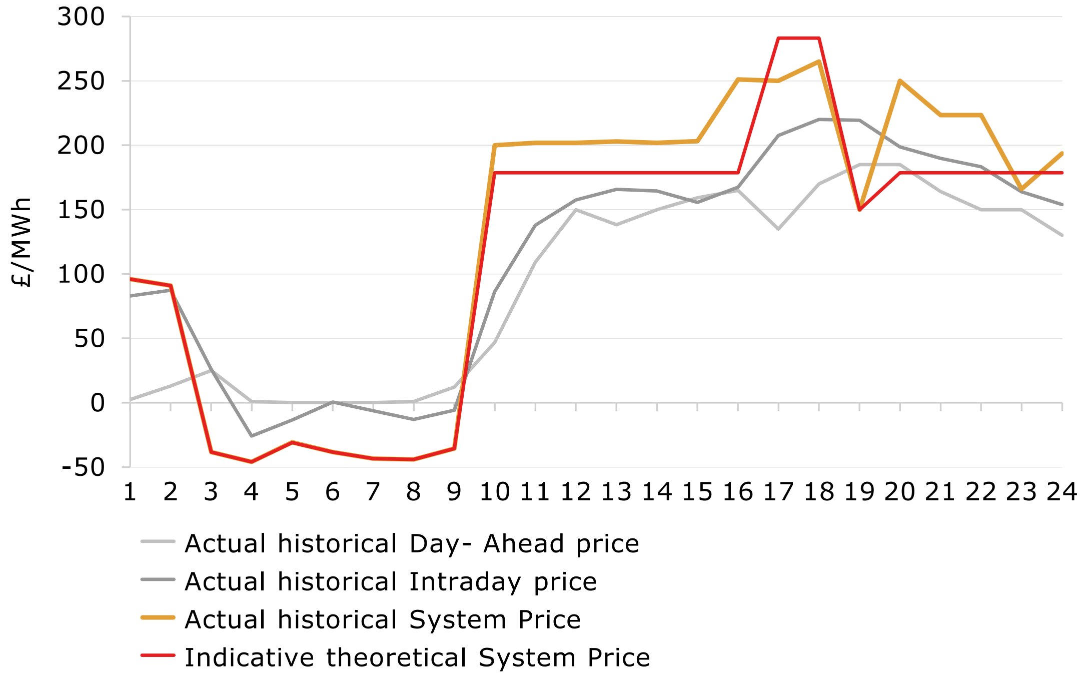
Overall, advance commitment decisions in the BM combined with beyond-the-wall protocols have the potential to make the System Price formation unclear. In turn, a system price not reflecting the system tightness at the adequate time can potentially dampen incentives for market participants to support system level energy balance and lead to under-remuneration for more flexible resources.

---

23 The BMU is theoretically able to change its Bid Offer Data for the subsequent periods, but shouldn't according to this balancing principle.

24 Run-through means to keep a unit synchronised between the planned desynchronisation from the market schedule, until the next planned synchronisation.

Exhibit 39 – Theoretical system price on the 01/01/2023



Note: The theoretical System Price assumes:
- Actual historical gas and carbon prices for the day
- 20% margin in the BM offer (beyond variable costs)
- Average efficiency of 48% HHV out the peak, 44% HHV during the peak
- CCGT start-up costs are recovered over 2 hours at the peak when additional synchronisation is needed

Source: AFRY Management Consulting

5.3 What are the impacts of these limitation on the key objectives?

We are only assessing the scheduling and dispatch arrangements and not the overall market design (including RES support, capacity market, operation of ex-ante markets). Our focus is therefore more on efficient operation.

Are decisions made at the right time and by the right actors?

A key strength of the status quo is the flexibility it allows to the market to respond to changes closer to real-time. This is particularly important in a world with increasing levels of renewables, embedded generation and demand side response.

The GB scheduling and dispatch arrangements are structured around a self-dispatch philosophy. However, we have seen that ESO is taking an increasing amount of balancing actions, including ‘early actions’, for a range of reasons, (managing congestion, ensuring there is sufficient operating reserve and meeting other system needs). Given the intended philosophy of the status quo, we could argue that a lot of the decisions are not taken by the appropriate actors, and potentially not through the most appropriate routes:

— the BM was never intended to be a congestion management process, but is increasingly used to manage network congestion; and

- ESO is taking synchronisation actions to procure operating reserve (headroom) through a process that was not designed to deal with intertemporal constraints.

In terms of timing, some actions are potentially taken too late, and some actions may be taken too early:

- when there is known congestion or other system needs, there is scope for managing this earlier;
- in cases of operating reserve, ESO may be intervening too early given the lack of visibility of what the market is doing and the risks attached to not having sufficient operating reserve in real time.

Are resources allocated appropriately across markets, timeframes and products?

Our quantitative analysis of the specific days we have looked at does suggest there is potential for some improvement in efficiency of dispatch if regulating reserve is procured earlier and in the case where the market is incentivised to respond to some key network limitations. From the small number of days modelled, we found that:

- there is an average social welfare (some of producer and consumer impacts) increase of around 0.1m£ per day with earlier procurement of operating reserve; and
- a 1.5m£ daily reduction of system costs, on average, from improved incentives to respond to network limitations in the ex-ante markets.

These improvements may not be large. However, it is, in any case, difficult for market design changes to bring about significant direct monetary impacts in short-run operating costs, as would be the case with changes in the underlying resource mix, movements in underlying commodity prices and large infrastructure development.

We do need to note that this only captures changes in variable operating costs and the modelling has been done on the assumption that there is no change in interconnector flows. Changes in long term investment and the cost of capital can have a material impact on overall system costs.

Are units positioned effectively for subsequent dispatch?

There are days that the status quo does deliver a ‘market schedule’ that has a large gap from a feasible schedule. ESO then takes multiple actions to ensure system security. The risks lie with ESO and it does not have complete information and access to all units on the system for subsequent dispatch.

Is trading and scheduling on the interconnectors efficient and seamless?

There are inefficiencies in the way interconnectors are used, but this is not necessarily a direct result of the GB scheduling and dispatch arrangements, rather other decisions over which ESO may not be in control of:

- flows are determined on the basis of the national GB price and ignore constraints on the GB power system;
- flows on all interconnectors are currently not determined based on implicit allocation and this inevitably will have some influence on efficiency of flows; and
- ESO cannot redispatch interconnectors through the BM and any redispatch can only be done through trades.

Exhibit 40 – High level assessment of 'status quo'

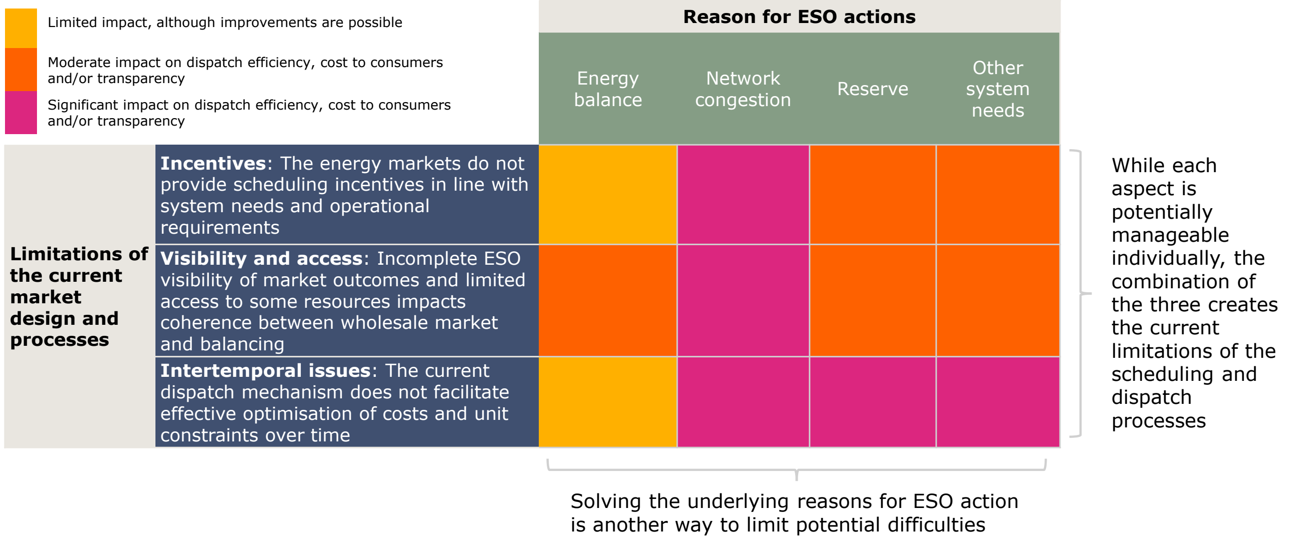
Objective	What works well?	What can be improved?
Efficient operation	Flexibility for market to respond to changes on a continuous basis within-day	— Scope of inefficiencies as market tries to solve the wrong problem at times and ESO uses a mechanism designed for residual balancing to manage constraints and takes decision with incomplete and potentially inaccurate information — Informal inter-temporal constraints/cost optimisation without proper bid structures and imperfect information — Potentially inefficient use of interconnectors: improved scope for I/C redispatch and cross-border balancing could improve dispatch efficiency — Potential underutilisation of energy-limited units in BM
Ensuring operational security	ESO has developed a range of tools and processes to manage the system securely	Weak incentives for market to help with operational security, driving larger volume of intervention by ESO
Support investment	This should be considered alongside the wider market design. However, incentives in particular in terms of location, could be improved, and absence of transparency due to improper inter-temporal optimisation combined with high redispatch can act as a barrier to investment	
Appropriate risk and cost allocation		ESO is taking on more risk than intended under the self dispatch philosophy of the scheduling and dispatch design
Supports competition and creates level playing field	Mitigations are put in place to manage market power	— There is scope for improving competition across different technologies — Potential to reduce market power around import constraints and voltage constraints
Adaptive	The market design lends itself to frequent adaptations in a relatively cost effective way	
Transparency and replicability	ESO is providing a lot of information to help explain the issues from a power system operation perspective	— BM actions are difficult to predict and replicate, even though a lot of information is made available — BM actions are often taken for several reasons (e.g. for both energy and system needs), and the tagging system doesn't fully reflect the reasons for actions — ESO does not have the right type of information and at the right time to ensure efficient scheduling and dispatch

6 Is there a case for change?

We have identified several limitations with the current scheduling and dispatch design. These are summarised alongside the underlying reason for ESO actions in Exhibit 1. The colour-coding attempts to highlight the impact of these limitations and the underlying drivers in terms of efficiency of dispatch and cost to consumers. We see some clear trends:

- network congestion is an important source of balancing action needs; and
- the way intertemporal issues are dealt with is also an important source of inefficiency.

Exhibit 41- Key limitations with current scheduling and dispatch market design



Given the changes in the power system – more RES, more storage, more distributed generation and more interconnection – ESO increasingly acts as a 'central scheduler' to manage risk in a market environment designed for a 'residual balancer'. While the need for balancing actions grows, ESO faces an increasing level of uncertainty and variability, compounding the difficulty and the potential for inefficient decisions. The BM was never designed to accommodate forward-looking decisions and optimisation over multiple timeframes.

The underlying conditions have changed since NETA was introduced, and the case for change of the status quo is clear. The current limitations impact both market participants and ESO, and potentially result in inefficient dispatch:

- **market participants do not have appropriate incentives to allocate resources to meet system needs**;
  - the market is solved for energy balance and market participants do not have appropriate incentives to meet network constraints or other system needs;
  - there is no formal recognition of the option value of reserve for 'using later' and some providers are – at times – used as 'free' reserve; and
  - the BM was not meant to provide forward-looking signals but to reflect the more reactive 'just in time' needs for balancing supply and demand; and
  - actions taken for proactive redispatch end up influencing the Imbalance Price, leading to poor interpretability and predictability of the market.
- **visibility of the market is incomplete and ESO cannot access some resources for balancing purposes;**
  - incomplete ESO visibility of market outcomes impacts coherence between wholesale market and balancing;
  - there are weak incentives for accurate and timely information sharing and this unnecessarily complicates system operation and potentially results in inefficiencies;
  - ESO cannot use some of the resources in the balancing timeframe- these resources include interconnectors and smaller units, which can both be cost-effective sources of flexibility; and
- **the current dispatch mechanism does not facilitate effective optimisation of costs and unit constraints over time;**
  - many balancing actions – including those to resolve system or energy balance – require ESO to take advance decisions outside the immediate balancing window for which it has inappropriate information or tools;
  - with a greater need for ESO-instructed synchronisations and more energy limited units on the system (which in turn arise from the scale of network congestion and the need for improved incentives on participants), it could be argued that the process to optimise intertemporal constraints and costs could be improved – this could also include changes which reduce the extent to which ESO must manage these issues in its scheduling activities .

The market design is not static, and many changes have been made since NETA Go-Live in 2001. There are however some changes underway that will help improve scheduling and dispatch. These are summarised in Exhibit 42.

