Report for ESO

Investigation into sequential and co-optimised procurement of energy, reserve and response services

Quantitative assessment

FTI CONSULTING nationalgridESO

23 July 2024

Disclaimer

This report has been prepared by FTI Consulting LLP ("FTI") for National Grid ESO as part of a Research and Development Collaboration Agreement entitled "Exploring the economic benefits of co-optimising procurement of energy, response and reserve" under the terms of the Agreement dated 05 October 2023 (the "Contract").

This report has been prepared solely for the benefit of National Grid ESO in connection with this project and relevant results can be disseminated through the industry (as defined under the relevant terms of the Contract), but no other party is entitled to rely on it for any purpose whatsoever.

This report is not to be referred to or quoted, in whole or in part, in any registration statement, prospectus, public filing, loan agreement, or other agreement or any other document, or used in any legal, arbitral or regulatory proceedings without the prior written approval of FTI. FTI accepts no liability or duty of care to any person (except to National Grid ESO under the relevant terms of the Contract) for the content of the report. Accordingly, FTI disclaims all responsibility for the consequences of any person acting or refraining to act in reliance on the report or for any decisions made or not made which are based upon such report.

The report contains information obtained or derived from a variety of sources. FTI does not accept any responsibility for verifying or establishing the reliability of those sources or verifying the information so provided.

Nothing in this material constitutes investment, legal, accounting or tax advice, or a representation that any investment or strategy is suitable or appropriate to the recipient's individual circumstances, or otherwise constitutes a personal recommendation.

No representation or warranty of any kind (whether express or implied) is given by FTI to any person as to the accuracy, completeness, reliability or adequacy of the report or of the information provided therein.

The report is based on information available to FTI at the time of writing of the report and does not take into account any new information which becomes known to us after the date of the report. We accept no responsibility for updating the report or informing any recipient of the report of any such new information.

Executive summary

FTI CONSULTING

Executive summary: Context, objectives and approach of the study

Context and objectives

- Wholesale (“WS”) energy markets and ancillary services (“AS”) markets in GB currently **operate independently (or ‘sequentially’)** of each other.
- However, there are **interdependencies between energy and AS markets** (e.g. participation in different markets is often mutually exclusive), so parties must decide which markets they participate in, based on (imperfectly informed) expected outcomes in subsequent markets.
- As such, sequentially designed markets have the potential to result in inefficiently allocated resources across WS/AS markets, leading to a knock-on effect on WS/AS prices.
- In this context, ESO wishes to investigate the merits of **co-optimising the scheduling of WS energy** and the **procurement of AS**, i.e. where a single optimisation process identifies the optimal allocation of resources across both WS and AS markets simultaneously.
- FTI has been commissioned to quantitatively evaluate the merits of co-optimisation of energy and AS, compared to a sequential design. This analysis complements a qualitative assessment of the merits of co-optimisation published alongside this report.

Approach and modelling setup

We deploy **our unit commitment dispatch model** for GB to examine the **merits of day-ahead co-optimisation of WS and AS market, relative to a sequential counterfactual.**

A. The **counterfactual (sequential) scenario** features four (simultaneously cleared) day-ahead AS markets: two markets for response (upwards/downwards), and two markets for reserve (upwards/downwards). These four markets are assumed to run ahead of (i.e. sequentially relative to) the day-ahead WS market.

B. The **co-optimised scenario** features a combined single optimisation process across all five markets (four AS markets and WS energy market) at the day-ahead stage.
   - We compare the two designs both historically (2022 as an illustrative year), and for selected future years (2025, 2030 and 2035).
   - For the forward-looking assessment, the model has been calibrated to the **FES 2022 Leading the Way (“LtW”) scenario.**
   - **Demand** for the **AS products** has been based on **ESO inputs.**

C. We perform this assessment under the current **national WS price design**, and we also test how co-optimisation would perform under a potential future **zonal WS price design** (featuring, illustratively, 7 zones and intra-GB transmission capacity based on ESO’s Holistic Network Design assumptions).

Using the methodology set out above, we quantify the **impact of co-optimisation** in terms of a **more efficient allocation of resources** between WS and AS markets, and its impact on **price formation**, system-wide **efficiency gains** and **consumer savings**.

Market design modelling overview

A — **Counterfactual scenario (“sequential”)**

B — **Co-optimised scenario (national design)**

C — **Co-optimised scenario (zonal design)**

For modelling tractability, the counterfactual represents a slightly simplified version of current GB arrangements

Impact of introducing co-optimisation of energy and AS

IMPACT MODELLED (historical and forward-looking)

Impact of introducing zonal pricing

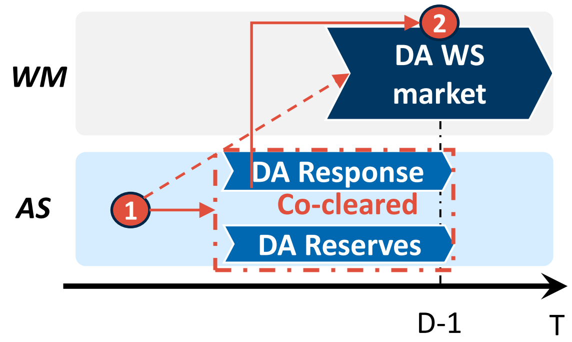
IMPACT MODELLED (forward-looking only)

Executive summary: Sequential and co-optimised designs have very different market dynamics which determines how efficiently resources are allocated

Sequential DA AS and WS markets

- In a sequential design, day-ahead ("DA") AS markets clear before the DA WS market, so the bidding strategy in the AS markets reflects participants' **expectation of the opportunity cost of being reserved in AS** markets (as successful participants in DA AS markets forgo the opportunity to bid in the subsequent DA WS market). This expectation, in turn, depends on **resources' forecast of the next day's DA WS energy prices and production schedules** (see 1 in the top right figure). The need to form such expectations leads to two inefficiencies:
  - **Imperfect WS price forecasting by individual resources** can lead to a **misallocation of resources** between energy and AS markets (e.g. unnecessarily costly mix of resources meeting the energy and AS demand).
  - There are difficulties in **reflecting the relevant opportunity cost in AS bids** (e.g. start-up costs for a thermal unit or arbitrage opportunities for storage across multiple hours are challenging to include in AS bids), which again can lead to a **misallocation of resources** between WS and AS markets.
- In the WS market clearing reservations for AS provision are respected (see 2 in top right figure), even though they may, ex post, turn out not to be the most efficient way of meeting demand for energy and AS.

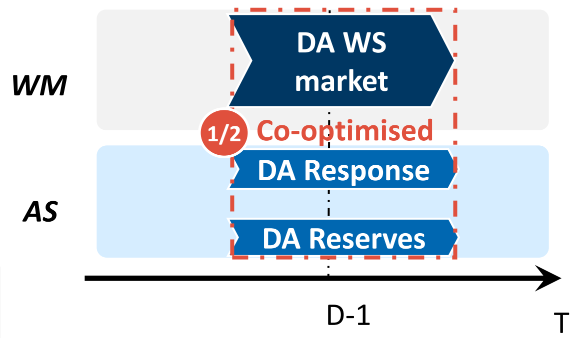
Sequential national DA market



Co-optimised DA AS and WS markets

- **With co-optimisation,** the clearing of the **DA AS market and DA WS market occurs simultaneously**[1] (i.e. the markets are "co-optimised") (see 1/2 in the bottom right figure).
- This means that **explicit AS bids no longer** need to be formulated for the AS markets and **market participants no longer need to form expectations of the DA WS market price**. As such, the two core sources of potential inefficiencies under the sequential design (i.e. the imperfect WS price forecasting and the challenges in incorporating all relevant opportunity costs in the AS bids) are removed.

Co-optimised national DA market



This report quantifies how any **misallocation of resources** between WS energy and AS markets, under a **sequential design relative to co-optimisation**, would lead to **higher-than-optimal WS scheduling costs**, and how it would impact **AS market price**, **WS market prices** and **overall consumer costs**.

We also show how co-optimisation enables both: (1) **DA AS reservation prices to reflect the opportunity costs of participating in the DA WS energy market**; and (2) **DA WS energy prices to reflect scarcity in DA AS markets**.

Note: (1) This may lead to a perception of price formation under co-optimisation being opaque, but stakeholder understanding can be improved through education and, over time, experience of outturn pricing outcomes, as explained in slide 44 of our qualitative report (published alongside this report).

Executive summary: Co-optimisation reduces system and consumer costs by improving scheduling of assets across energy and ancillary services

1	**Co-optimisation generates benefits in all years**	▪ Co-optimisation **consistently generates benefits for GB consumers and for the system** overall, across all modelled years. ▪ The **magnitude of benefits varies** depending on specific market circumstances.
2	**Energy market impacts dominate… …while AS markets impacts are second-order**	▪ The **main benefit of co-optimisation is in the wholesale energy market**: it helps identify a **less costly marginal unit** compared to a sequential design. ▪ Even a relatively modest price impact in the wholesale market has a large aggregate effect as it drives the price for the entire GB demand. ▪ The **benefits of co-optimisation for AS markets are an order of magnitude smaller** due to the size of the AS markets relative to energy. ▪ We estimate that average energy prices fall by 3-5% in most years[1], while AS prices fall by 25-90%, yet energy market represents the majority of total consumer cost savings.
3	**Consumer benefits exceed system cost savings**	▪ **Consumer cost savings** are the main benefit of co-optimisation in most years[2] because the re-allocation of resources between WS and AS markets, and associated reduction in energy prices, affect all generating units. ▪ **System cost savings are typically an order of magnitude smaller,** as the same re-allocation of resources mainly affects marginal units.
4	**Co-optimisation helps in both thermal- and battery-dominated systems**	▪ In a system with thermal generation (e.g. 2022), **co-optimisation mitigates "incorrect" reservations of thermal plants** for AS, which would otherwise inefficiently increase the energy market clearing prices. ▪ In a battery dominated system (e.g. 2030-35), **co-optimisation improves the coordination of battery participation** across AS and WS markets and, in turn, reduces the need to run thermal units in the WS markets.

Modelled impact of co-optimisation across WS and AS markets

	System cost savings	Consumer cost savings[1]
2022	£178m ↓	£1,423m ↓
2025[2]	£831m ↓	£60m ↓
2030	£189m ↓	£704m ↓
2035	£137m ↓	£884m ↓
Total PV (2025-2035)	£3.4bn ↓	£4.9bn ↓

These savings represent up to 6% of annual consumer costs

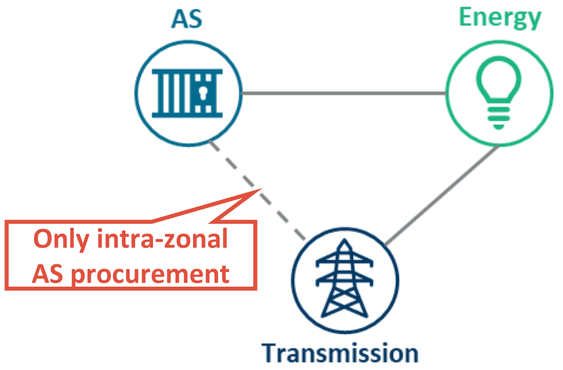
Notes: (1) We calculate consumer cost savings based on reductions in WS energy and AS prices. Accounting for offsetting changes in subsidy payments for generators on Contract for Difference schemes leads to a reduction in annual consumer cost savings of 7% in 2022, 11% in 2025, 63% in 2030 and 48% in 2035; (2) 2025 is an outlier year with very tight AS markets, which, unusually, drives slightly higher energy prices under co-optimisation relative to sequential design. This occurs due to two reasons: firstly, scarcity in AS market is being reflected in WS prices under co-optimisation but not under sequential markets. These price signals encourage entry into AS under co-optimisation, facilitating new investment through a 'dynamic efficiency' effect. Secondly, under sequential markets, generators make costly mistakes in forecasting uncertain WS prices and incur inefficient costs (e.g. start-up), but also self-cannibalise the resulting WS prices. This leads to economically inconsistent outcomes under the sequential design, and also leads to substantially higher system costs relative to co-optimisation. We do not consider the second effect to be representative of the typical outcomes for sequential vs co-optimised designs.