Exhibit 42 – Limitations managed with the "to be" arrangements

Recent or planned changes potentially mitigating identified issues

	Recent or planned changes potentially mitigating identified issues
Incentives	– **Ongoing network capacity expansion** – **Balancing Reserve will pre-contract some resources to provide reserve availability** – **Half-hourly settlement** – **Ofgem compliance engagement with storage regarding TCLC**
Visibility and access	– **GC117 proposal to reduce BMU threshold to 10MW** – **Local constraint market (pilot for B6) will allow ESO access to more resources**
Intertemporal issues	– **Balancing reserve will reduce the need for pro-active scheduling actions in the BM** – **Potential submission of data on energy limited units (within Gate Closure only)** – **Ofgem inflexible offers licence condition**

These changes can help with a lot of the limitations of the "status quo", but may not be sufficient to ensure that the scheduling and dispatch market design is fully adapted to the realities of the future power system. Some limitations will persist, and there will be a need for further action.

What is less clear is what to change to. There are two high-level approaches:

- giving market participants better incentives and better information to support system operation, which cold include some or all of the following:
  - shorter imbalance settlement intervals;
  - smaller zone sizes;
  - improve signals for ancillary services;
  - improved information sharing between market participants and ESO;
- formalising ESO's de facto role by giving greater control at an earlier stage. Effectively allowing ESO to coordinate unit commitment and operation of energy-limited units, as well as within-day positions.

In the context of the investment needed to deliver on the decarbonisation agenda, the choice between the two is finely balanced. These two high-level options have complex trade-offs, and need to be considered as part of the wider ongoing Review of Electricity Market Arrangements

Annex A Modelling of historical days

A.1 Introduction

As part of the overall assessment, we performed a quantitative analysis based on selected actual historical days. The objective was to understand:

— how the market and ESO behaved on these days; and

— determine a different dispatch with different incentives in place.

A.2 Key study days

Through engagement with various ESO teams, we selected some key study days for further assessment. These were selected to showcase a range of issues. While each key study day is intended to highlight a specific issue, in practice, multiple conclusions and insights can be drawn from each day.

Exhibit A.1 – Overview of key study days and limitations they highlight

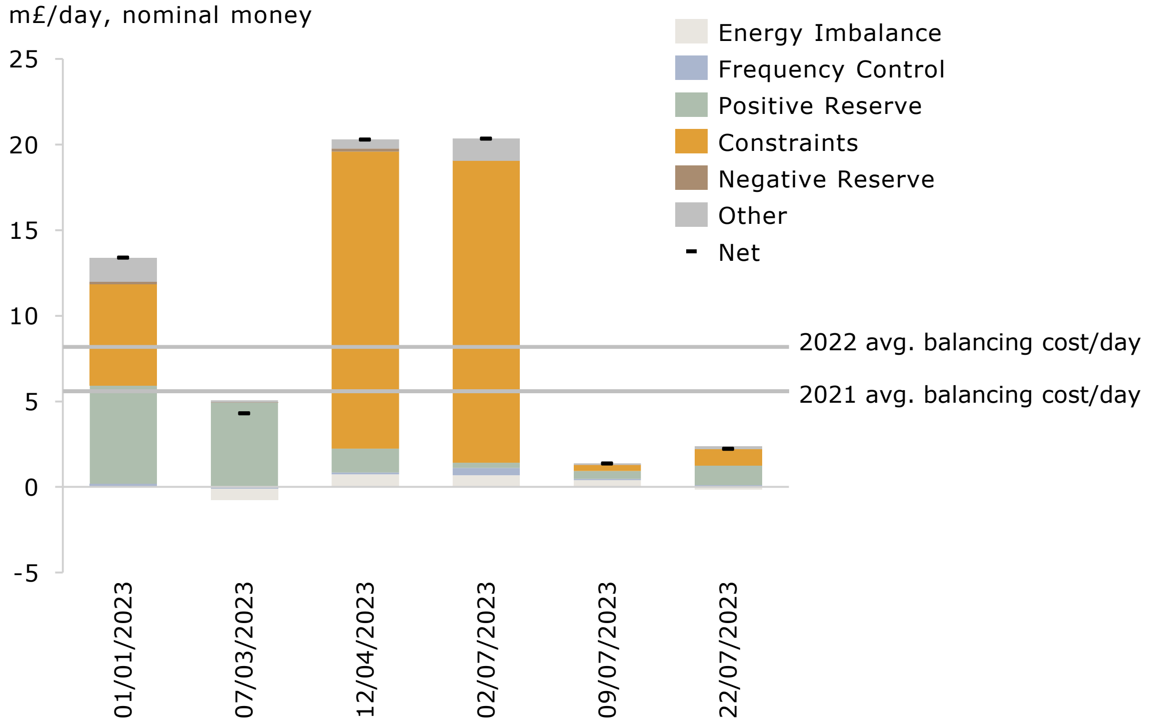
Date	Main event and driver	Significance of the key limitations		
		Incentives	Visibility and access	Intertemporal issues
01/01/2023	A day with very low net demand overnight with wind dropping during the day and a need for gas generation to ramp up. Need for ESO action to manage inertia and voltage overnight and to manage reserve for the evening peak			++
7/03/2023	A day with low wind output, fairly high demand at peak and poor unit availability. Most of the balancing costs were due the procurement of sufficient regulating reserve for the evening peak. Expensive BM offers at £750/MWh and £1950/MWh had to be accepted		+	++
12/04/2023	A day with high wind generation throughout the day and particularly high output in Scotland. A large volume of balancing actions had to be taken in the morning to solve transmissions constraints	++	+	+
02/07/2023	A day with a combination of high wind , high solar PV and low demand. A large volume of actions, relatively constant through the day, was taken to solve thermal transmission constraints and other system needs (inertia and voltage)	++	+	+
09/07/2023	A day with generally low net demand, and average wind generation. The overall volumes of balancing actions was low; the main difficulty came from large change in interconnector schedules close to gate closure		++	

22/07/2023	A day with generally low demand at night and average wind generation during the day. Interconnector sell trades were executed by ESO to solve import constraints in the South East	+		+

Source: ESO

Not all days selected are 'high balancing cost' days. The actual historical balancing costs are shown in Exhibit A.2 and include SO trading and BOAs, but exclude ancillary service costs (such as ORPS payments for mandatory reactive provision and DA procurement of reserve). The daily balancing cost in some of the selected days is markedly high compared to the annual average, driven primarily by constraints (congestion, voltage, inertia etc.).

Exhibit A.2 – Actual historical daily balancing costs for key study days



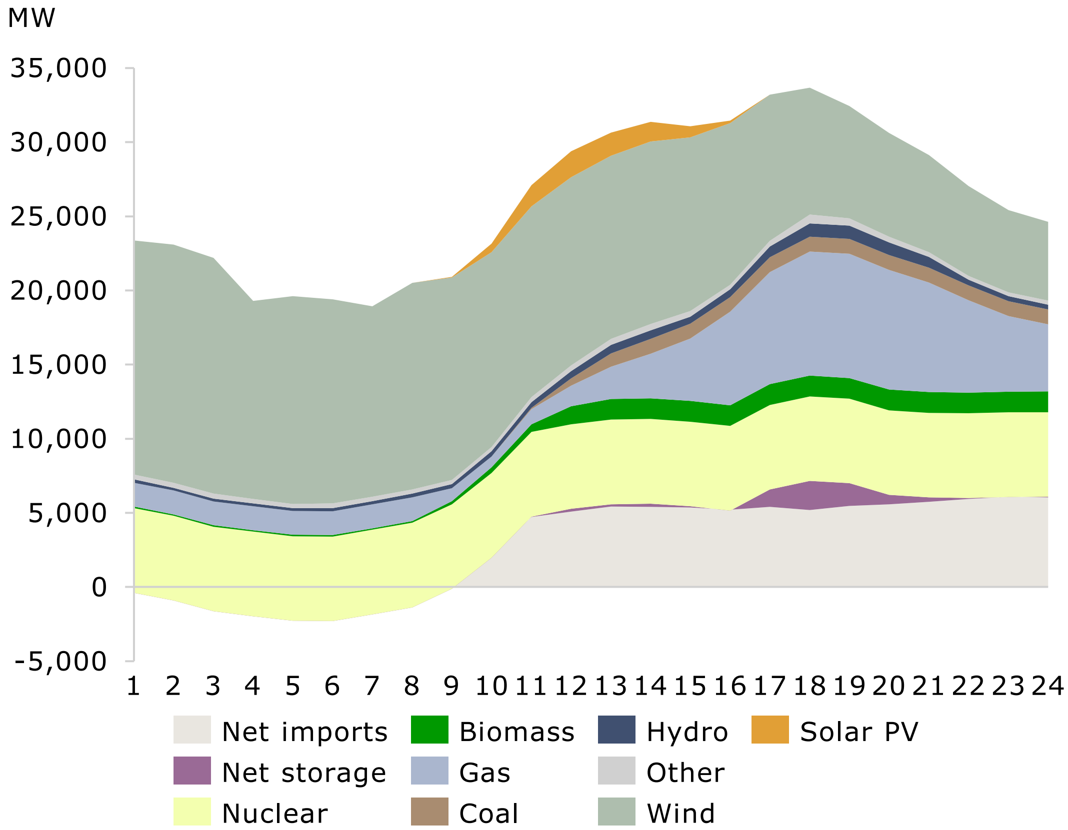
Source: ESO, AFRY analysis

The rest of the section provides an overview of the market and balancing actions for the key study days.

A.2.1 1 January 2023

Typically, on New Year's day electricity demand is fairly low. This combined with above average wind output means that the net demand overnight is very low. As a result, there are only limited CCGTs operating under market conditions overnight. With demand picking up and wind dropping during the day, there is a need for more gas generation to ramp up to meet demand and ensure there is sufficient reserve available. Some biomass is also scheduled in the morning and storage is used at the peak. GB is exporting overnight and importing during the daytime.

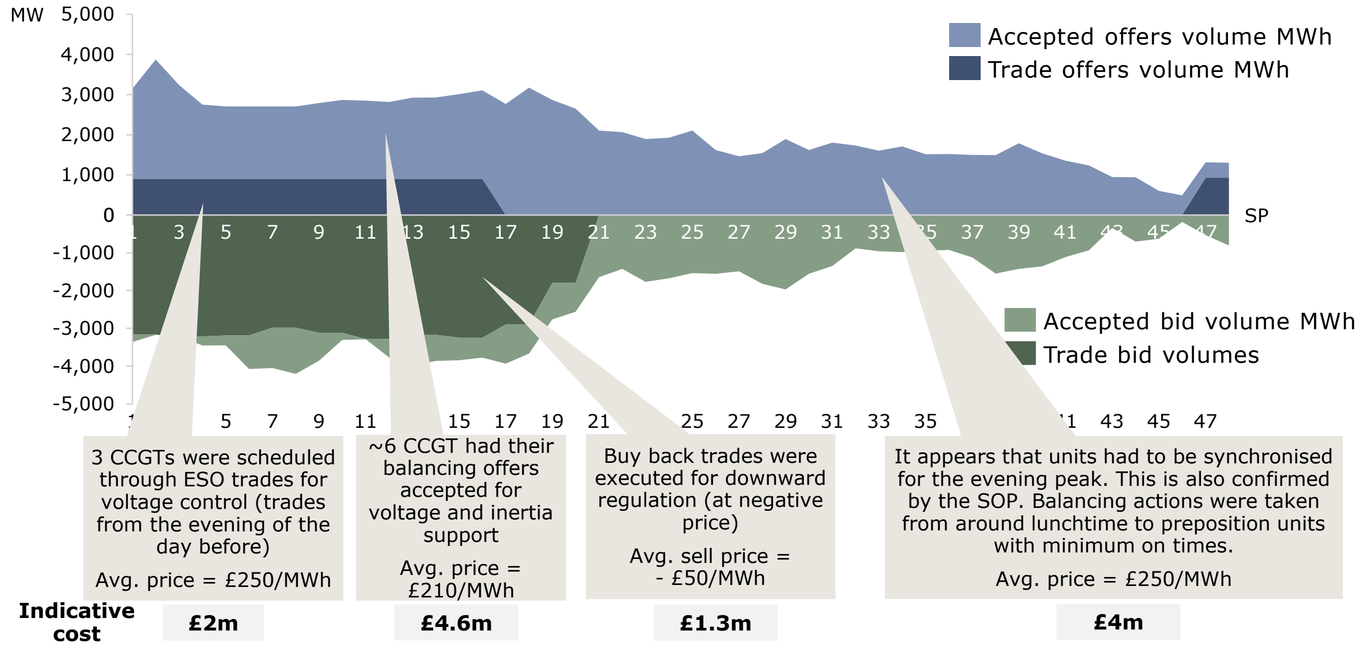
Exhibit A.3 – Aggregated FPN + RES forecast on 01/01/2023



Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

The 'market schedule' was not feasible from an operational perspective. ESO synchronises CCGTs to manage inertia and voltage overnight, whilst changing interconnector flows and curtailing some of the wind. Additional synchronisations were needed later in the day to provide reserve.