Executive summary: In the zonal variant of the co-optimised model, reserve and energy prices vary across zones, and response prices are also affected

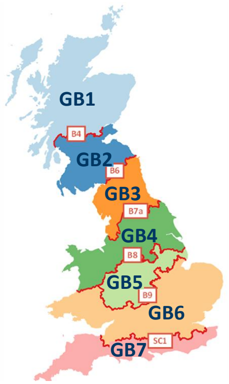
Methodology

- In the zonal model the WS market is split into **seven zones** (right chart below), originally derived for our work on locational design for Ofgem. The physical attributes of the services mean that **reserve markets are zonal** while **response remains procured nationally**.
- The zonal model uses the same demand and generation capacity inputs as the National model (FES 2022 LtW), allocated across zones.
- **Conservatively**, we assume that:
  - The **aggregate zonal requirements for reserves equal the national requirement for reserves**; and
  - Zonal reserve requirements need to be **fulfilled with assets from within the relevant zone.**

Zonal co-optimisation



Seven-zone model



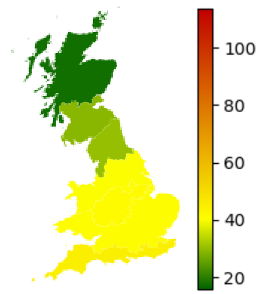
- **The objective of the zonal assessment differs** from the national assessment. We do not compare a **zonal sequential vs a zonal co-optimised market.** Instead, we compare **AS reservations and AS price dynamics** between a **zonal co-optimised** DA WS and AS markets relative to **national co-optimised** markets from 2025 to 2035.

Key findings

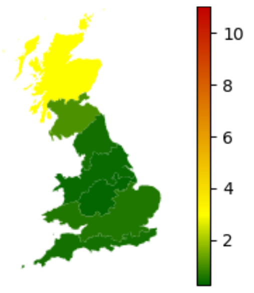
1. **Both Reserve (Up) and energy prices vary significantly by zone, reflecting** location-specific supply-demand balance. **Reserve (down)** prices are zero for the vast majority of the modelled hours (similar as under national pricing).

2. Typically, **energy prices are highest in the South**, driven by low wind capacity relative to demand (and transmission capacity). By contrast, **Reserve (Up) prices are highest in the North,** where high wind capacity increases reserve needs while the supply of reserve is limited.

Energy weighted average prices by zone (£/MWh, 2035)

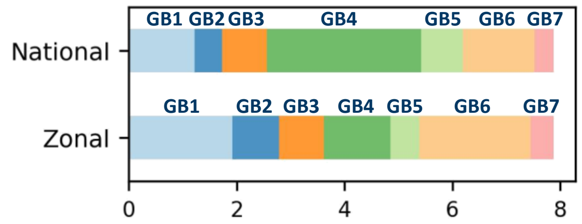


Reserve Up weighted average prices by zone (£/MW/h, 2035)



3. In the **national model** reserve is procured from the cheapest source across the country (e.g. GB4), while the **zonal model procures reserves where it is most needed** (e.g. GB6).

Reserve Up provision by zone (GW/h, 2035)



4. In the **zonal model**, Reserve Up is provided by a **different technology mix**: more thermal plants (13% of provision in 2035 vs 5% in National) and less batteries (55% in 2035 vs 60% in National) and pumped storage, partly due to the need to meet each zone's Reserve Up demand using intra-zonal sources.

5. **The zonal reserve procurement, in turn, impacts response prices** (even though the latter remain national in the zonal model). For example, lower battery provision of Reserve Up in a zonal market frees up battery capacity to provide Response Up, thus pushing down Response Up prices (by c.45-60%).

Executive summary: The modelling results are based on several key assumptions, many of which we consider to be very conservative

Conservative assumptions

Factors which could lead to merits of co-optimisation being underestimated

- **Product structure:** We only model four AS products that are jointly cleared in the Sequential design; this simplified market structure reduces information asymmetries.
- **Gaming or market power:** Only economic factors are considered; in practice co-optimisation could reduce gaming potential as fewer human decisions need to be made.
- **Modelling sequence:** AS reservation markets run just before DA WS energy market, with only a short time lag; a longer lag would increase WS price forecast errors under Sequential.
- **Zonal design**: We assume the same total reserve requirements as under a National market design, and do not model inter-zonal procurement (only intra-zonal). In practice total needs for AS would likely fall under a zonal design and inter-zonal procurement would reduce AS and WS prices.

General limitations

Could have positive or negative impact on merits of co-optimisation

- **Central scheduling:** As standard in power market modelling, we model a centrally scheduled wholesale energy market (and do not examine the merits of central relative to self-scheduling).
- **Modelled scenario:** We use the FES 2022 LtW scenario for capacity and demand forecasts, and the HND and NOA 2021/21 Refresh transmission network assumptions.
- **Demand for AS:** Need for AS products has been forecast based on ESO inputs.
- **Supply of AS:** Participation of individual technologies (e.g. demand-side response (DSR) and renewables (RES)) could differ from the assumed projections.

Overestimation

Factors which could lead to merits of co-optimisation being overestimated

- **Trading:** We do not consider the potential of intraday trading in reducing DA misallocations of resources under Sequential.[1]
- **Product structure:** We model one-hour AS products, though multi-hour products could facilitate reflecting opportunity costs under Sequential.
- **Changes in subsidy payments:** We do not account for any changes in subsidy payments, which apply to a subset of resources and may offset reductions in consumer cost savings (e.g. increases in Contract for Difference payments to generators when energy prices fall).[2]

This is a key area of interest for ESO (e.g. see recent ESO work here).
As a next step, ESO may wish to investigate the **impacts of self vs central scheduling on consumer costs**; and how this may **interact with the design of the Balancing Mechanism** and/or the wholesale market **locational granularity**.

Notes: (1) Intraday trading might lead to a reduced gap in system costs between sequential markets and co-optimisation but is not expected to impact consumer savings from co-optimisation; (2) However, co-optimisation causes the largest reduction in WS energy prices during peak hours (see slide 45), i.e. hours in which subsidised RES units are less likely to generate.

Table of contents

Report conventions

Unless stated otherwise:

- Prices, revenues and costs are expressed in GBP and in 2022 real terms. As the model is run in real EUR terms, we convert these values to GBP at an exchange rate of 1 GBP = 1.17 EUR.
- Modelled wholesale prices are presented as annual average load-weighted prices.
- All present values ("PVs") are discounted to 2024, at a rate of 3.5% as per the HM Treasury's Green Book (link)
- We have linearly interpolated between modelled years to derive values for intermediate years (e.g. 2027, 2028 etc.).
- All results are presented for calendar years beginning 1 January.

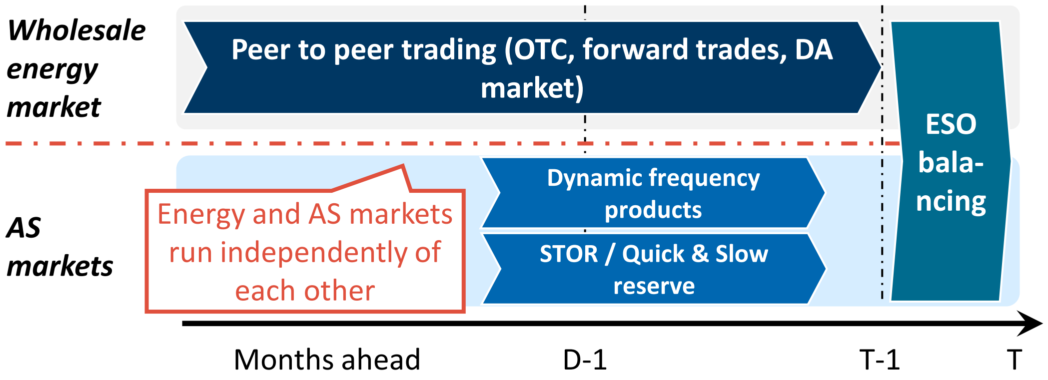
Section 1: Background and context

FTI CONSULTING

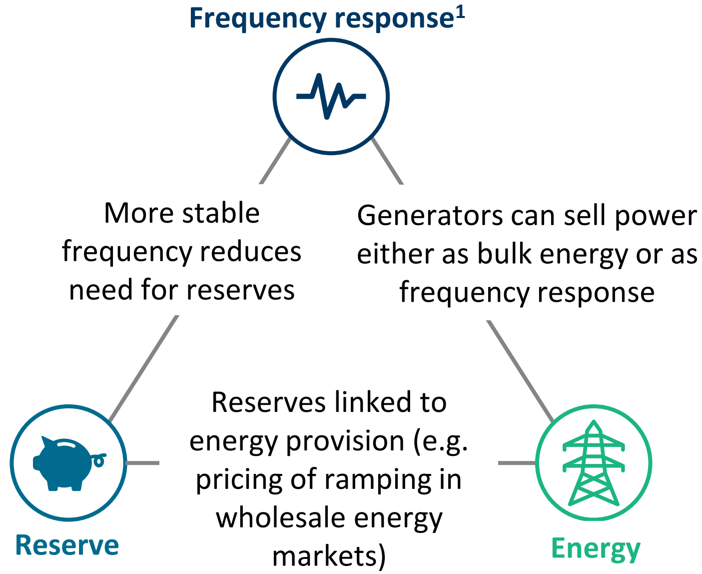
The current GB electricity market design features separately operating energy and ancillary services markets, despite inherent links between the two

WS and AS markets in GB currently operate independently of each other

Simplified current GB design – without co-optimisation



However, there are close interrelationships between energy, frequency response, and reserve markets



Markets operate independently and close sequentially, implying that:

1) Resources may need to **choose which markets to participate in**;
2) Resources' bids reflect the **estimated opportunity costs** of participating in the selected market (as by virtue of providing a service in one market assets are not able to provide another service); and
3) The **need to form such expectations** introduces **significant information imperfections**, with potential **adverse impacts on the price formation process** and resulting market outcomes.

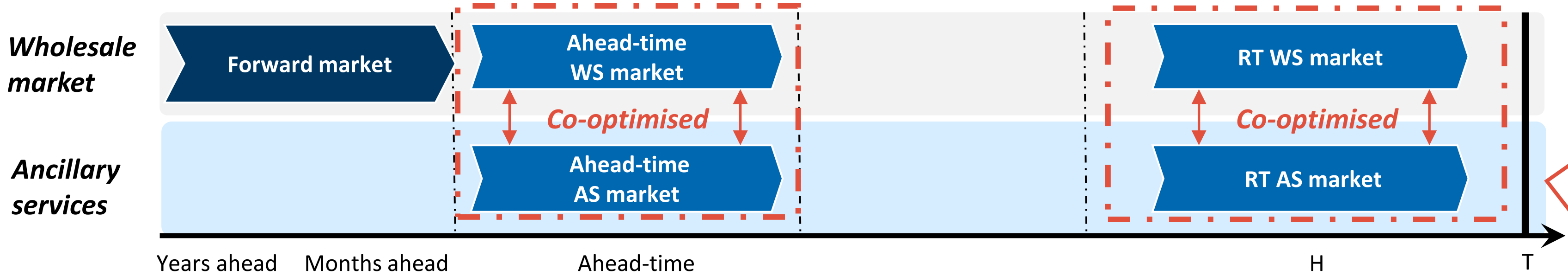
Market participants must first **forecast DA WS energy prices** and their resulting production schedule to understand their own **opportunity cost** of providing AS. Forecasts are complicated by **uncertainty** on the final energy demand, renewable production, possible outages etc., as well as the fact that the bidding of other market participants, and hence final market outcomes, will be dependent upon their opportunity cost estimations.

Notes: (1) We use "frequency response" or simply "response" interchangeably in this report.

ESO wishes to investigate whether there is a case for systematically considering these interactions when scheduling energy and procuring AS

- ESO wishes to investigate the potential merits of **co-optimising the scheduling of energy (via WS markets) and the procurement of AS.**[1] This is a process where a **single optimisation process** identifies the optimal outcome across **both WS and AS markets** simultaneously.
- The expectation is that this process would lead to, among other results, a **more efficient allocation of resources** based on the relative needs and conditions of each market and an **efficient formation of prices** across each market. This could be particularly important in future years as new AS markets are launched and AS requirements increase, increasing the potential efficiency gains of co-optimisation.

Simplified schematic of a hypothetical co-optimised market design (with central scheduling[2] and an ahead-time market[3])



This assessment explores the impact of WS and AS markets operating in an integrated, co-optimised manner.

We evaluate the impact of removing the need for resources to choose which markets to participate in and/or perform complex estimations of the **opportunity costs of participating in different markets**.

- In this report we focus on the potential co-optimisation of WS energy and AS markets. Alternative degrees of integration between markets are possible, which we discuss in more depth in the terminology section.
- Co-optimisation can be performed at forward timeframes and/or real time ("RT"). In this report, we **refer to the forward time frame as 'day-ahead'** because:
  - Currently the reference WS market timeframe in GB is the DA market;
  - Currently US markets co-optimise the WS and AS markets at DA stage.
- However, in the future, co-optimisation could also be performed at the **intraday timeframe** augmenting and/or replacing co-optimisation at the DA timeframe, as the importance of renewable generation forecasting overtakes the long-duration commitment decisions of thermal units.

Notes: (1) ESO has started examining these issues, at a high level, in its recent work on Net Zero Market Reform (link); (2) The merits of moving from GB's current self-scheduled market design to a centrally scheduled one are not examined in this report, for more discussion see slide 17; (3) Here we provide a simplified hypothetical market design, in practice there would be the possibility to introduce WS markets at multiple timeframes e.g., a day-ahead WS market preceding multiple intraday WS auctions.

ESO has commissioned FTI to undertake a qualitative and a quantitative assessment of the co-optimisation of energy and AS

- FTI has prepared two reports, published in parallel, setting out our approach, assessment and key findings regarding the sequential and co-optimised procurement of energy, reserve and response services.
- Our assessment comprises the following elements:

1. Qualitative assessment of the pros and cons of co-optimising energy and AS in GB;

This report focuses on the qualitative assessment. We describe, qualitatively, the benefits (e.g. **more efficient price signals** and **more efficient allocation of resources** across markets) as well as **potential challenges and risks** arising from a potential co-optimisation of energy and AS in the GB context.

2. Quantitative assessment of the historical cost savings of co-optimised procurement of energy and AS in GB;
3. Quantitative assessment of the potential efficiency savings from the co-optimised procurement of energy and AS in GB from 2025 to 2035;
4. Quantitative assessment of the potential efficiency savings from co-optimisation in a locational market design.

This report focuses on the **quantitative** assessment of potential impacts of co-optimisation of energy and AS (specifically reserve and response) in GB.

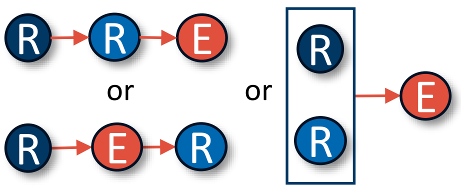
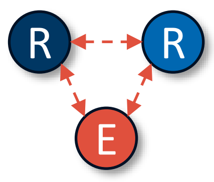
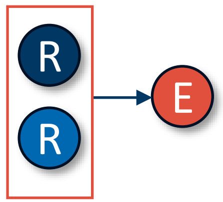
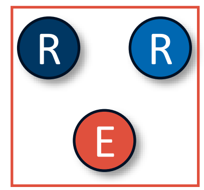
Section 2: Co-optimisation terminology and central dispatch

FTI CONSULTING

Co-optimisation, as a term, has multiple meanings, and this report refers to "full" co-optimisation of energy, response and reserves markets

- The value of co-optimisation is derived from efficient allocation of the potential services a single resource can provide.
- From a market design perspective, there are varying degrees to which a joint optimisation process can be used to identify the optimal outcome (e.g. minimum system costs) across energy and reserve and response.
- There is no uniformly accepted terminology, and the term "co-optimisation" can refer to different aspects of an optimisation process. For example, it can refer to a "manual" process where multiple factors are considered in decision-making (i.e. with human intervention process) or a completely automated process in which total system costs are minimised while respecting ancillary service requirements.
- Here we consider the required hourly volumes of reserve and response products to be predetermined. In other contexts, co-optimisation can also refer to holding different quantities of different response and reserve products (where they are at least partly substitutable) based on their relative cost. Here we assume that the volumes are predetermined.[1]
- For the purposes of this report, we define and use the following spectrum of co-optimisation options:

We set out, in the methodology section, how these designs have been reflected in our modelling approach.

	Separate ("sequential")	Implicit co-optimisation	Partly co-optimised	Fully[2] co-optimised
Key features	▪ Markets clear separately, typically at different times ▪ Interdependencies between markets reflected through estimated opportunity costs of participants	▪ Implicit consideration of multiple requirements ▪ Typically, with some degree of human decision making	▪ Automated (algorithm-based) optimisation of multiple requirements ▪ Only some requirements are considered (e.g. subset of AS)	▪ Automated (algorithm-based) optimisation of multiple requirements ▪ Considering all relevant requirement
Real-world example	▪ GB DA energy ("E") market versus Reserve and Response ("R&R")	▪ GB cost-minimisation (subject to SQSS) by ESO, post gate closure (Energy, Reserve and Redispatch)	▪ GB DA Enduring Auction Capability (EAC) auction: Reserve and Response	▪ International examples (but not in GB) ▪ Energy, Reserve and Response
Stylised representation				

Notes: (1) An example of the latter is the reduced need to procure fast response services when there is abundant inertia in the system and vice versa; (2) The term "fully" in our report refers to co-optimisation that includes currently feasible range of services – notably energy, frequency response and reserves. We recognise that co-optimisation could in theory also include voltage, inertia and other system services, but we do not include them in our report, as such co-optimisation would be breaking new ground from a market design perspective.

Central scheduling is expected to facilitate the introduction of full co-optimisation between energy and ancillary services

- The merits of moving from GB's current self-scheduled market design to a centrally scheduled one are not examined in this report. (This potential reform is being considered by the Department for Energy Security and Net Zero ("DESNZ") as part of its Review of Electricity Market Arrangements ("REMA")).[1]
- For the purposes of this report, we consider that a potential transition to a **centrally scheduled market design is likely to facilitate the introduction of the full co-optimisation of energy and AS markets** (response and reserves) for two reasons:

1

Bidding complexity

- Under self-scheduling, market parties that operate resources that can be active in both the energy and AS markets need to formulate bids for these markets separately.
- However, the bids need to be interlinked, i.e. for a particular market time unit, if a bid is accepted in the AS market, the volume of electricity that can be sold in the energy market is impacted (and vice versa).
- Providing explicit linkages between the separate energy and AS bids while having the bids reflecting the technical characteristics of the resources is extremely complex.
- Under central-scheduling, multi-part unit-based bids remove the need for such linked bids as the assets' technical capability and costs can be considered directly.

2

Improved SO information

- Under central scheduling, the system operator has access to significantly improved information regarding the technical characteristics of all resources.
- This information allows for a more efficient allocation of resources between energy scheduling and AS provision needs.
- As well, a wider option space for resources participating to AS can be expected as explicit participation is not required.

- In our modelling setup, we are implicitly assuming a centrally scheduled market design in place - although we do not express any views on its merits and implementation.[2] This is the convention due to the use of a unit commitment dispatch model to represent the power system, which is standard practice.
- International examples of co-optimisation are also commonly (but not always) accompanied by centralised markets and locational pricing.[3]
- A complementary technical assessment for ESO (published alongside this report) discusses in more detail how bidding languages in centralised and self-scheduled markets differ, and the consequent impact on co-optimisation.[4]

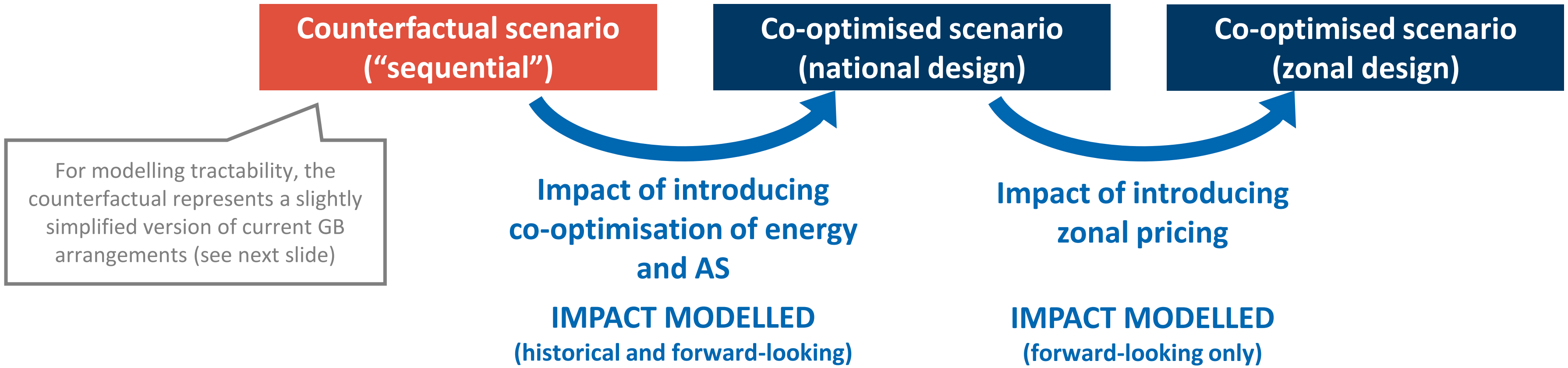
Notes: (1) In the second consultation of REMA, published on 12 March 2024, both self and central scheduling options are still being considered (link); (2) For example, impacts on procedures to schedule interconnectors and the organisation of intraday markets under central scheduling would need to be reviewed; (3) For further details, see slide 64 of our qualitative report (published alongside this report); (4) See also the explanatory note from N-SIDE about the SDAC MSD Co-optimization Roadmap Study from 20/10/2020 (link).