Exhibit A.4 – Balancing volumes on 01/01/2023



Source: ESO, AFRY analysis

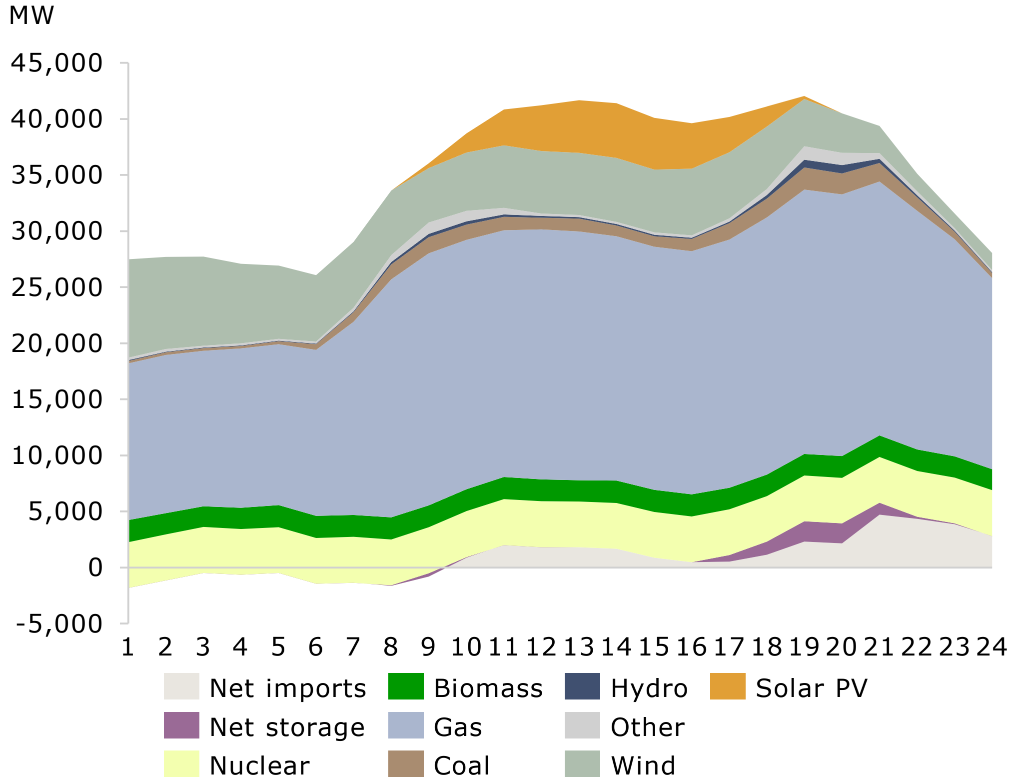
This day illustrates the importance of intertemporal considerations in ESO decisions. Various commitment decisions were taken, including the run-through of units, keeping a unit synchronised between the planned desynchronisation from the market schedule, until the next planned synchronisation and additional synchronisation for inertia and reserve needs.

A.2.2 7 March 2023

On 7 March 2023, power demand was fairly high, and wind generation was generally low. There was however a significant wind forecast error with more than 1GW of additional wind becoming available in the afternoon.

Poor overall unit availability led ESO to issue warming notices to four coal units on the previous day and an EMN.

Exhibit A.5 – Aggregated FPN + RES forecast on the 07/03/20233

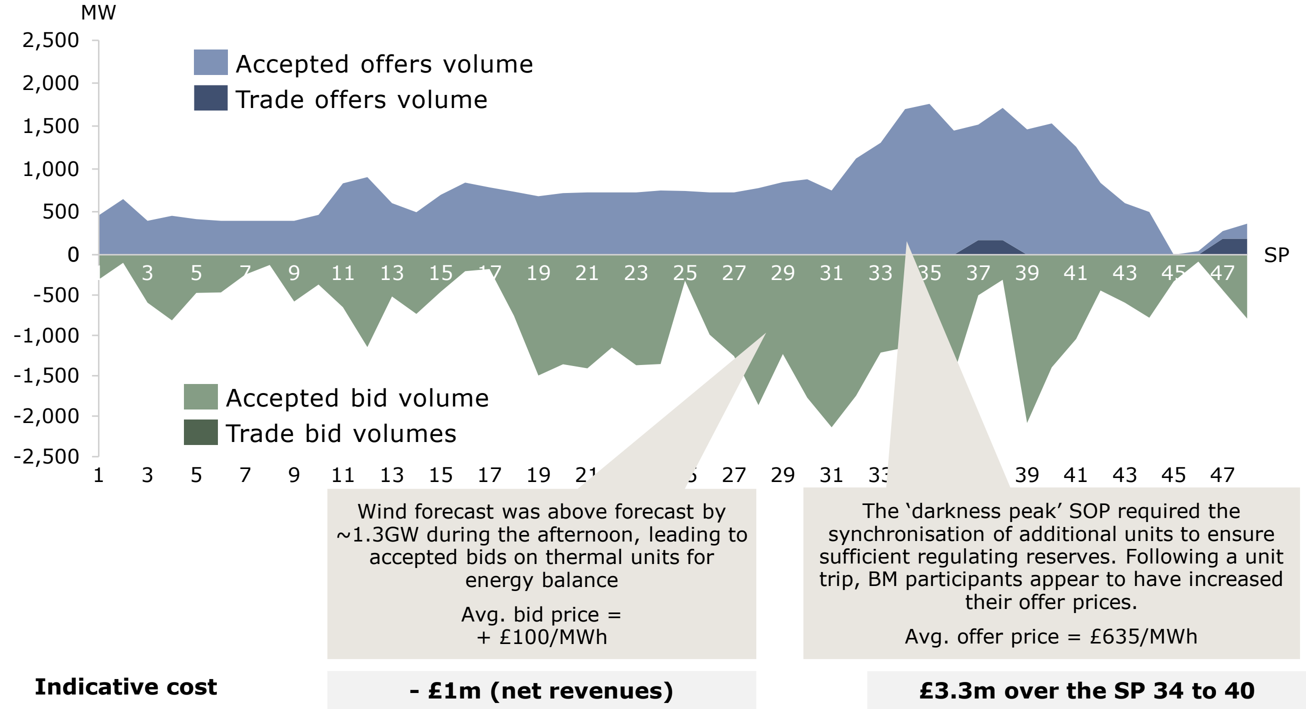


Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

In the afternoon, bids were accepted because of higher than forecasted wind generation, while offers were accepted to ensure sufficient regulating reserve. This day is a representation of a typical day with 'system tightness'. Most of the balancing costs were due the procurement of sufficient regulating reserve for the evening peak (and indirectly sufficient available energy to meet demand). Due to plant unavailability during the afternoon and evening, expensive offers at £750/MWh and £1950/MWh had to be accepted.

This high balancing cost is to some extent reflective of scarcity, and is to some degree desirable in a well-functioning market to deliver appropriate operational and investment signals. From a pure 'capacity margin' perspective, however, the overall price level and cost appears somewhat disproportional.

Exhibit A.6 – Balancing volumes on the 07/03/20233



Source: ESO, AFRY analysis

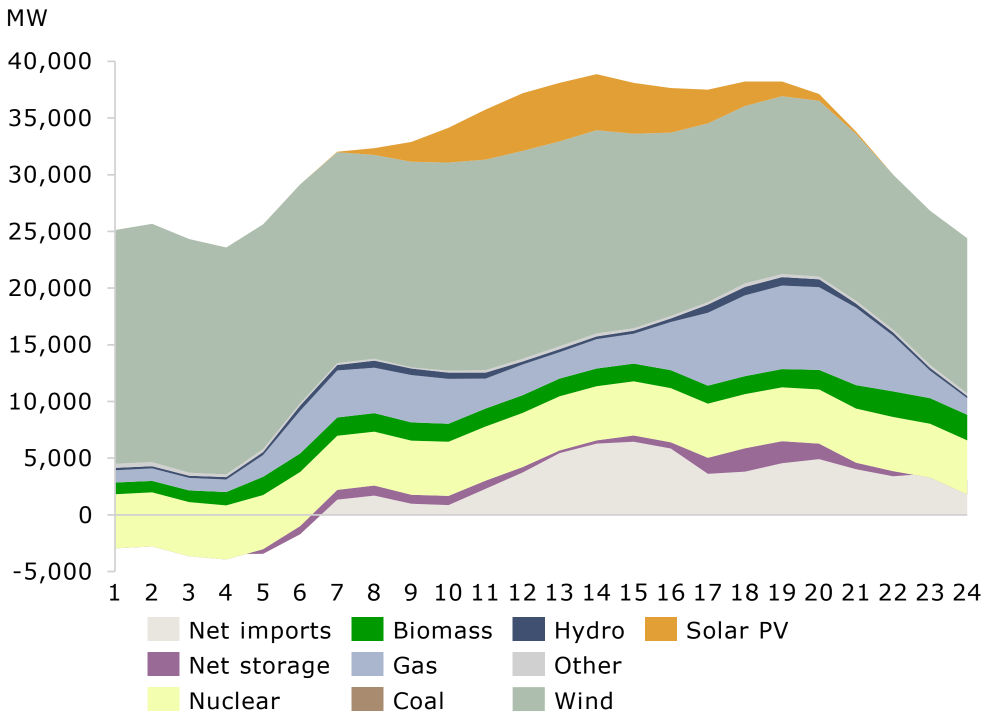
On this day, thermal units were bid down in the afternoon while others were synchronised over the same periods. This highlights the challenges ESO faces in terms of forecast uncertainty and management of intertemporal constraints: ESO anticipated a margin shortfall at the peak and synchronised units in the afternoon to ensure sufficient headroom; in the meantime, the wind generation was above the forecast, meaning other thermal units were bid down (to their SEL) in the mid-afternoon.

This day also illustrates the impact of limited visibility and access: the margin shortfall at the peak did not materialise to the anticipated level because national demand did not reach the forecast level, potentially due to controllable embedded generation and price responsive demand behaviour.

A.2.3 12 April 2023

The north-south export constraints were active through the day. From around 8am, B4, B6 and B7 constraints were active. Wind bids had to be accepted to resolve thermal limit constraints. The wind in the north ends up having to be replaced by thermal generation in the south. Pumped storage behind constraints submitted positive PNs and had to be bid down. This happened for most settlement periods of the day.

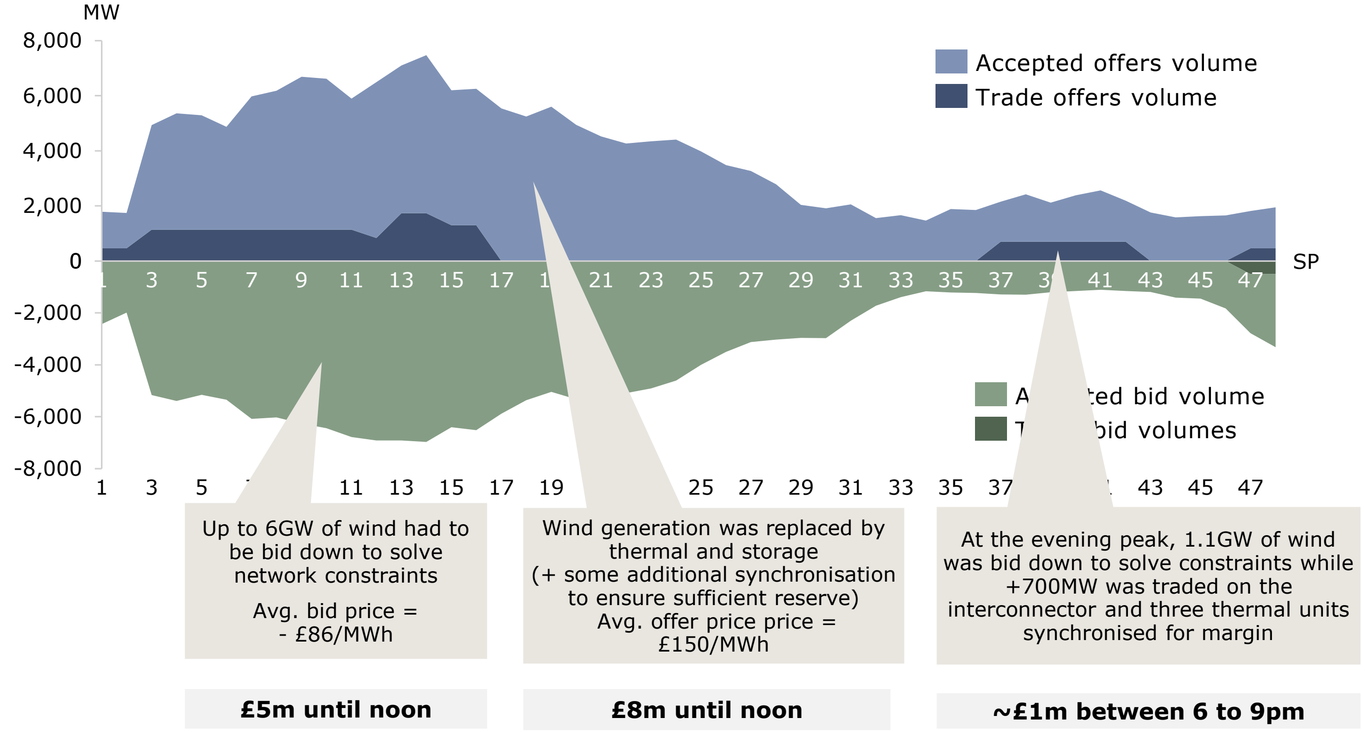
Exhibit A.7 – Aggregated FPN + RES forecast on 12/04/2023



Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

A large volume of actions (>6GW) had to be taken in the morning, to resolve transmissions constraints.