Section 3: Overview of the methodology

We assess full co-optimisation relative to a sequential design under historical and forward-looking horizons; and zonal and national wholesale designs

- We examine full co-optimisation, relative to a sequential counterfactual (see slide 16 for definitions), under two-time horizons:
  - **Historical** perspective, using **2022**[1] as an illustrative year; and
  - **Forward-looking** perspective, looking at **2025, 2030 and 2035**[2].
- We perform this assessment under the current **national** WS price design, and we also test how co-optimisation would perform relative to a sequential counterfactual under a future **zonal** WS electricity price design.
- Our assessment is summarised in the figure below.

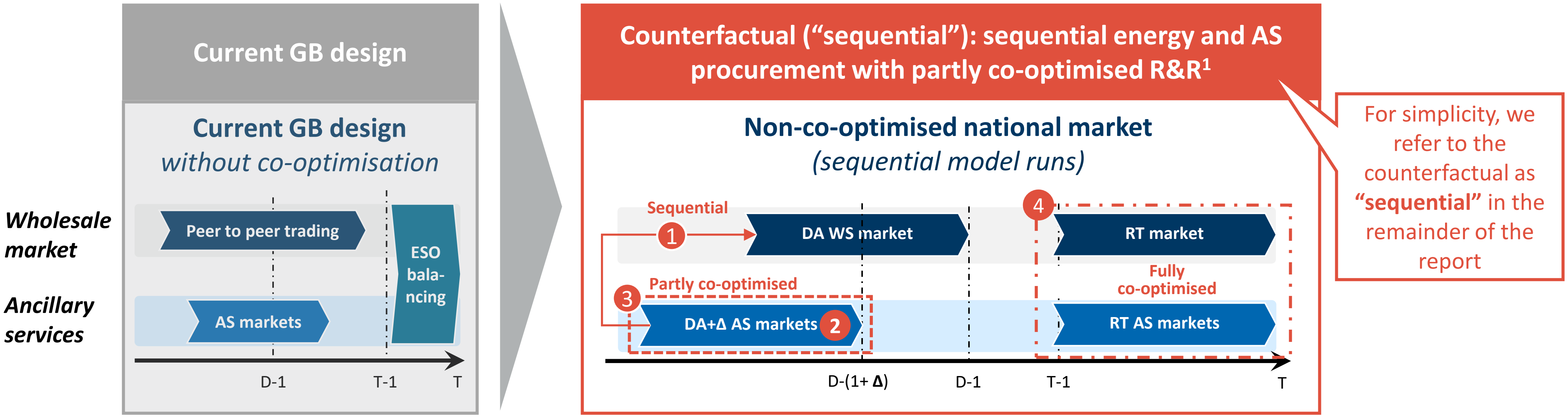
Market design modelling overview



- The remainder of Section 3 sets out:
  - The definition of the **counterfactual "sequential" scenario**, including how this differs from the current GB design;
  - The definition of the **co-optimised scenario** (under **national and zonal designs**), as applied in our modelling; and
  - The **quantitative metrics** we use to compare the actual and counterfactual scenarios.

Notes: (1) In selecting year 2022, we reflected a trade-off between modelling a recent year vs. a year with "average conditions"; (2) The modelling period was agreed with ESO and is not necessarily indicative of potential timeframes to introduce a co-optimised market design. In selecting 2035 as the end year for our modelling, we considered a trade-off between assessing over more years and the level of uncertainty regarding key inputs relating to GB system in later years such as generation, storage, transmission and historical market fundamentals (commodity prices, demand, weather etc).

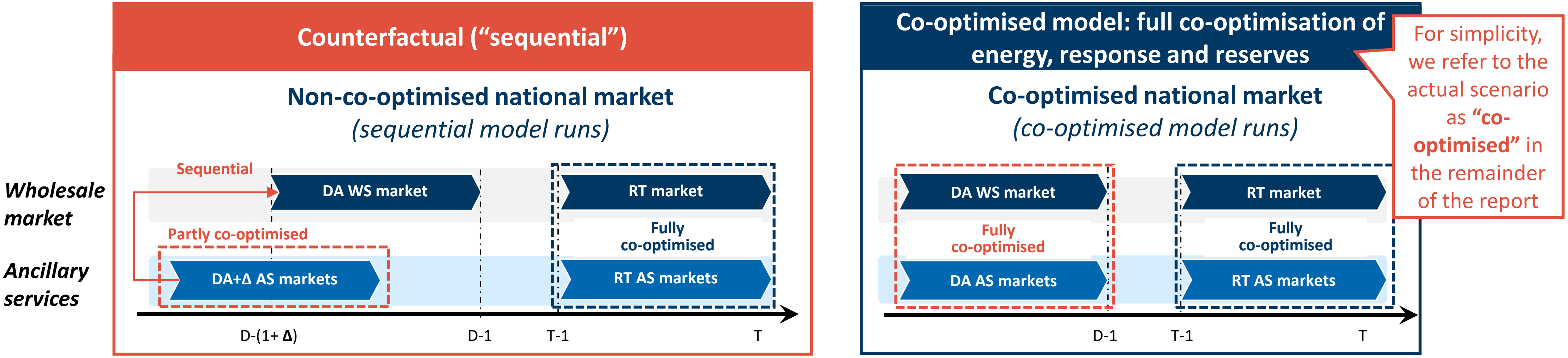
The counterfactual reflects the current GB market design: "sequential" energy and AS procurement with partly co-optimised reserve and response



- Our approach to modelling the counterfactual scenario, in terms of the WS and AS markets, is summarised in the red box above.
- To model the current GB design of energy, reserve and response markets in a tractable manner, we make several key assumptions:
  1. We assume that AS reservation markets run ahead of the DA WS market[2,3], such that market participants estimate the opportunity cost of participation in the DA energy market when bidding in the DA AS markets.
  2. The modelled DA AS market only has four markets: two markets for response (upwards/downwards), and two markets for reserve (upwards/downwards), rather than ESO's planned six response and six reserve products.[4]
  3. Building on ESO's Enduring Auction Capability ("EAC") at DA stage, launched in November 2023, we assume the procurement of response and reserve products is co-optimised ("Partly co-optimised").
  4. At RT stage, while in practice currently energy, reserve and response are managed in an 'implicit way', for modelling tractability we assume a full co-optimisation of the RT energy, redispatch and reserve markets in each hour.[5]
- Each of these assumptions is discussed in detail in the next slide.

Notes: (1) See Section 2 for the definitions of "separate", "implicit", "partly co-optimised" and "fully co-optimised" market designs; (2) Reserve markets operate with a small time lag ("Δ"); (3) We recognise that co-optimisation can occur at forward timeframes and/or real time. See slide 13; (4) The current response products (both up and down) in place are dynamic containment (DC), dynamic regulation (DR), dynamic moderation (DM) and Static FFR (up only). The previous reserve markets 'Fast Reserve' and 'STOR' are being replaced by quick reserve (QR), slow reserve (SR) and balancing reserves (BR); (5) To limit the computational burden, we use hourly settlement periods at the DA and RT stage.

We compare the counterfactual sequential model with a fully co-optimised model that clears energy and AS markets simultaneously at DA stage

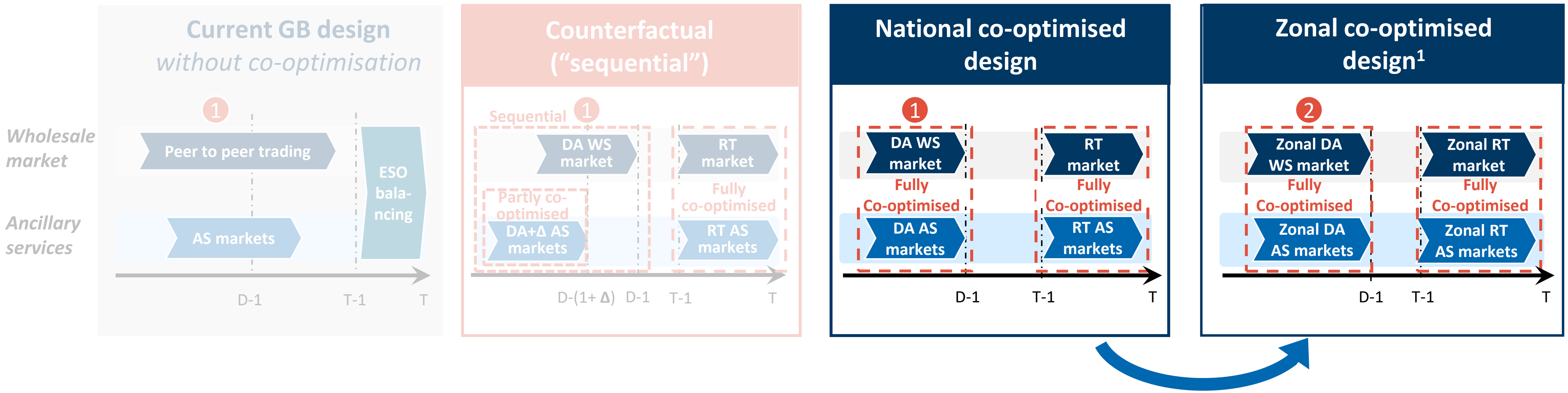


- The key difference between the modelled counterfactual and co-optimised ("actual") design is at the **DA stage**:
  - the counterfactual design features **sequential procurement of energy and partially co-optimised AS,** so assets must estimate opportunity costs of participating in AS markets relative to the WS Energy market…
  - … while in the co-optimised design the energy scheduling (via the WS market) and AS procurement are **fully (and simultaneously) co-optimised,** meaning the model provides the lowest cost solution to meet the combined WS energy demand and AS requirements.
- From a technical perspective, under both models the **objective function of the DA WS market is the maximisation of welfare;** the difference lies in the **additional response and reserve constraints that need to be satisfied in the DA WS market clearing under the co-optimised model**.[1]
- By contrast, there are **no reserve and response constraints in the DA WS market clearing under the sequential model**. Instead, reservations for response and reserves are made in the preceding DA AS reservation market. The objective function of the DA AS reservation market, where all modelled reserve and response products are co-cleared, is to maximise welfare. In the subsequent **DA WS market clearing** the generation / consumption capacity of all assets is **adjusted according to their commitments in the DA AS markets** (see Section 4).
- In line with the counterfactual model, full co-optimisation of energy and AS is assumed at the RT stage. However, as the resources reserved in the DA-stage can be different, the **resources that are activated in RT to solve imbalances are not necessarily the same**.

Note: (1) The maximisation of welfare equals the minimisation of system costs in our modelling as we do not represent demand via utility functions but instead have price-inelastic DA AS demand and DA WS energy demand. However, in the DA market clearing we consider shiftable demand ("smart demand") and tranches of interruptible demand ("demand side response"). The latter is assumed to have no disutility for shifting and for the former the disutility equals its variable cost.

We also assess the incremental impact of transitioning from national to zonal wholesale pricing under co-optimisation

We start from a national WS market design and examine how a zonal WS electricity market design impacts co-optimisation.



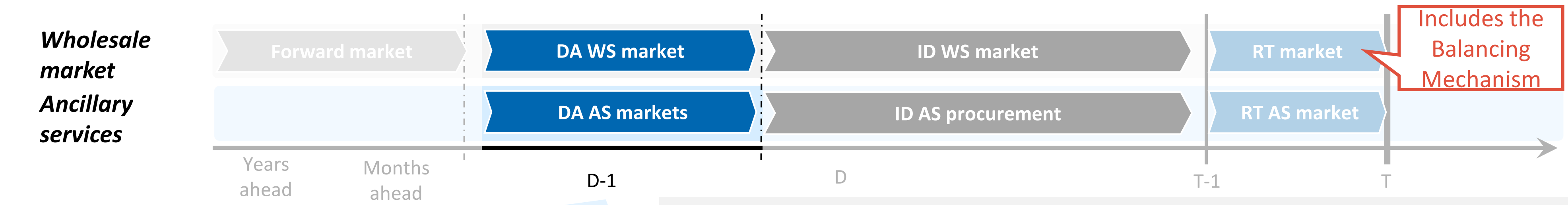
Impact of introducing zonal pricing

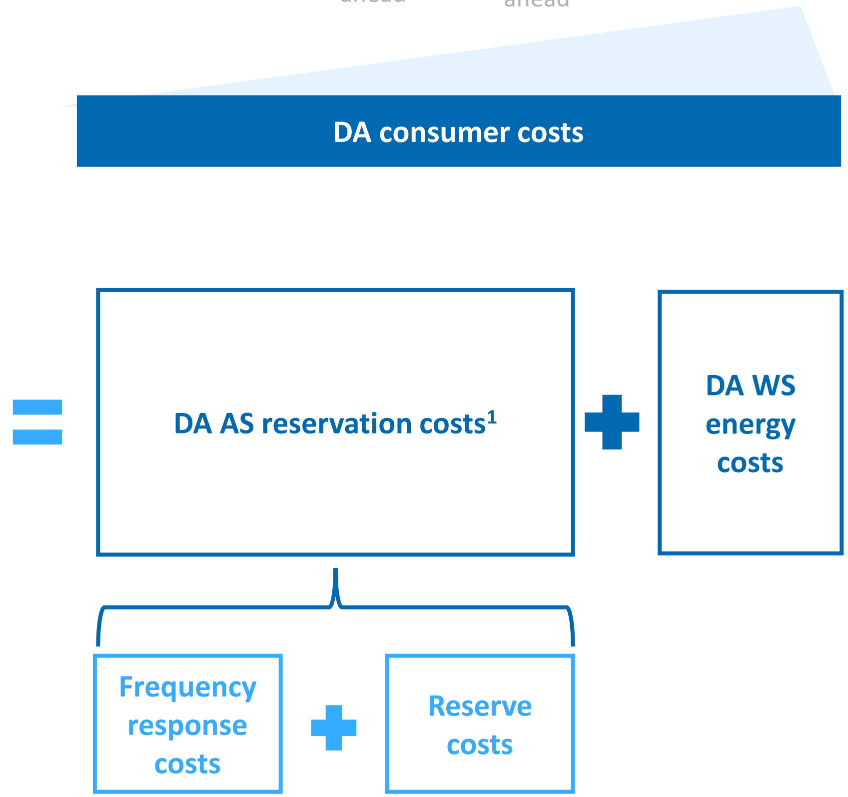
1 National DA WS market	▪ The GB transmission network is not reflected in the DA WS and/or AS markets (in line with how these markets currently operate) ▪ Pricing is **national** for the WS energy and AS markets
2 Zonal DA WS market[1]	▪ Inter-zonal transmission limits are considered in the co-optimised zonal DA WS energy and zonal DA reserve markets, leading to **zonal** WS energy and **zonal** reserve prices ▪ Response continues to be procured at a national level (i.e. ignoring intra-GB transmission) and is still priced nationally[2] ▪ Trading of energy in the WS market between zones is possible at DA up to the limits of the inter-zonal transmission capacity ▪ The total zonal reserve requirement (aggregated over all zones) is the same as the reserve requirement under national pricing – the allocation of reserve requirements per zone is discussed in Section 5 ▪ Only resources located in a zone can be procured to fulfil the zonal reserve market requirements (i.e. no inter-zonal trading of reserve)

Notes: (1) This is not necessarily reflective of how ESO sees AS potentially being procured in a sequential zonal market, which is out of scope; (2) Typically a reserve margin of all transmission network elements is reserved to ensure the delivery of response resources when being activated.

A key output of our modelling is the estimate of **consumer costs** under each market design, of which **DA wholesale costs** represent the largest share

Key elements of wholesale and ancillary services markets





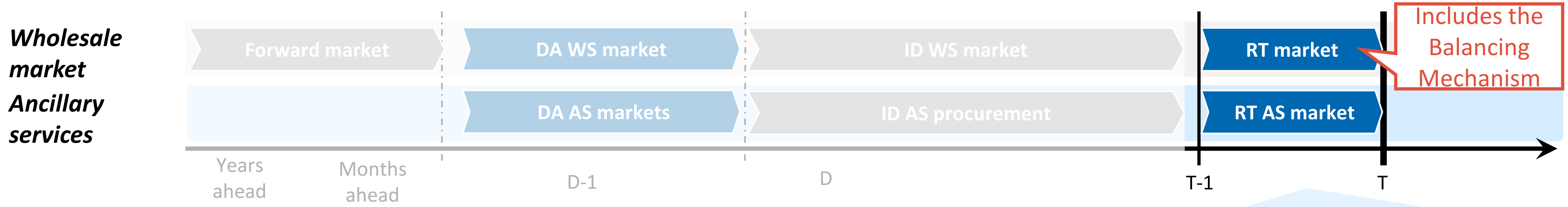
- Total consumer costs are assessed across the **four DA AS markets** and the **DA WS market.** The consumer costs per market are computed as the **hourly price per market multiplied by the hourly demand per market** plus **possible uplifts per market.**
  - **Uplifts** are required as the marginal clearing price, which reflects the additional system costs incurred to meet a marginal increase in the demand for energy or the relevant AS product, might not be sufficient for thermal units to recover all their operating costs (e.g. start-up costs and potential losses from running at minimum load) when scheduled in the WS or providing AS.
- Under the **“sequential” model**:
  1. The **hourly price per AS reservation market** is determined based on AS bids from market participants that are co-cleared simultaneously to identify the least-cost way of meeting the **pre-determined response and reserve demand**.
  2. Market participants reflect start-up costs or losses from running at their minimum stable load in their AS bids, as such there are **no uplifts in the AS markets**. Uplifts are only possible in the DA WS market.
  3. The **hourly DA WS electricity price** is determined in the DA WS market clearing. Units that were successful in the DA AS markets must respect its AS commitments, e.g. a unit successful in reserve up cannot bid in its full generation capacity, but only the share of the capacity that is not reserved. **Uplifts are possible** in **the WS market.**
- Under the **co-optimised model**, the **hourly AS price** in each AS market represents the **additional system cost of a marginal increase in the reserve requirement of that AS. Uplifts** are included in **both the DA AS and WS markets.**[2]

Notes: (1) We do not model other ancillary services such as stability, voltage control or black-start; (2) See Appendix 12 for further detail.

FTI CONSULTING

Real-time consumer costs are expected to be relatively similar between the "sequential" and the co-optimised model (compared to day-ahead costs)

Key elements of wholesale and ancillary services markets



- As shown in the figure to the right, real-time costs can be conceptually split into[1]:
  - **RT energy activation costs** to solve energy imbalances (most relevant to co-optimisation);
  - **redispatch costs** to relieve thermal constraints; and
  - **RT AS reservation costs** incurred to reserve additional units when the reserves are depleting due to activations.
- However, in practice, RT actions are considered jointly by ESO and are difficult to distinguish. In terms of magnitude:
  - Most **RT costs are incurred due to redispatch under a national WS market design**. Without locational pricing, co-optimisation will not inherently favour or disfavour redispatch costs. However, random impacts from reservations lead to **minor differences** in redispatch costs between the sequential and co-optimised models.
  - **RT energy activation costs** can be **slightly different** under the counterfactual and co-optimised model because different units are reserved at the DA stage.
  - **RT reservation costs** are expected to be **similar** under both models.
- RT redispatch costs are c. 3x larger than RT energy activation costs (see Table on the right), and **differences in RT energy activation costs** due to the co-optimised rather than the sequential DA procurement of AS reserves **are not directly observable**.

Real-time consumer costs

Changes in wholesale electricity costs after day-ahead	+ **RT energy activation costs**	+ **Redispatch costs**	+ **RT AS reservation costs**
Impacted by DA co-optimisation of energy and AS?	To some extent	Not systematically	No
Proportion of BM costs (April '23- March '24)[3]	18%	54%	0-5%

Impacts on RT energy activation costs conflated with random deviations in redispatch costs

Notes: (1) We do consider real-time activations for stability, voltage or inertia purposes here; (2) Modelling "implicit" co-optimisation is not tractable; (3) The BM costs not included here are costs for reactive power, restoration, response and other minor cost components. More details are provided in Section 6.

Our two main evaluation metrics of the merits of co-optimisation between energy and ancillary services are **consumer costs** and **total system costs**

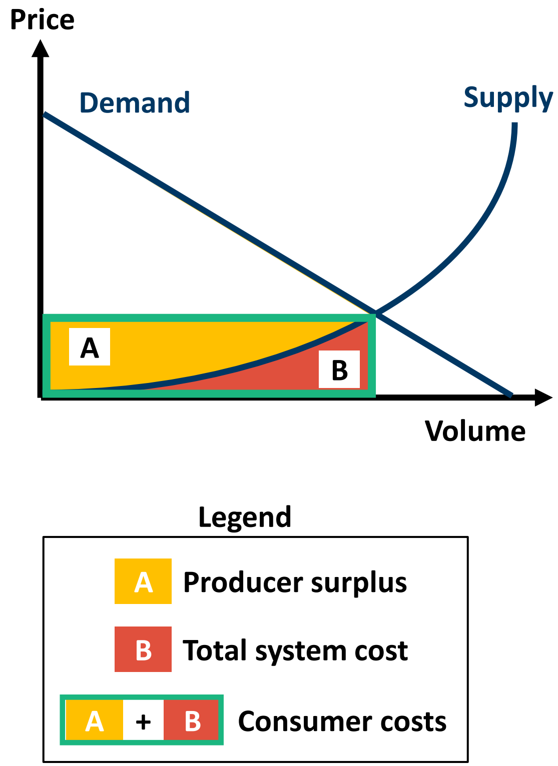
- Our **qualitative report**, published alongside this report, identifies multiple benefits and perceived risks of co-optimisation. Of these, the critical element is the allocative efficiency arising from co-optimisation, i.e. lowering total system costs by improving the allocation of resources across different markets.
- In this report, we quantify the **allocative efficiency impact** from both consumer and societal perspectives.
- We also analyse **changes in DA WS energy and AS reservation prices** under co-optimisation, as a proxy for long-term price signals (and, hence, **dynamic efficiency**).
- **Consumer costs** consist of the DA WS energy costs and the DA AS reservation costs.
  - DA WS energy cost equals the WS energy price multiplied by the load plus the required uplifts.[2]
  - DA AS reservation costs equal the AS reservation prices multiplied by the reserve demand for each product. Under the co-optimised model, uplifts are also applied to DA AS prices.