Exhibit A.8 – Balancing volumes on 12/04/2023



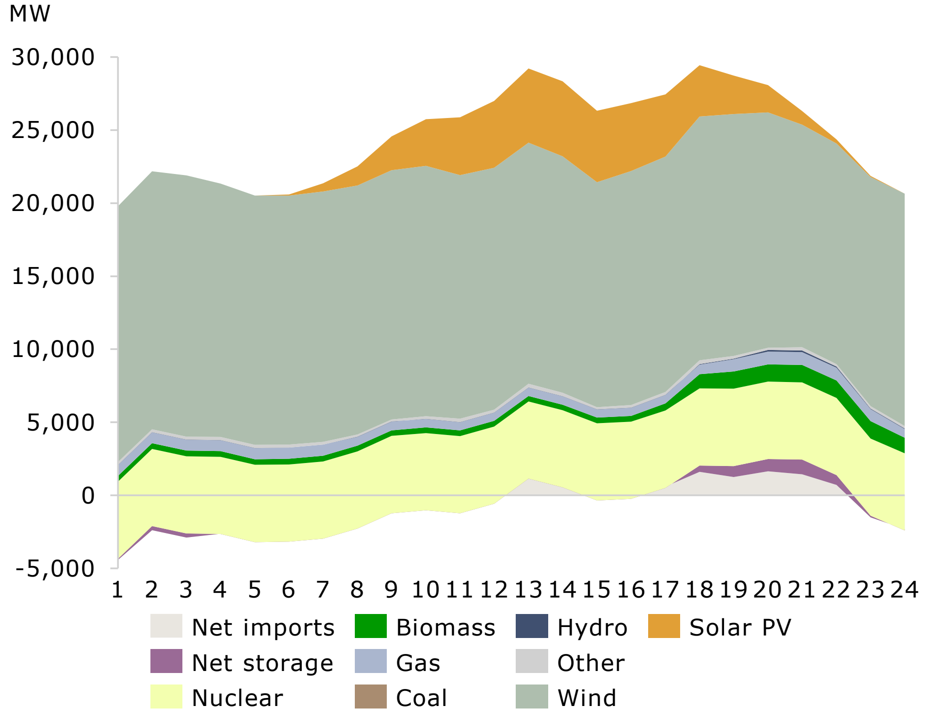
Source: ESO, AFRY analysis

This day is an example where market participants incentives were not aligned with the operation of the system, leading to large volumes of redispatch. The operational difficulty of the redispatch was compounded by intertemporal issues (due to additional synchronisations) and poor visibility of the system.

A.2.4 2 July 2023

The north-south constraints were binding. The market had delivered very little thermal generation (other than nuclear). There was a need to curtail wind and replace with thermal generation to both manage the thermal limit constraints and also ensure there is sufficient inertia and regulating reserve.

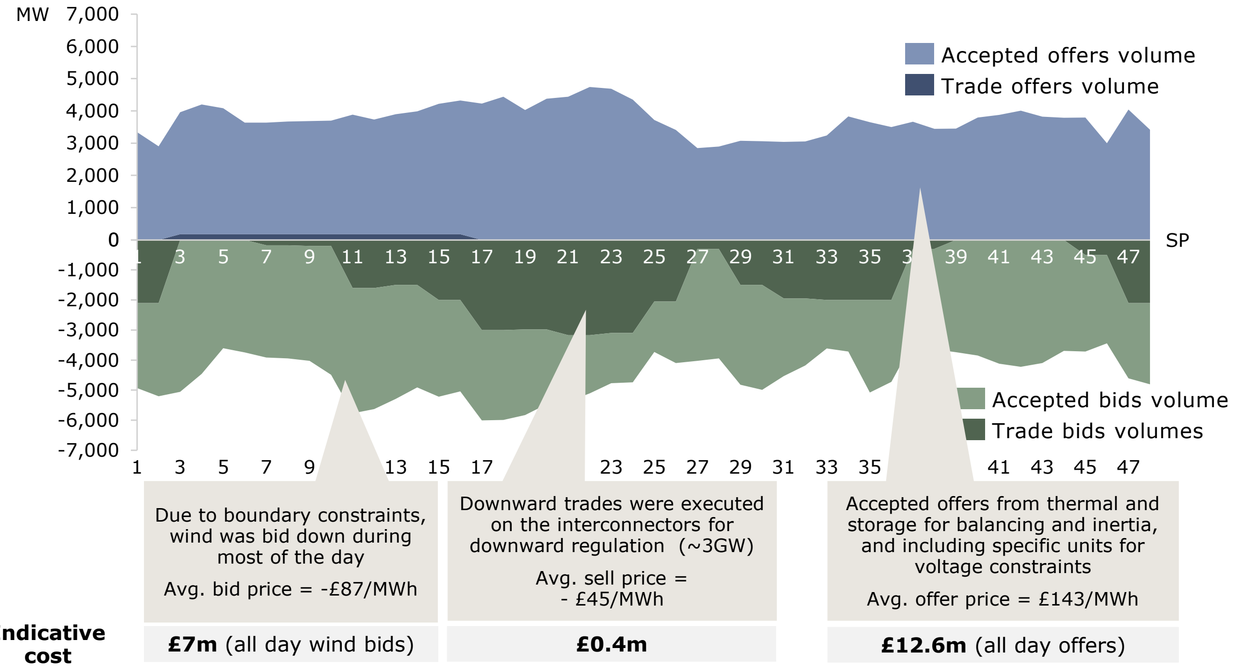
Exhibit A.9 – Aggregated FPN + RES forecast on 02/07/2023



Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

A large volume of actions, relatively constant through the day, was taken to solve thermal transmission constraints. Day ahead and intraday prices were negative in most settlement periods.

Exhibit A.10 – Balancing volumes on the 02/07/20233



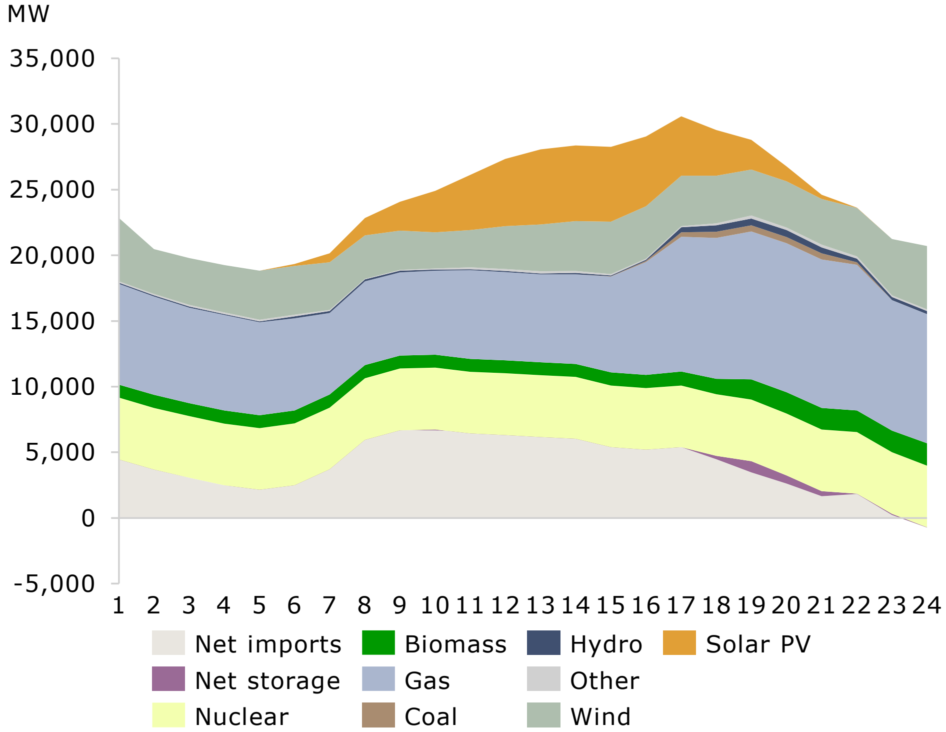
Source: ESO, AFRY analysis

This day is an illustration of the insufficient incentives on market participants to support the system operability, and the resulting large volume of actions in the BM.

A.2.5 9 July 2023

This was a day with generally low net demand, and average wind generation. For the evening peak, with wind generation quite low and solar PV dropping, there is a need for more gas generation to ramp up to meet demand and reserve needs.

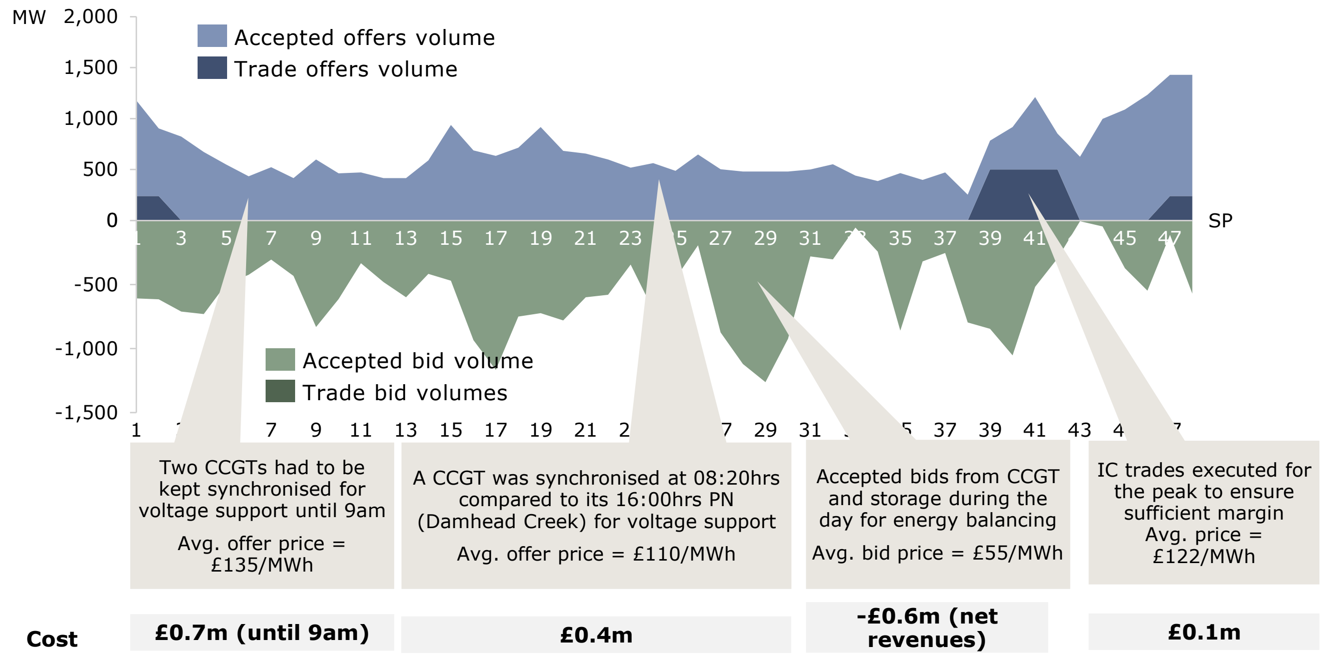
Exhibit A.11 – Aggregated FPN + RES forecast on 09/07/2023



Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

There was generally a low volumes of balancing actions on this day. A CCGT had to be synchronised during most of the day, for voltage support and provision of reserve.

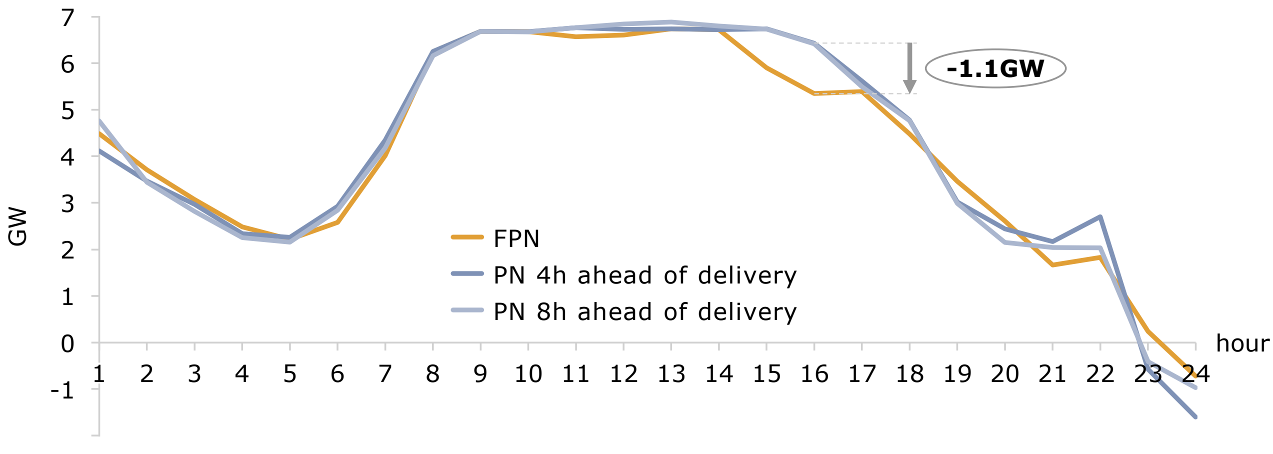
Exhibit A.12 – Balancing volumes on 09/07/2023



Source: ESO, AFRY analysis

On this day, the interconnector schedules changed significantly close to gate closure for delivery at 3 and 4pm, as presented in the exhibit below.

Exhibit A.13 – Total interconnector physical notifications (net imports to GB) on 09/07/2023



Source: ESO, AFRY analysis

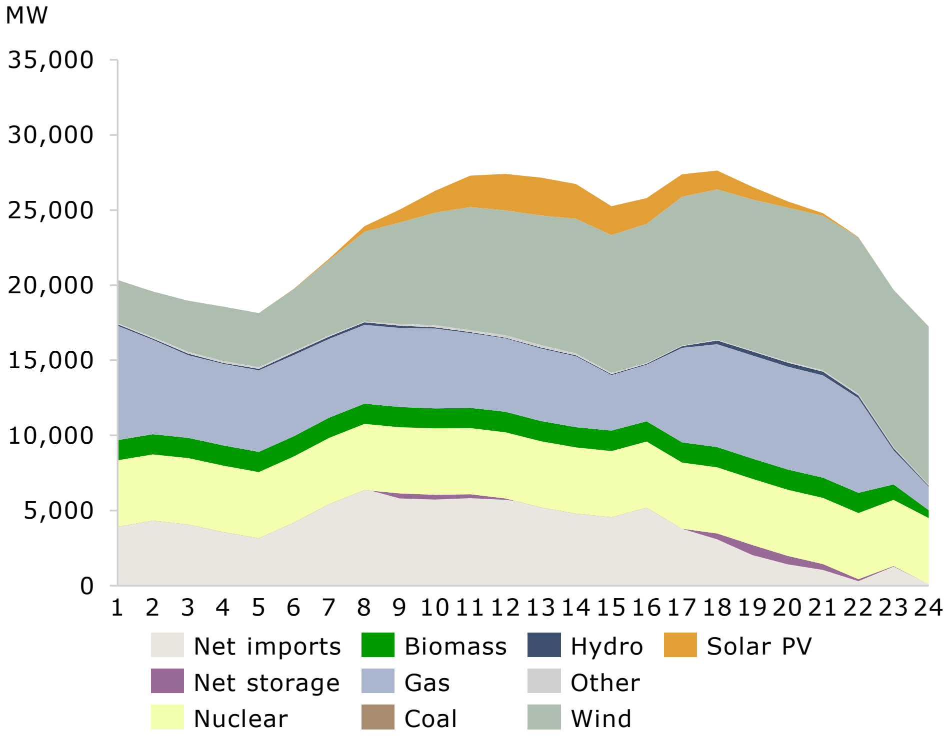
This level of change close to delivery leads to difficulty for the control room as it materially changes the overall position of the system.

On this day, the national demand did not materialise as expected in the forecast in the afternoon. The likely explanation of the situation on this day is that non BM units increased their generation and traded the output on the interconnectors. This illustrates the limitation arising from the limited visibility of the market by the system operator.

A.2.6 22 July 2023

The 22/07/2023 was a day with generally low demand at night, average wind generation during the day, and significant interconnector imports.

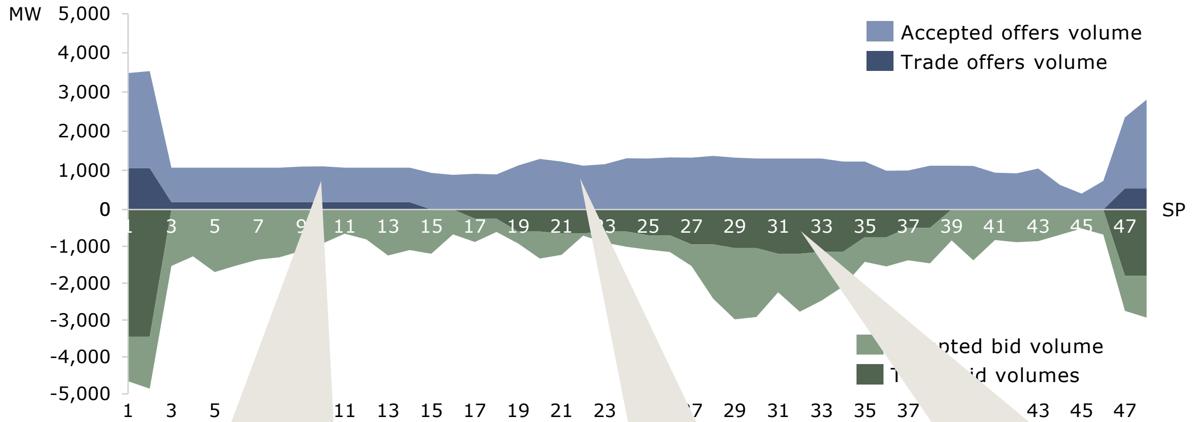
Exhibit A.14 – Aggregated FPN + RES forecast on 22/07/2023



Note: generation mix on the day based on FPNs for all units (except wind BMU); and forecasts for wind (embedded + BMUs) and solar PV forecast (1h ahead)
Source: ESO, AFRY analysis

On the 22/07/2023, interconnector sell trades were executed to solve import constraints in the South East. CCGTs were kept synchronised all day for voltage and reserve needs.

Exhibit A.15 – Balancing volumes on the 22/07/2023



	Three units were delayed desync (from the day before), one until the morning, the other two through all day, for voltage support initial and reserve later in the day Avg. offer price = £125/MWh	High wind generation compared to forecast leading to accepted bids on thermal and storage Avg. bid price = £25/MWh	Sell trades were executed on IFA1, Eleclink and NEMO to solve import constraints in the South East (BOSELEX) Avg. sell price = £25/MWh
Indicative cost	**£1.8m**	**- £0.2m (net revenue)**	**- £0.2m (net revenue)**

Source: ESO, AFRY analysis

On this day, delay desynchronisation decisions were taken, which had with intertemporal consequences (potentially impacting the formation of the imbalance price). This day is also an illustration of the incentives on the market not reflective of system constraints: import constraints were binding in the South East due to 'excessive' import flows on the interconnectors. ESO had to countertrade to solve the import constraints.

A.3 'What if' modelling of historical days

A.3.1 Inputs

We used actual historical data from the key study days in our power market model, BID3, including:

- renewable output and demand forecasts for the day;
  - however, we intentionally assumed no wind forecast errors between the initial dispatch and redispatch runs to avoid capturing changes because of forecast errors;
  - the ability to redispatch in response to changes in the supply and demand balance is key in power markets and the scheduling and dispatch market design should support this;
- historical BOD and cost-reflective data for the market run, based on historical commodity prices for each month (gas, coal and UK ETS)
- nuclear availability based on actuals and interconnector flows based on FPNs;
- regulating reserve needs calculated based on the wind generation; and
- historical boundary constraint availability for each day.

A.3.2 BID3 model

AFRY's proprietary BID3 model assesses both physical operation and economic behaviour across integrated electricity markets. BID3 is an optimisation model which minimises the system cost in a year subject to constraints. It models all 8760 hours of the year and accounts for varying renewables, demand-side management, hydro and pumped/battery storage.

BID3 considers key plant costs and technical parameters, such as:

- incremental cost of production;
- start-up costs;
- no-load costs;
- Minimum Stable Generation; and
- minimum on- and off-times.

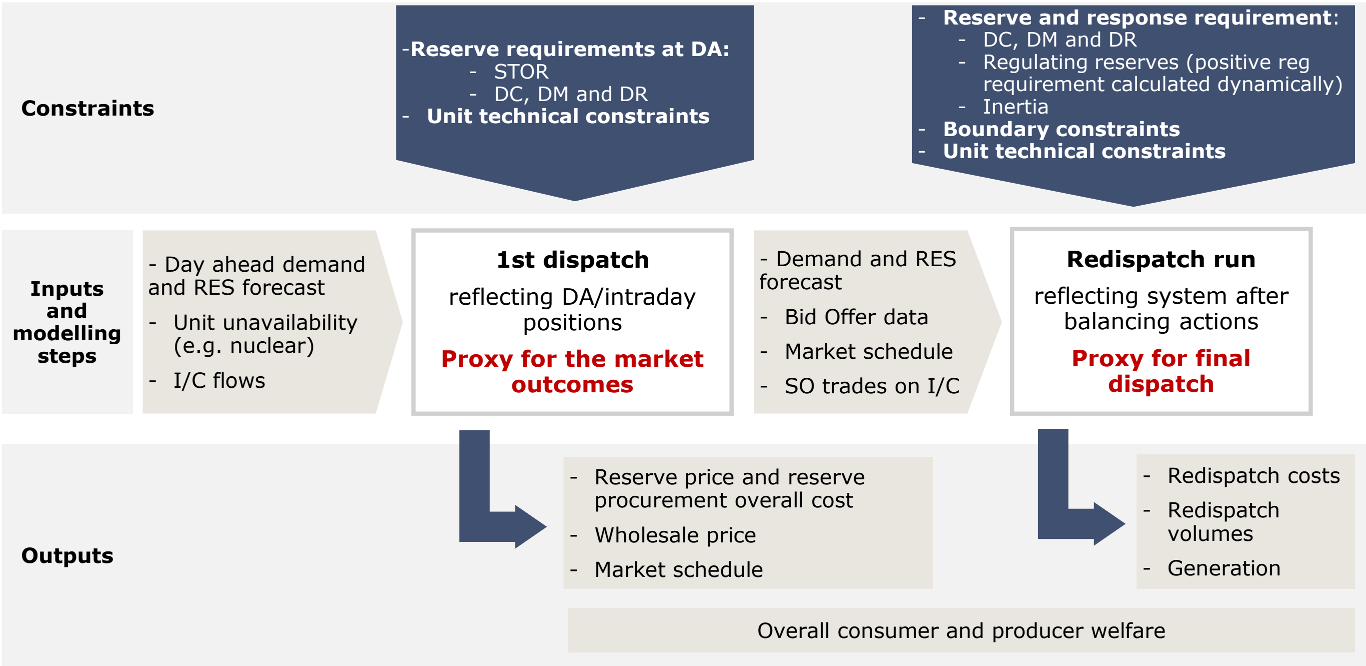
BID3 has been specifically designed to address the intermittency of renewables, to capture reserve requirements and other system requirements (e.g. inertia), and to reproduce redispatch actions taken in the Balancing Mechanism.

A.3.3 Modelling flow

The model is run successively to represent the 'market schedule' (proxy for market positions) first and the resulting dispatch accounting for the various system constraints next (proxy for final dispatch through the BM).

Exhibit A.16 shows the sequence of model runs, the relevant inputs and constraints, and key outputs.

Exhibit A.16 – Overview of the modelling flow and outputs



Notes: the specific constraints and inputs vary under the different modelled cases

A.3.4 Modelled cases

Four cases are modelled to show the potential incremental benefit of alternative scheduling arrangements. The table below details the assumptions and objective of each modelled case.

Cases		Description and objective
Status-quo		This case reproduces the scheduling and dispatch decisions taken on the key study day It serves as : – a backcast exercise to calibrate the model – a counterfactual against which to compare results of alternative arrangements – indication of potential inefficiencies in decisions made on the actual day
Alternative cases changes to the incentives or earlier control by ESO	**Early procurement of positive regulating reserve**	A case where some of the procurement of positive regulating reserve is done at (or close to) the DA. This can be viewed as either the introduction of 'balancing reserve' product, or co-procurement of regulating reserve at day-ahead. 1,000MW of positive regulating reserve is held early.
	Early management of boundary constraints	A case where the market schedule reflects boundary constraints (zonal day ahead market).
	Early management of boundary constraints + early procurement of positive regulating reserve	Combine the two cases above: – 1,000MW of positive regulating reserve is held early, – market schedule reflects boundary constraints

A.4 Results

A.4.1 Model calibration against actuals

AFRY's BID3 model has been calibrated in order for the status-quo modelled run to produce similar outcome as the actual days, in order to serve as counterfactual. The exhibit below presents the accuracy of modelled results compared to actuals, in terms of overall balancing costs.

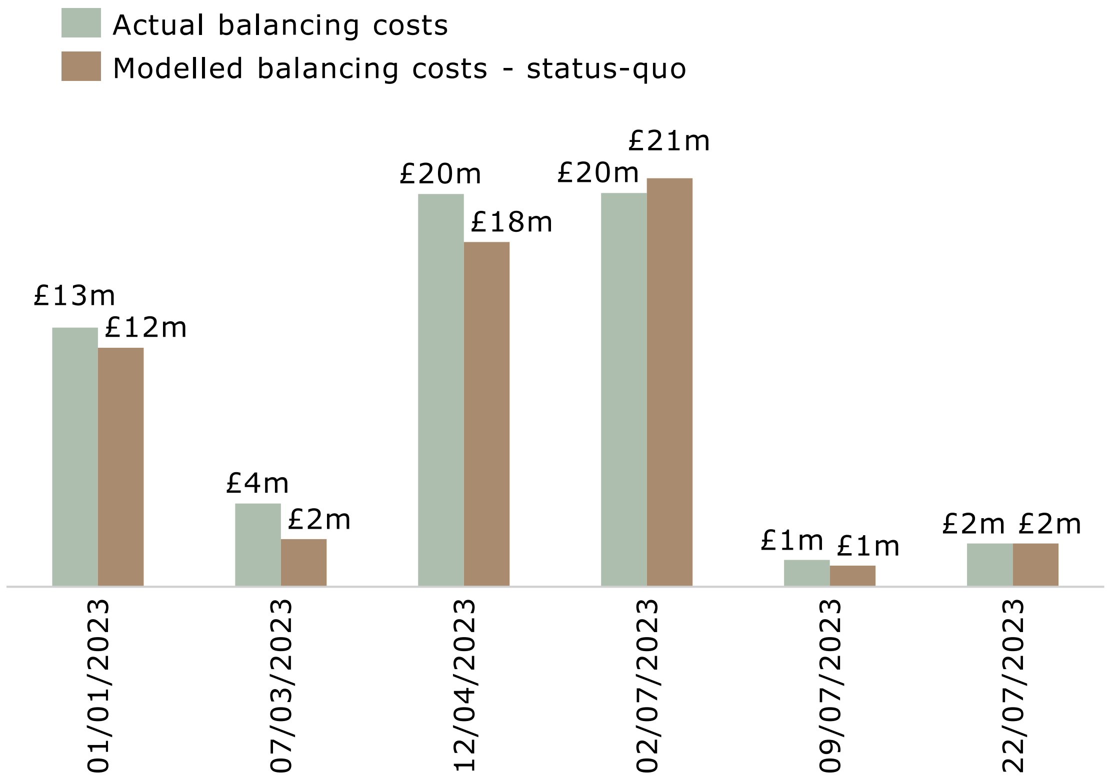
Differences between the modelled results and actuals will predominantly be a result of:

- information that the model cannot capture, such as localised voltage needs; and
- perfect foresight when scheduling storage.