Consumer costs = DA WS energy cost + DA AS reservation costs

- **Total system costs** equal the scheduling costs for generators required to meet GB demand (i.e. adjusted for imports and exports). Changes in system costs are the key metric to measure allocative efficiency gains/losses.
  - Scheduling costs are the operational costs resulting from the DA WS energy schedule, namely fuel and carbon, variable maintenance and start-up costs.
  - AS reservations indirectly impact scheduling costs. When being reserved, certain resources need to adjust their DA energy schedule to comply with reserve requirements.
  - For example, if a thermal unit is selected to provide reserve up, it must ensure that it is scheduled in the DA WS market at a generation level between its minimum stable load and its maximum generation capacity minus the reserved capacity.
- The formula used to adjust system costs for import and export is provided in Appendix 2.

Total system costs = DA GB scheduling costs + Adjustments for imports/ exports

Schematic representation of the relationship between consumer costs, total system costs and producer surplus[1]



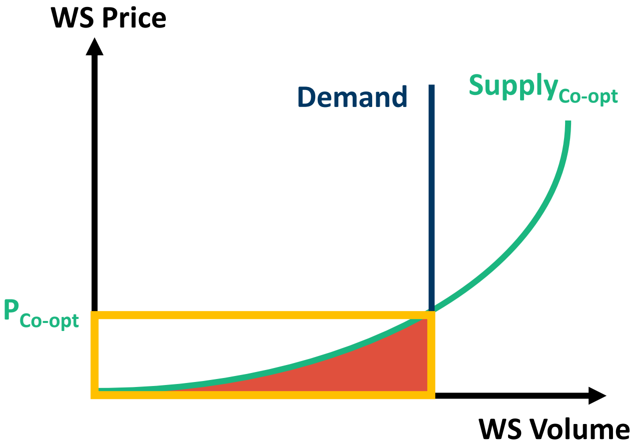
In the next slide we provide an illustration of the link between total system costs and consumer costs

Notes: (1) Producer surplus represents the revenues of GB power producers and storage from the DA WS energy and AS reservation markets minus their DA scheduling costs. The diagram excludes interconnector impact, for simplicity; (2) Details on the calculation of uplifts are provided in Appendix 12.

FTI CONSULTING

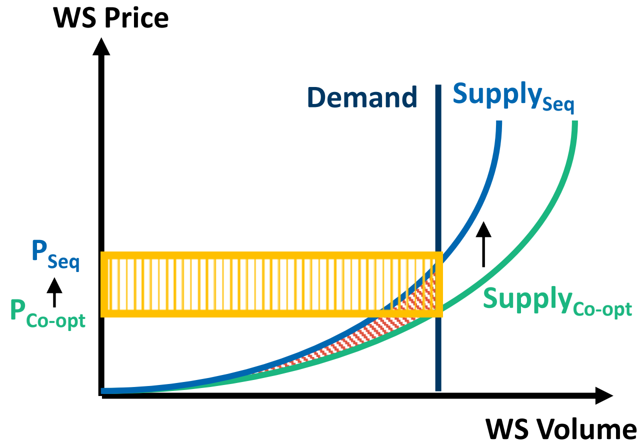
Inefficient scheduling in Sequential markets increases total system costs, but may or may not increase consumer costs relative to Co-optimisation

Outcome A: If Sequential scheduling is efficient, outcomes will be identical to Co-optimisation



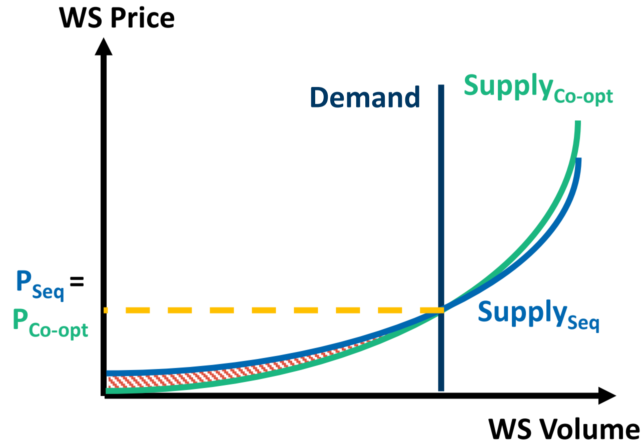
- By design, co-optimisation delivers an efficient schedule which optimises over WS and AS markets to **minimise total system costs**.
- Sequential markets may not impact scheduling, if all units accurately forecast their opportunity costs. This can lead to identical system and consumer cost outcomes under Co-optimisation and Sequential design.

Outcome B: Inefficient scheduling in Sequential markets can increase WS consumer costs more significantly than system costs



- Inefficient scheduling in Sequential markets distorts the merit order and thereby shifts the supply curve upwards.
- If the distorted units would have been **marginal** in the WS Energy market under Co-optimisation, then the distortion will increase the WS Energy price in Sequential.
- This can **significantly increase WS consumer costs**, with potentially a **smaller increase in total system costs**.

Outcome C: Inefficient scheduling in Sequential markets may have no impact on consumer costs, but still increase system costs



- In contrast, if the distorted units are **infra-marginal**, then the WS energy price will remain the same across the Sequential and Co-optimised models.
- In this case **WS consumer costs** will also **remain unchanged**, but there will again be a **small increase in total system costs**.

Legend Total system cost | Consumer costs | Increase in total system cost | Increase in consumer costs

See Appendix 3 for a series of worked-through examples of the dynamics of price formation.

Note: (1) We use a perfectly inelastic demand curve in the diagrams above for simplicity. The qualitative results remain unchanged with an elastic demand curve.

Section 4: Day-ahead sequential design and co-optimisation under national pricing

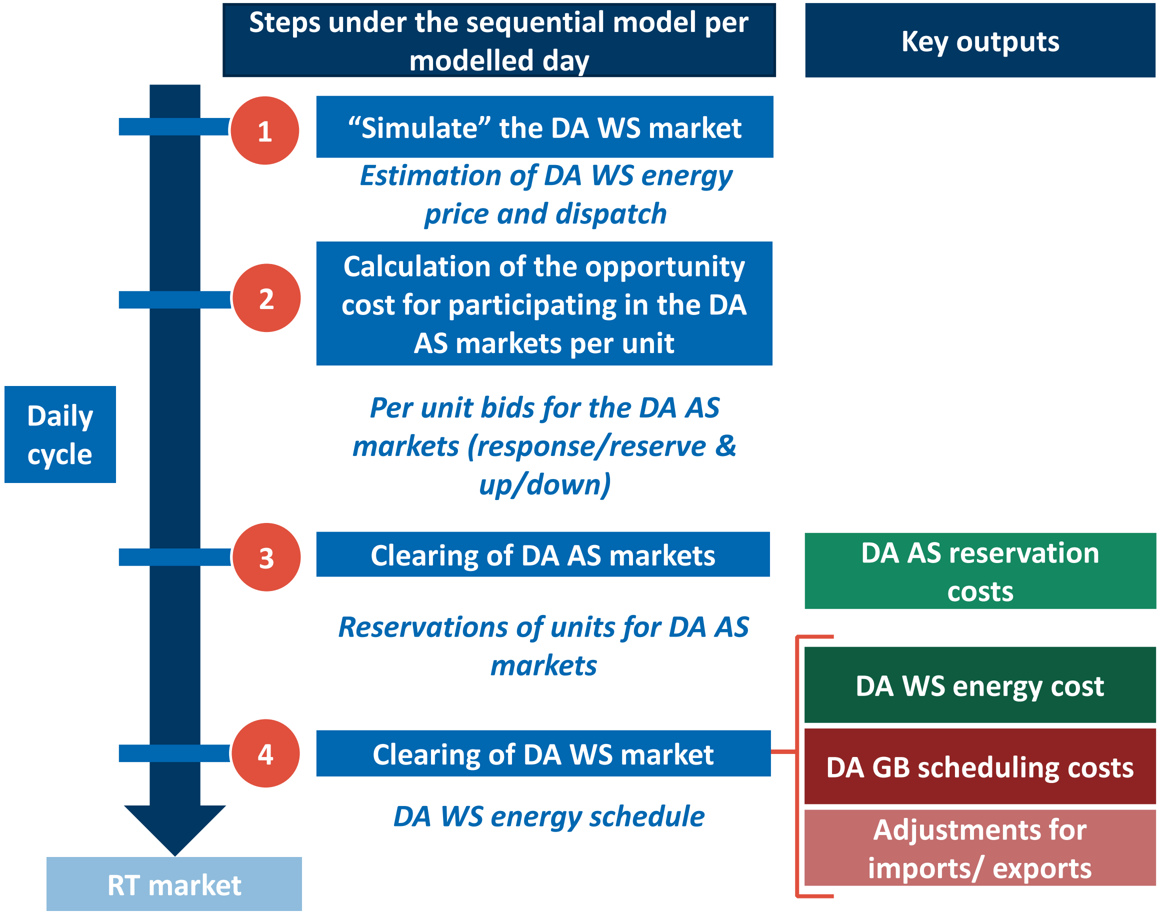
FTI CONSULTING

Implementation of the methodology

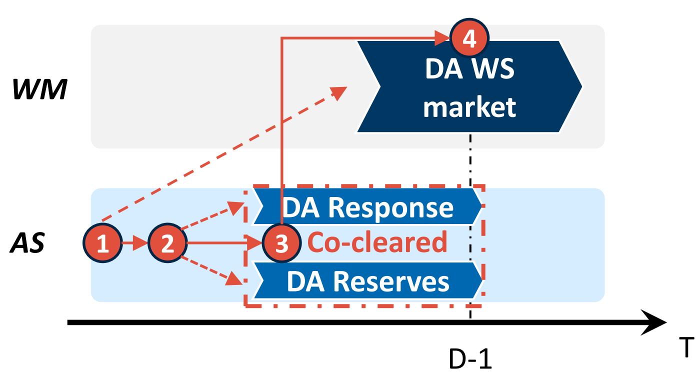
FTI CONSULTING

In the sequential model, market participants in the DA AS markets need to form expectations on the clearing prices of the (subsequent) DA WS market

- The counterfactual "sequential" model setup consists of **four steps** as illustrated in the figure below. For each modelled day, the same four steps are repeated. In step 4, the daily DA WS energy market clearing (with hourly resolution) considers a one-day look-ahead to avoid batteries depleting at the end of the day and not discourage start-up decisions.
- We outline each step in further detail in the following slides.



Sequential national DA market *(counterfactual)*



Note: (1) For thermal units we consider start-up costs, ramping rates and minimum stable load ("MSL") levels. We do not consider shut-down costs or minimum running requirements. Ramping constraints are also considered when determining how much unit upward/downward reserve a unit can provide.