Some of the differences may be highlighting some inefficiencies with the 'status quo'. However, the below balancing cost differences between actuals and modelled should, by no means, be considered as conclusive evidence that ESO actions were inefficient on these days.

One day does stand out though – on 7 March 2023, the model chooses to avoid expensive offers for the evening peak for redispatch, and tends to rely on smaller units to meet the energy and operating reserve needs. This leads generally lower balancing costs in the model for this day, and does suggest there may be scope for having more access to smaller units for balancing purposes[25].

Exhibit A.17 – Daily balancing costs: modelled status-quo versus actuals



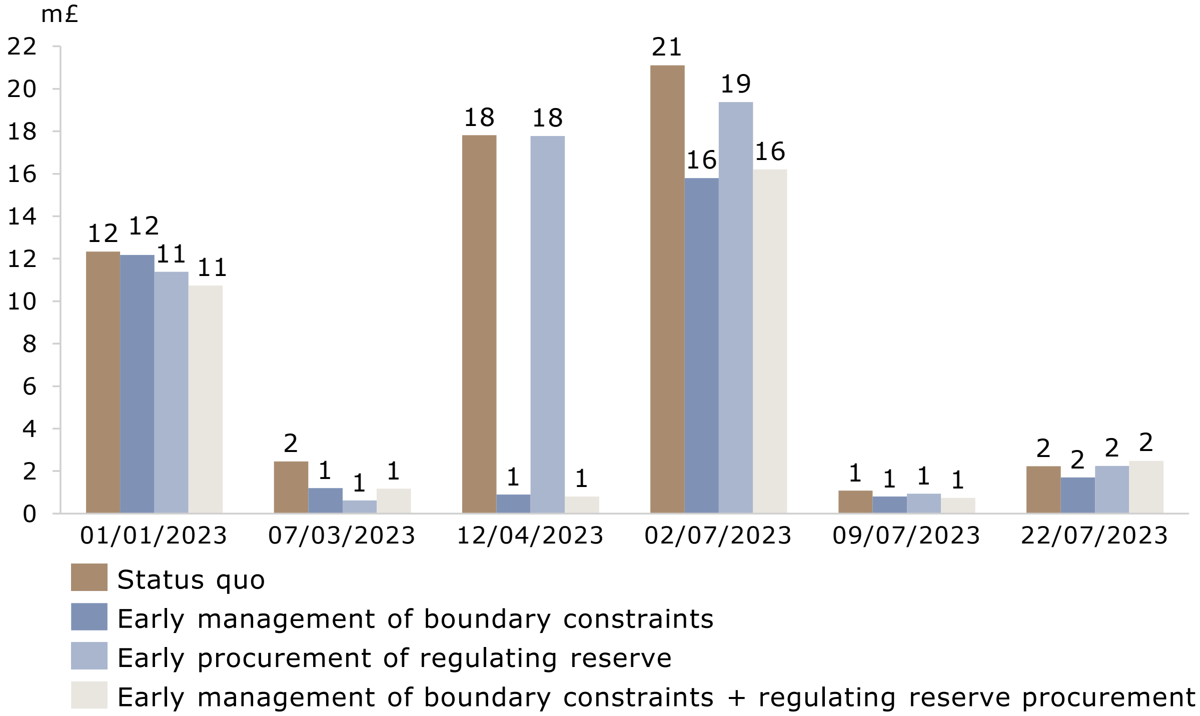
A.4.2 Change in balancing cost

In the three alternative cases, more constraints are 'added' to the market schedule (either through holding more reserve or reflecting boundary transmission constraints). Because more of the system constraints and needs are respected in the market schedule, less redispatch actions are required to obtain the final dispatch position. As expected, the resulting balancing costs in the alternative cases are lower than in the status quo, in particular on days with high balancing costs overall.

Exhibit A.18 presents the balancing costs in the four modelled case.

[25] ESO's Open Balancing Platform, introduced at the end of 2023, has been developed to facilitate bulk dispatch of small BMU assets. We can expect this development to allow to rely more on smaller BM units for balancing purposes.

Exhibit A.18 – Daily balancing costs in the status-quo and assuming alternative arrangements



The largest reduction in balancing costs can be seen on days with significant transmission constraints (12/04/2023 and 02/07/2023).

The early holding of positive regulating reserve led to the largest reduction in balancing costs on days when the procurement of regulating reserve was the main source of balancing costs (01/01/2023 and 07/03/2023).

A.4.3 Socio-economic welfare impact

In addition to the reduction in balancing costs in the alternative cases, other changes in system costs and market prices drive the overall distribution of consumer and producer surplus. This section present the impact on consumer and producer welfare in the alternative cases against the status-quo.

Key outputs are extracted from each run in order to calculate the difference against the status-quo:

- **Wholesale price impact**: we mode the impact on ex-ante market prices from different incentives compared to the status quo. In the case with early management of boundary constraints, the consumer price is the demand weighted average zonal prices, while producers are settled against the respective zonal price.
- **Change in balancing costs**: transfer between consumers and producers
- **Change in CfD payments**: higher or lower wholesale prices leads to a change in opposite direction of the CfD payments to generation under support. Additionally, a change in the final metered generation under support drive a change in the amount of CfD payments.

— **Additional reserve costs/revenues**: in the cases with early procurement or reserve, the additional cost to consumers (revenues to producers) to hold the reserve early. The payment is equal to the shadow prices of the regulating reserve constraint.

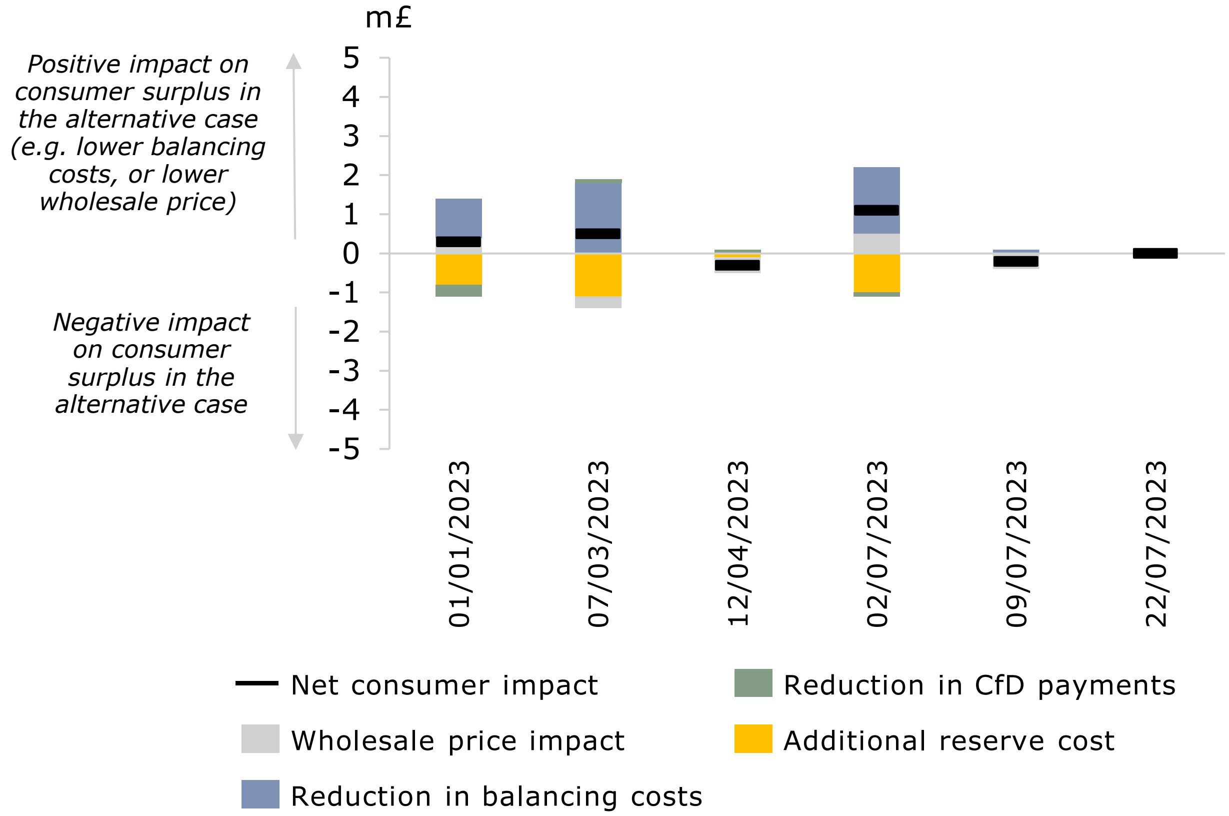
— **Change in generation costs**: difference in the final dispatch leads to difference in the overall generation costs (fuel and carbon costs).

— **Cross zonal congestion rent:** in the case early management of boundary constraints via introduction of market zones, congestion rent between the zones is assumed to be captured by the SO and passed on to consumers. In practice, with the introduction of zones, it can be assumed that locational TNUoS charges would be modified, which would offset some of the welfare transfer presented for in this section

A.4.3.1 Early procurement of positive regulating reserve

The chart below shows the difference in consumer surplus assuming earlier procurement of positive regulating reserve compared to the status quo.

Exhibit A.19 – Change in consumer surplus considering early procurement of positive regulating reserve (delta compared to status-quo)



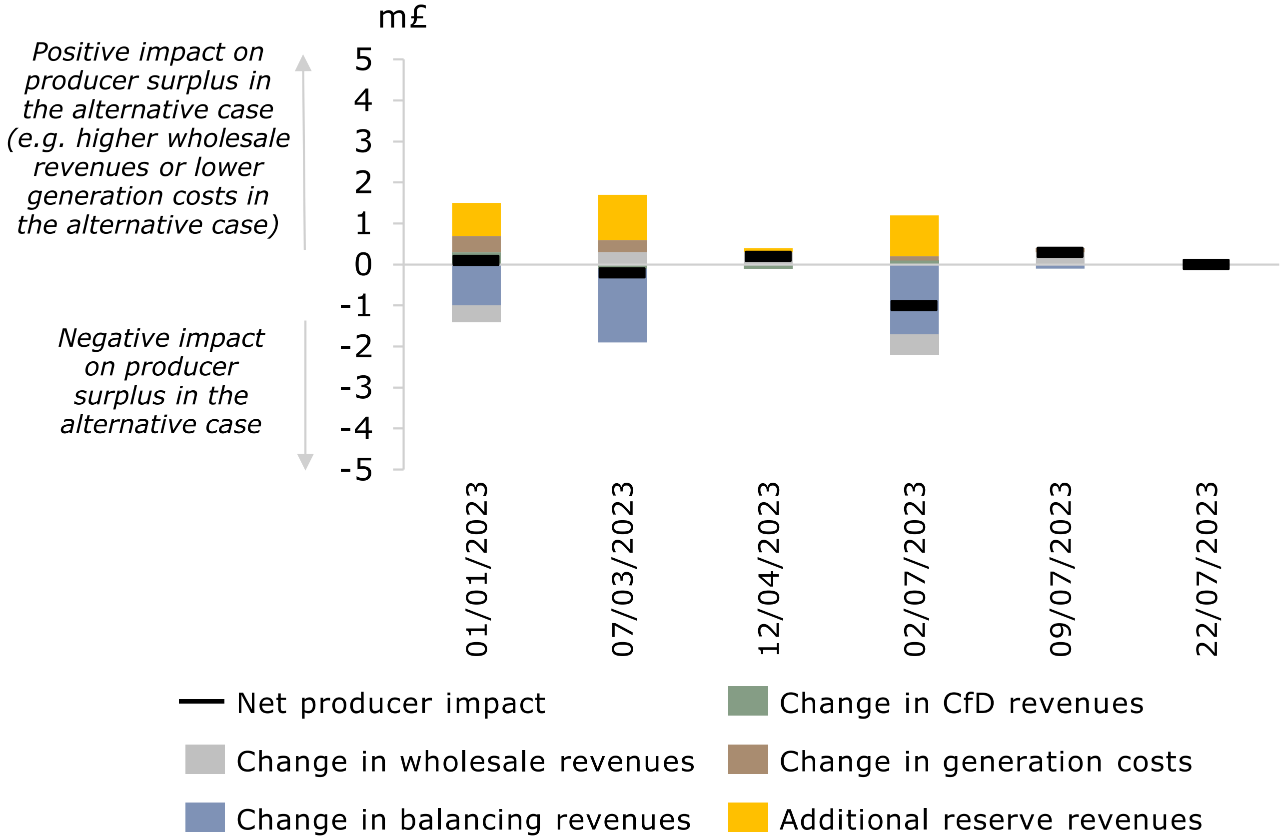
The changes in consumer surplus for each day are as follows:

— **01/01/2023**:operating reserve held earlier leads to lower balancing costs, offsetting the additional cost to procure reserve compared to the status quo. The net consumer surplus impact for this day is positive.