The opportunity cost of being reserved in AS depends on assets' expectation of the next day's DA WS energy prices and production schedules

DA AS bids are dependent on the opportunity costs arising from expected DA WS energy prices and schedules

- The provision of energy and AS are partial **substitutes**. A generator's provision of one product **limits its ability to provide another**.
- Being reserved to provide AS therefore incurs **opportunity costs**, linked to the **revenues foregone** in WS market, and it may also change the planned production schedule and associated costs.
    - For units that expect to be 'in merit', the higher the expected WS energy price, the more rent would be missed out by being reserved upwards. In other words, the higher the **opportunity cost** and hence the higher the unit's bid for being reserved up.
    - Generators must also have a view of their likely **production schedule**. A thermal generator does not need to internalise its **start-up costs** in its DA AS bid when it expects to be scheduled in the DA WS energy market (as it will already be scheduled).
- This means **market participants** must first **forecast DA WS energy prices** and their **resulting expected production schedule** to understand their **own opportunity cost of providing AS** .
- Forecasts of the DA WS market outcome are complicated by **uncertainty** on the final energy demand, renewable production, possible outages etc. A further complication is that the **bidding of other market participants** in the DA energy market will be a function of which DA AS markets other participants will tender in and their **likelihood of being successfully cleared in the DA AS markets**.

Modelling assumptions of forecasts of the DA energy prices and production schedules

- **Imperfect forecasting** under **sequential procurement** and the resulting **misallocation of resources** between energy and AS markets is a **key driver** of the **benefits** case **of co-optimisation.**
- As a **conservative assumption**, in our modelling we use the **DA WS energy prices and scheduling decisions from the co-optimised model as the basis of market participants' forecast.**
- This assumption is conservative as the WS energy prices and schedules under the sequential and the co-optimised model would be **identical in case all market participants would have perfect information**.
- **In practice, not all market players will make the same forecasts**. To reflect this, we add **minor random noise[1] to the DA WS energy price forecasts** used to calculate opportunity costs in step 2.
- Forecasted energy schedules used by participants are all the same – these are the WS energy schedules resulting from the co-optimised model.

On the next slide we provide an illustration of the implications of a **market participant making a forecast error in the DA energy prices**.

Note: (1) In our modelling scenarios we have used a mean absolute percentage error (MAPE) of 5% between the DA WS energy prices resulting from the co-optimised model (used as a basis for the forecasts) and the forecasted DA WS energy price used to calculate each market participants' opportunity costs. Each thermal generator receives a different forecast, storage resources are organised as portfolios for computational reasons. We provide more details in Appendices 4-5.

Incorrect estimation of the next day DA energy prices and/or schedules can cause suboptimal reservation of resources in the DA AS market

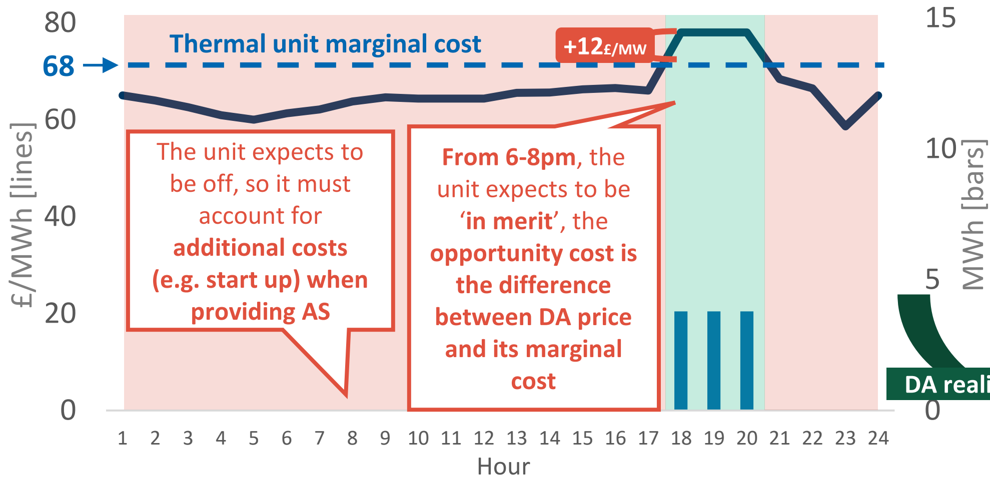
Illustration of how a thermal unit estimates the DA WS energy prices and its production schedule, and the implications for its opportunity costs

- In the stylised example below, a thermal unit **expects to generate at full capacity of 5MW for three hours** (6-9pm) and be offline for the rest of the day.
- In those three hours (highlighted in green), the unit expects to be '**in merit**' and faces an **opportunity cost** of upward AS provision equal to it inframarginal rent (80 £/MWh – 68 £/MWh =12 £/MWh).[1] The unit does not need to internalise start-up costs in its DA AS bid (as it already expects to be online).
- In other hours (highlighted in pink), its opportunity cost for upward AS provision is zero (no foregone wholesale revenue), but there are **additional costs** that need to be accounted for (e.g. start-up and/or losses for running at minimum load when started up earlier than 6pm or shut down later than 9pm)

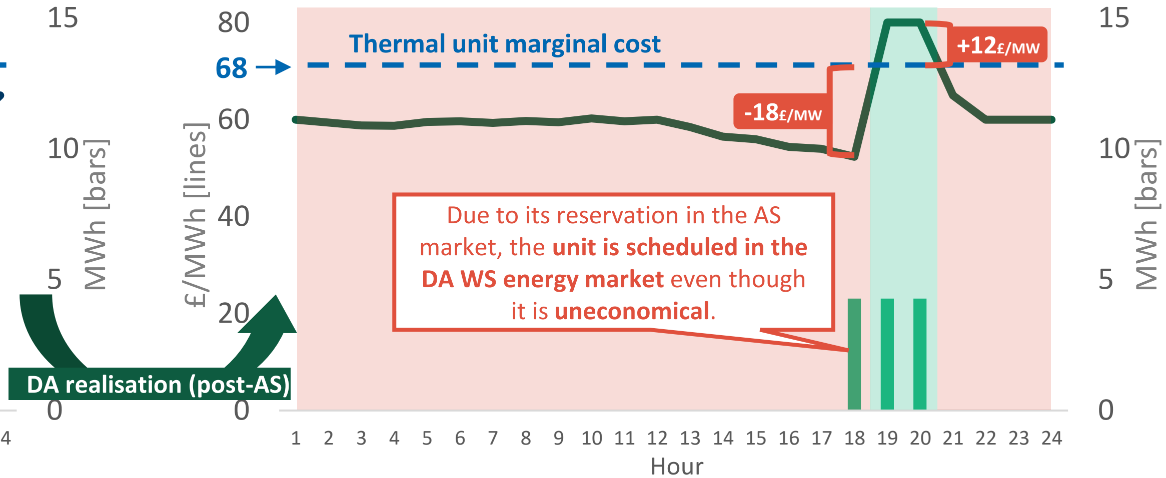
Illustration of how incorrectly estimated DA WS energy prices impact a thermal unit's revenues and overall system costs

- Based on its expectations of WS outcomes, the thermal unit has **bid 12 £/MW** in the upwards reserve market for 6-9pm for its entire available capacity. We assume that the **unit is cleared in the reserve upwards market in all 3 hours for the entire capacity**.
- This implies that the unit **needs to be scheduled at its minimum stable load ("MSL")** during 6-9pm; it will **bid in that capacity at zero** (see also slide 34).
- But its **price expectations were formed incorrectly and the WS energy price rises to the maximum one hour later** (i.e. 7pm). This has two important consequences:
  - The unit is now potentially making a loss between 6-7pm (**earning £12 per MW reserved in AS but unexpectedly losing £18 per MWh scheduled in the WS market**) Another lower cost unit could have been scheduled at 6pm leading to **increasing system costs and possibly impacting WS energy prices relative to co-optimisation**

Forecasted DA WS energy prices used to estimate opportunity costs



Realised DA WS energy prices and dispatch of the thermal unit



Estimated DA WS energy price (£/MWh)
Expected schedule of thermal unit (MWh)
Opportunity costs for AS provision ('in merit')
Additional costs for AS provision ('out merit')
Realised DA WS energy price (£/MWh)
Realised schedule of thermal unit (MWh)

Note: (1) Here we assume the unit would be made whole (i.e. avoid making losses) in case the start-up cost would not be recuperated over the course of the day via the marginal WS energy price.

Different technologies face different opportunity costs which depend as well on the direction of the AS market

	Upward AS *'Increase output or reduce load if called upon by ESO'*	Downward AS *'Reduce output or increase load if called upon by ESO'*
Thermal generators	■ To provide upward AS, units must **withhold generation capacity** from the **DA energy market**. ■ If expected to be '**in merit**' in WM, opportunity cost is **revenue foregone** in WM **minus its short-run marginal costs ("SRMC")**. ■ If '**out of merit**', opportunity cost is zero, but the bid must reflect **start costs and/or losses** from being scheduled at MSL in the WM.[1]	■ To provide downward AS, units must be able to **reduce output** if called upon by ESO. ■ If expected to be **'in merit'** in WM and scheduled above its MSL, opportunity cost is zero. ■ If **'out of merit'**, bid price must reflect both the costs of being **scheduled at MSL** and the **costs of the generation above MSL** to be made **available for turndown** in the WM.
Storage-based units	■ **Opportunity cost** reflecting that AS reservation **restricts discharging** during **highest-priced hours** (referred to as **"power" opportunity costs**). ■ **Opportunity cost** when AS reservation requires **maintaining state-of-charge, reducing the opportunity to discharge in earlier hours**, e.g. during peak-price hours (referred to as **"energy" opportunity costs**).	■ **Opportunity cost** reflecting that AS reservation **restricts charging** during **lowest-priced hours** (referred to as **"power" opportunity costs**). ■ Occurs when AS reservation **requires maintaining storage headroom, reducing the opportunity to charge** during **earlier hours,** e.g. during off-peak hours (referred to as **"energy" opportunity costs**).
Demand-side units	■ If scheduled to consume in the WM, the **opportunity cost of AS reservation is zero at DA**. However, DSR may bid a high price for activation in RT.	■ If not scheduled at its maximum consumption level, the opportunity cost of **AS reservation is zero at DA.**
Renewables	■ Intermittent renewables are **not considered to provide upward AS.**[2] Run-of-river hydro can provide upwards reserve; its opportunity cost is the **estimated DA price**.	■ When scheduled to produce, opportunity cost **of AS reservation is zero. Less than full production capacity is offered** due to uncertainty in final production levels.

Notes: (1) Start-up costs are allocated in all consecutive hours after a shutdown and before a start-up, which can be considered a risk-averse bidding strategy. Also, if the cost of anticipating an expected start-up and running the plant is cheaper than having two separate start-ups, the start-up cost will equal the former. Must-run thermal units (e.g. combined heat and power (CHP) plants with a heat demand) have no opportunity costs in case the forecasted energy price is higher than their SRMC – in that case they will be producing at MSL. More details are provided in Appendix 5; (2) Due to the high opportunity cost of foregone subsidy under the current CfD scheme.

In the sequential model, AS bids are jointly cleared for all AS markets ("partly co-optimised") for the least-cost solution respecting units' physical constraints

Illustration of how units prepare AS bids in the sequential model

- In the sequential model, **all units** prepare AS bids for the **markets that they are eligible for**.
- In line with **EAC design**[2], these bids can be **mutually exclusive**. This means that at the bidding phase, units **do not have to choose which market they participate in**.