- **07/03/2023**: holding reserve early on this day helps avoid ~£2m of balancing costs. Despite the additional cost of procuring this reserve, the net impact for consumers is positive for this day.
- **12/04/2023**: holding regulating reserve early on this day leads to marginally higher wholesale prices, and slightly lower balancing costs. The net impact on consumers is small (slightly negative).
- **02/07/2023**: holding regulating reserve early on this day helps to avoid ~£1.7m of balancing costs. It also leads to lower wholesale prices on this day with high renewable generation: the units synchronized for the provision of reserve 'shift' the merit order, leading to slightly less expensive marginal generation (the value for these generators comes from the reserve cost and not the wholesale price). The net impact for consumers is positive by ~£1m.
- **09/07/2023**: holding regulating reserve early on this day leads to marginally higher wholesale prices, and slightly lower balancing costs. The net impact on consumers is small (slightly negative).
- **22/07/2023**: on this day with average thermal generation, holding regulating reserve early has minimal impact on wholesale prices and balancing costs.

The exhibit below presents the difference in producer surplus assuming earlier procurement of positive regulating reserve compared to the status quo.

Exhibit A.20 – Change in producer surplus considering early procurement of positive regulating reserve (delta compared to status-quo)



The changes in producer surplus tend to follow the opposite direction as the change in consumer surplus, demonstrating a welfare transfer. Only on the 01/01/2023 is the welfare impact positive for both consumers and producers. Drivers for each study day are as follows:

- **01/01/2023**: reserve held earlier leads to higher wholesale prices and additional reserve revenues for producers. Generation costs are lower thanks to a more 'optimal' schedule reached early with the early procurement of reserve: only the required thermal units to provide reserve are synchronised instead of additional synchronisations in the status quo. Overall, the net welfare impact for producers is small (slightly positive) for this day.
- **07/03/2023**: the reduction in BM revenues for generators is offset by lower generation costs and additional revenues for provision of regulating reserve. The net impact for producers is small (slightly negative).
- **12/04/2023**: holding regulating reserve early on this day leads to marginally higher wholesale prices. The net impact on producers is small (slightly positive).
- **02/07/2023**: the additional revenues for the provision of regulating reserves are lower than the reduction in balancing revenues combined with the reduction in wholesale price; it leads to a net negative welfare impact for producers on this day.

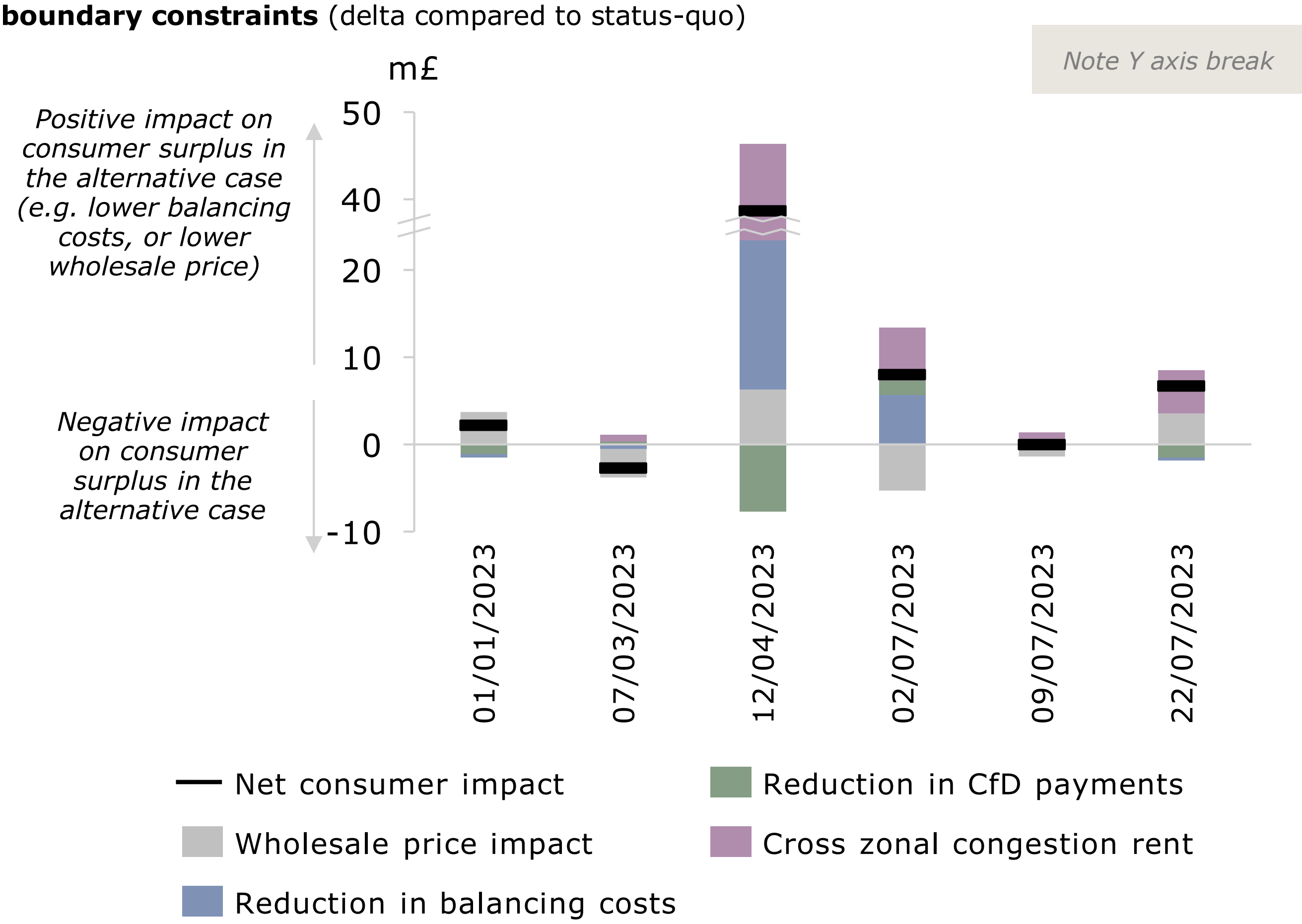
- **09/07/2023**: holding regulating reserve early on this day leads to marginally higher wholesale prices. The net impact on producers is small (slightly positive).
- **22/07/2023**: minimal change in wholesale prices and balancing costs compared to status quo.

It is worth noting that, on some days, the earlier procurement of positive regulating reserve leads to positive socio-economic impacts for both producers and consumers. For instance, on the 01/01/2023, the increased dispatch efficiency (that can be seen through the change in generation cost) leads to a positive socio-economic impact overall and for consumer and producer separately as well.

A.4.3.2 Early management of boundary constraints

The chart below presents the difference in consumer surplus assuming that the market schedule reflects boundary constraints (zonal market) compared to status quo. The model considers the main boundary constraints. In practice, it means that 11 zones are reflected in the model.

Exhibit A.21 – Change in consumer surplus considering early management of boundary constraints (delta compared to status-quo)



Note: "Cross zonal congestion rent" is, in a zonal market, the congestion rent between the zones captured by the system operator and transferred to consumers (e.g. through a reduction is network charges)

Early management boundary constraints via a zonal market has a very large impact on balancing costs and overall welfare transfer on days with network congestion, like the 12/04/2023 and the 02/07/2023:

- **12/04/2023**: market schedule respecting boundary constraints leads to a significant reduction in balancing costs on this day. The notable impact on wholesale prices is the significant decrease in wholesale prices in zones 'behind' the constraints. The CfD payments to wind under support are higher in the alternative case due to lower wholesale prices in the zones with high RES generation. Consumer surplus is further increased by the cross zonal congestion rent: in the case early management of boundary constraints via introduction of market zones. It is assumed the congestion rent between the zones captured by the system operator will be transferred to consumers (e.g. via lower BSUoS costs)[26]. The net consumer surplus impact is very positive on this day, reaching ~£38m.
- **02/07/2023:** market schedule respecting boundary constraints leads to a reduction in balancing costs on this day. However, wholesale prices are higher, driven by higher prices in zones in front of constraints. The net consumer surplus impact for this day is positive.

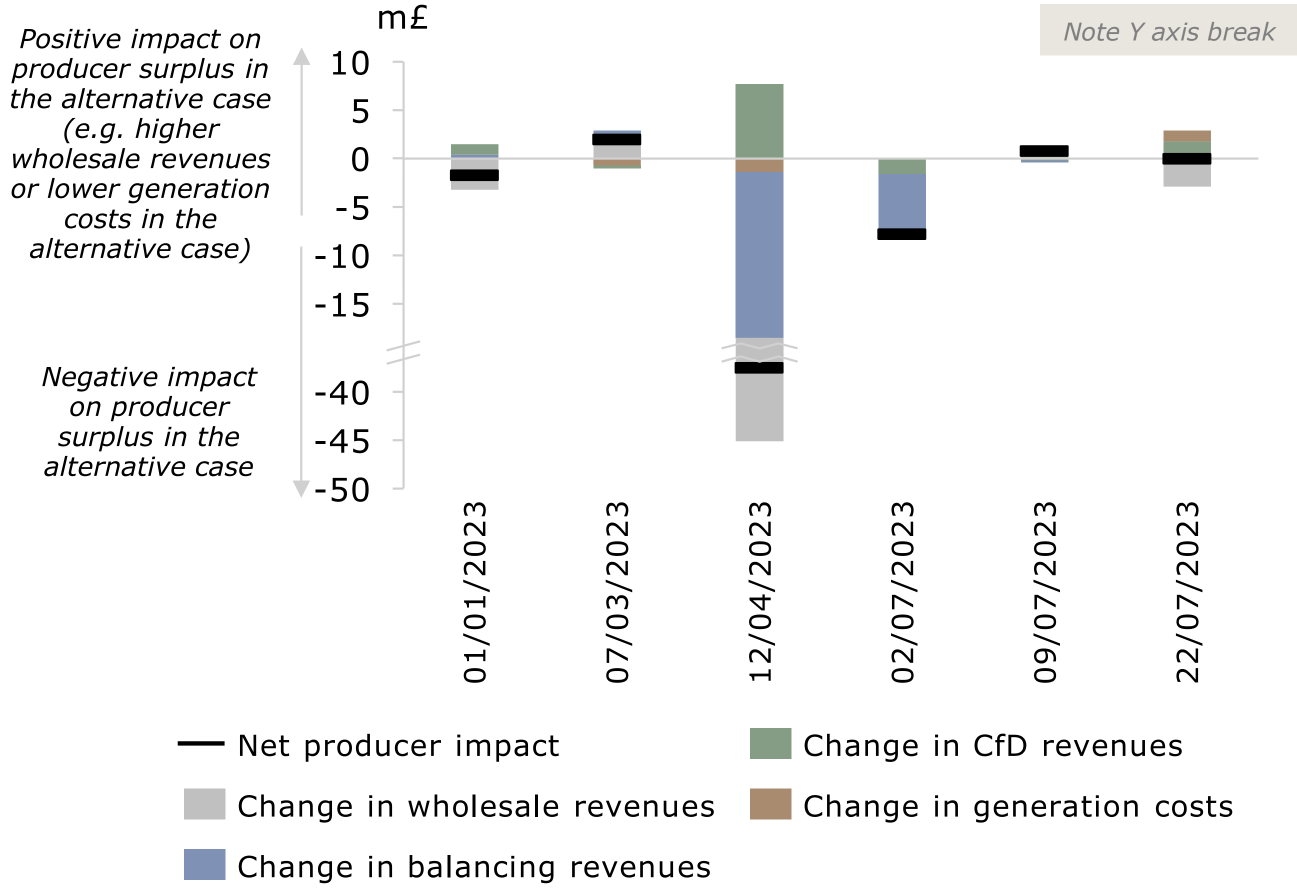
On 01/01/2023, 07/03/2023 and 09/07/2023 no major transmission constraint were binding. However, the market model shows changes in the wholesale price compared to the status quo. This is mainly due to the model optimisation horizon of three days . With the introduction of transmission constraints, the optimal market schedule is different than in the status quo case leading to different generation patterns, including for units with temporal constraints, such as pumped storage. For instance, on the 01/01/2023, pumped storage generation is higher in the case with boundary constraints compared to status quo case, as the PS units are not be able to generate as much in following days because they are located behind a constraint. This change in generation (due to constraints on following days and not the study day itself) leads to lower wholesale prices compared to the status quo case, and positive consumer surplus impact.

On 22/07/2023, ESO executed trades on the interconnectors to solve South East import constraints. These constraints are reflected in the zonal market results (before considering ESO interconnector trades), and therefore lead to a positive cross zonal congestion rent for consumers.

The chart below presents the difference in producer surplus assuming that the market schedule reflects boundary constraints (zonal market) compared to status quo.

[26] In practice, with the introduction of market zones, it can be assumed that locational TNUoS charges would be removed, which would offset some of the welfare transfer between consumers and producers presented in this case.

Exhibit A.22 – Change in producer surplus considering early management of boundary constraints (delta compared to status-quo)

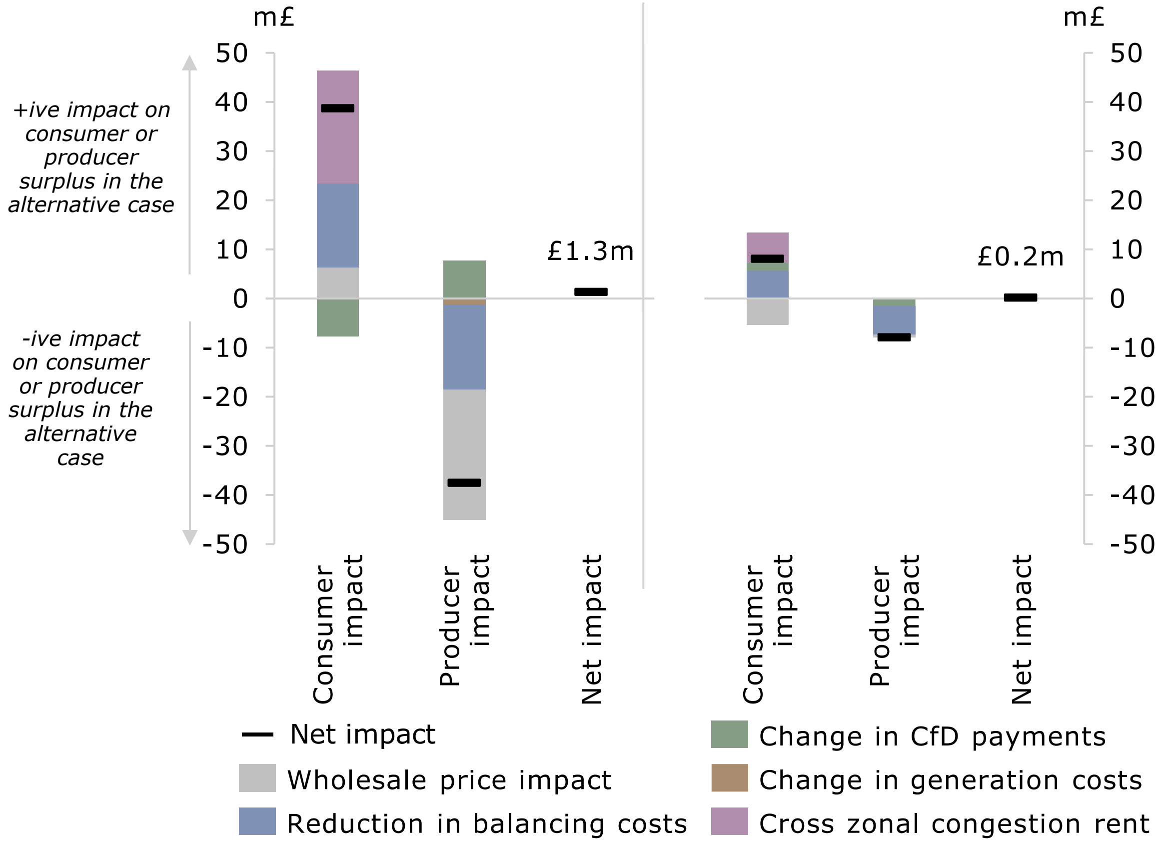


Drivers for the days with the largest change in balancing costs and overall welfare transfer are the following:

- **12/04/2023**: market schedule respecting boundary constraints leads to a significant reduction in balancing revenues. Lower market prices in the zones located behind the constraints lead to a significant decrease in revenues to generators located in these zones. This is particularly the case for inflexible generation such a nuclear located in the north of England and in Scotland that becomes exposed to zero or negative prices. CfD payments are higher given the decrease in market prices in the zones with high renewable generation. The resulting net producer surplus impact is very negative for this day, around -£37m.
- **02/07/2023:** the change in wholesale revenues are minimal, given prices are generally low in all zones. Most of the impact comes from a reduction in balancing revenues, leading to a negative producer surplus impact for this day

When considering the two days when significant transmission constraints were binding (12 April and 2 July 2023), it can be noted that the early management of boundary constraints leads to an improvement in the overall dispatch efficiency. This is reflected through the lower generation costs, and through the overall positive socio-economic welfare impact as presented in the exhibit below.

Exhibit A.23 – Change in system costs and welfare distribution with earlier management of boundary constraints
(delta compared to status-quo)

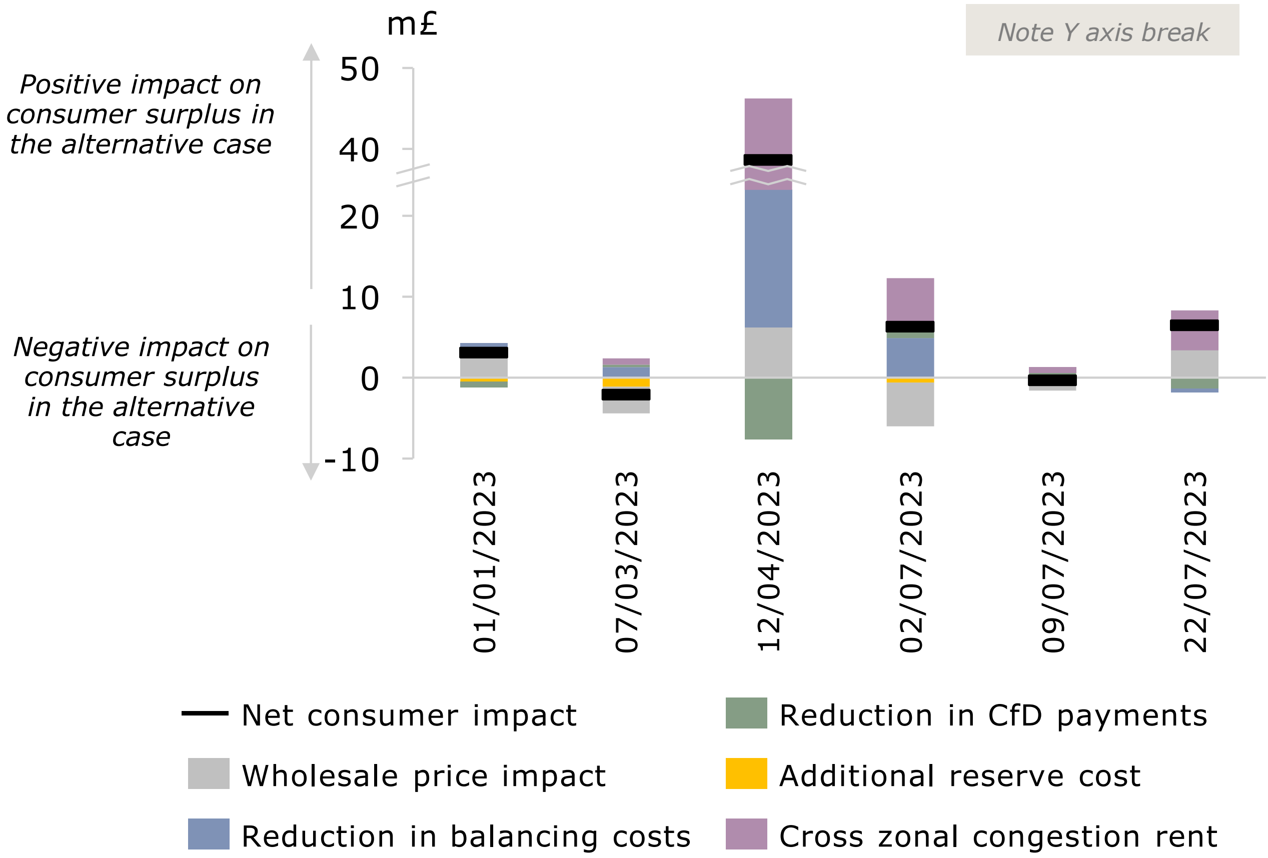


A.4.3.3 Early procurement of positive regulating reserve + management of boundary constraints

This modelled case combines the assumptions of the two cases presented above, i.e. 1,000MW of positive regulating reserve is held early and market schedule reflects boundary constraints.

The chart below presents the difference in consumer surplus for this modelled case compared to status quo.

Exhibit A.24 – Change in consumer surplus considering early procurement of positive regulating reserve + management of boundary constraints
(delta compared to status-quo)



The drivers for this modelled case are similar to the drivers for the respective individual two cases presented above. It can be noted the overall welfare transfers are larger on days with transmission constraints.

The chart below presents the difference in producer surplus for this modelled case compared to status quo.

Exhibit A.25 – Change in producer surplus considering early procurement of positive regulating reserve + management of boundary constraints (delta compared to status-quo)

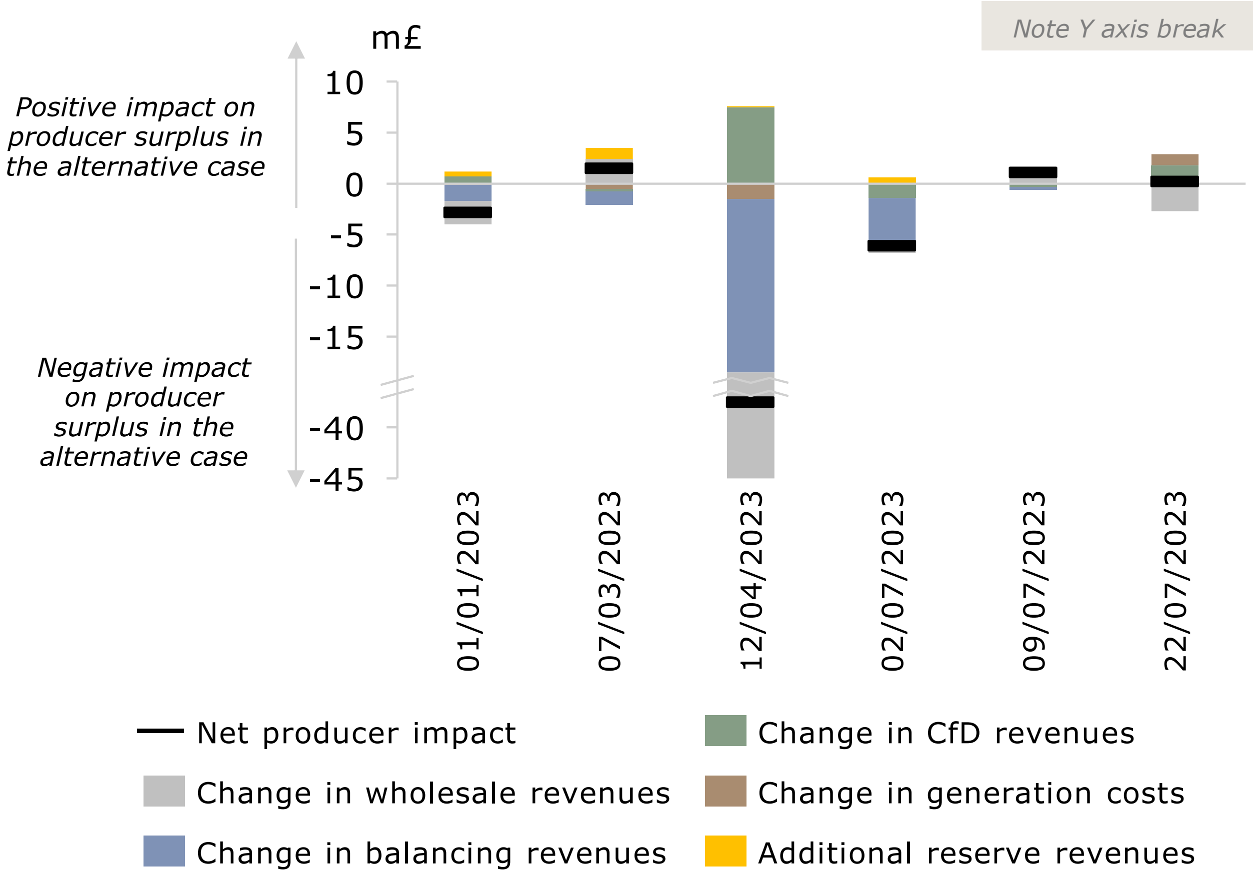


TABLE OF EXHIBITS

Quality control

Roles	Name	Date
Authors		
Approver		

Document control

Report Identifier: 2023/0120017/A

Version	Date	Comments

ÅF and Pöyry have come together as AFRY. We don't care much about making history.

We care about making future.

AFRY Management Consulting Limited
King Charles House
Park End Street
OX1 1JD
Oxford
United Kingdom
[Email]

Tel: +44 1865 722 660
afry.com

E-mail: consulting.energy.uk@afry.com