Illustrative bidding by a battery unit

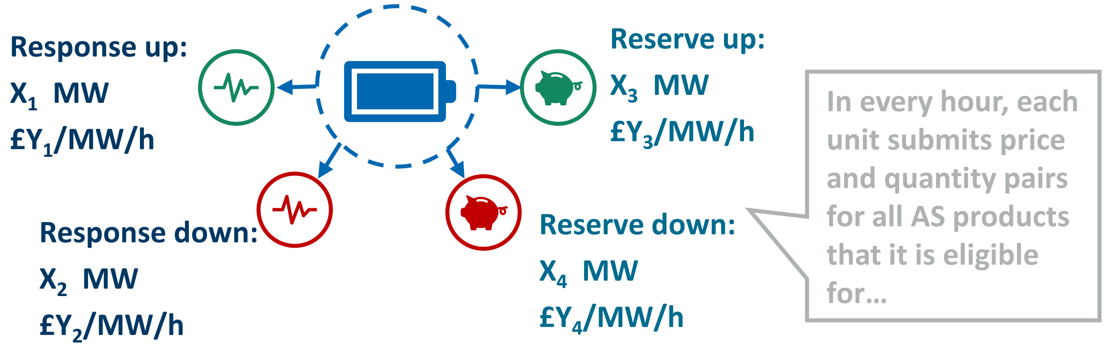
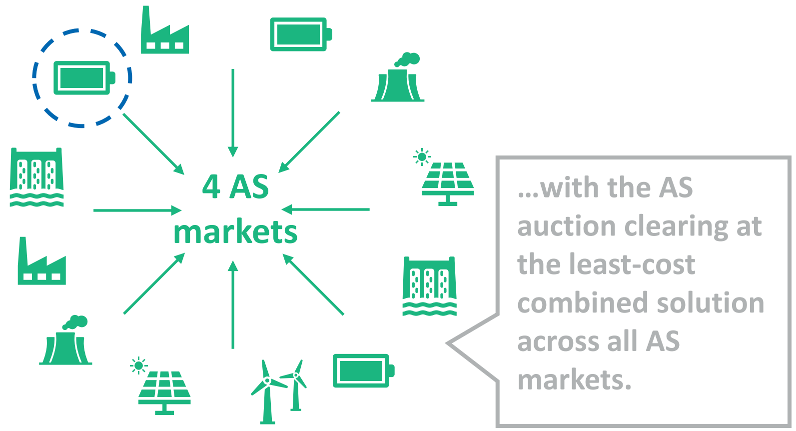


Illustration of the joint-clearing of AS markets in the sequential model[1]

- The AS bid and offer price and quantity pairs from all units are fed into a **combined AS market clearing**, which seeks to **minimise total AS reservation costs across the four AS markets**.
- Potential **incompatibilities** of each unit's bids are **addressed in the joint-clearing**.



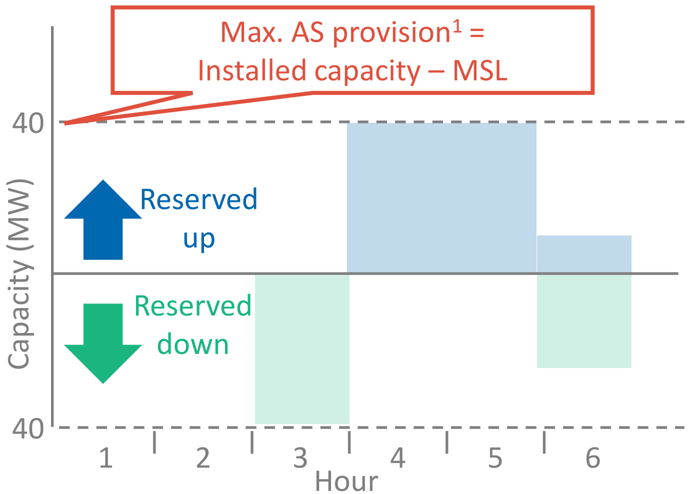
The hourly combined clearing of AS markets is bound by four key constraints:[3]

1. **Demand for each AS market must be met in all hours.** The shadow value of these constraints determines the AS price.[4]
2. **Reserved capacity must physically feasible.** The solution respects the technical properties of each generator (e.g. MSL level).[5]
3. **Reserved capacity across all markets must be compatible.** Reservations in one AS market restrict the available volume in another AS market (e.g. the total procurement volume needs to be lower than the available capacity).
4. **No unit is forced to provide AS in any hour.** There is no mandatory provision of ancillary services by any units.

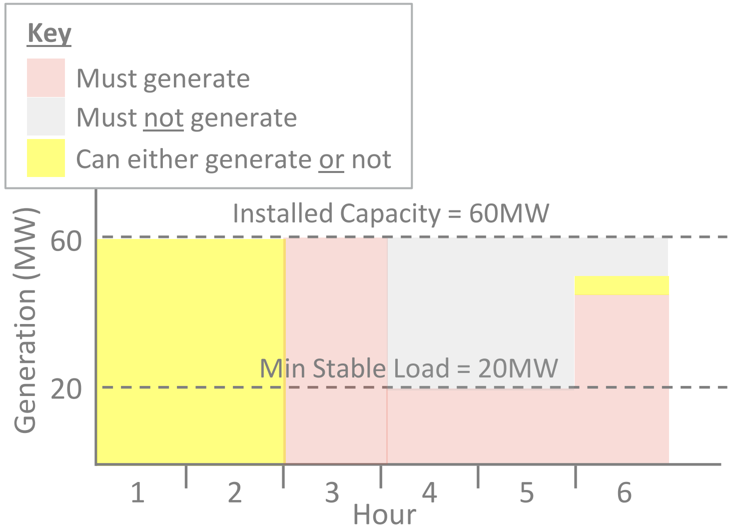
Notes: (1) In the co-optimised model AS markets are also jointly cleared, but the process is extended to also incorporate energy scheduling (via the WS market); (2) Enduring Auction Capability ([link]); (3) A more detailed mathematical description is provided in Appendix 6; (4) The shadow price indicates the additional reservation costs when procuring a small increment of additional response/reserves, i.e. it represents the marginal price of the AS product; (5) Storage technologies also require a certain level of energy available to be able to provide AS. The lower the duration of the storage, the stricter this requirement. More details are provided in Appendix 7.

The resulting DA AS market reservations impact the bidding of reserved units in the DA energy market and consequently DA energy prices

Illustration of the reservation of a thermal unit from the clearing of the joint AS market (step 3)



Reservations in DA AS market impact the scheduling in the subsequent DA WS energy market (step 4)[2]



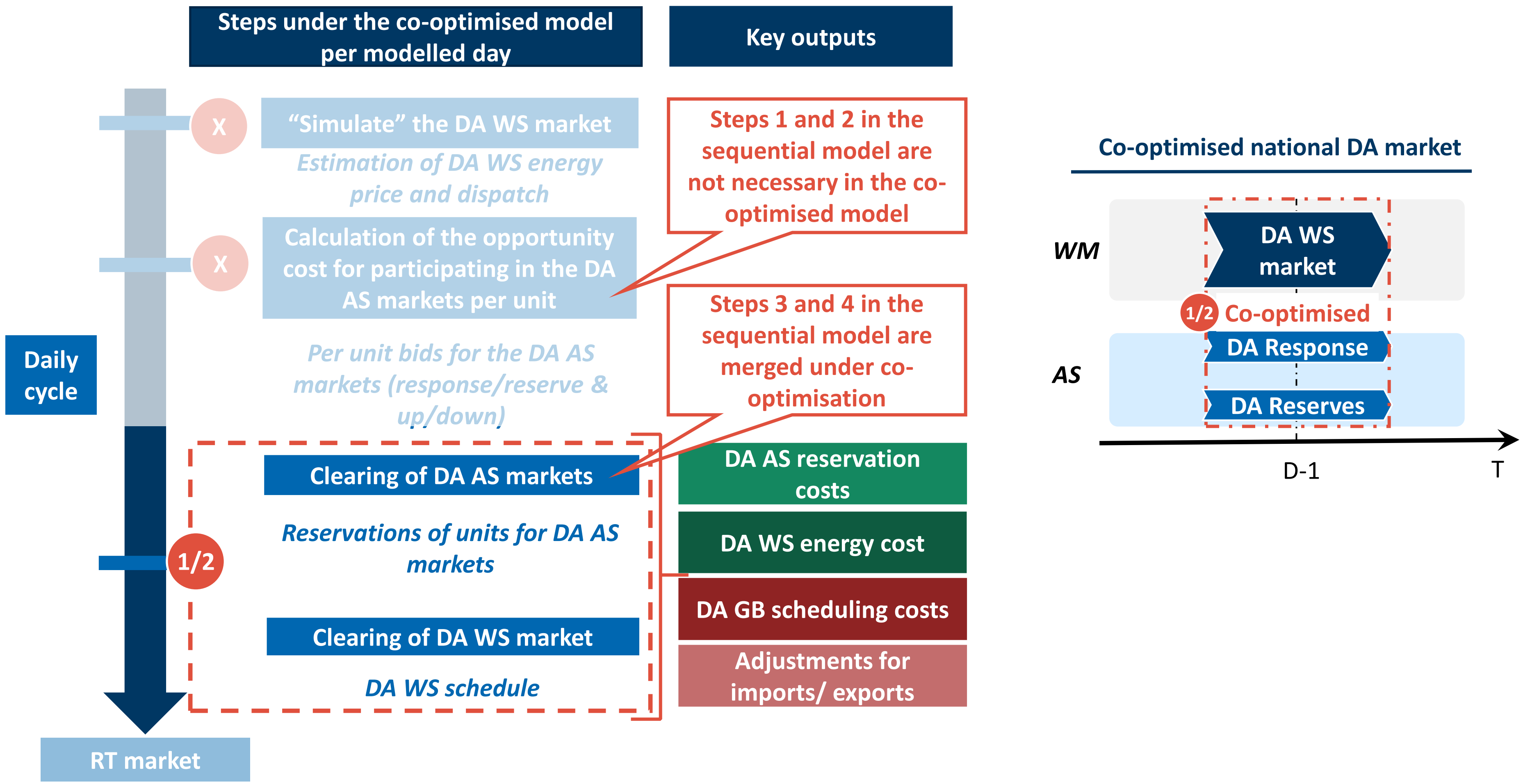
- The **constraints resulting from AS reservations** are reflected in the DA WS energy schedule to reflect how **the reserved AS units can bid in the WS**.
- In the example above, the unit has an installed capacity of 60MW and an MSL of 20MW:
  - In **hours 1-2**, the unit is not reserved in any DA AS market, hence it is free to either not generate, or generate at any level between 20-60MW (i.e. between its MSL and installed capacity).
  - In **hour 3**, the unit has 40MW reserved for downward AS, hence it must be scheduled in the DA WS market at 60MW (i.e. its full installed capacity) to be able to reduce generation to its MSL if activated (i.e. by 40MW). This implies the unit will bid in WS its entire capacity at 0 £/MWh during this hour.
  - In **hours 4-5**, the unit has 40MW reserved for upward AS (the maximum possible for this unit), hence it must generate in the wholesale market at MSL (i.e. 40MW below is full capacity). The unit will bid MSL capacity at 0 £/MWh in WS (to be online in those hours); but it cannot offer more in the WS.
  - In **hour 6**, the unit has 25MW reserved for downward AS and 10MW reserved for upward AS, hence it must be scheduled in the WS market between 45-50MW. The unit will bid its MSL capacity at 0 £/MWh and the 5 MW of available capacity at its SRMC.

AS reservations constrain a unit's WS market bidding strategy and can therefore impact the DA WS energy prices and DA WS scheduling costs.

Notes: (1) Technically a unit can provide its entire capacity as downward AS but doing so could require a unit to shut down and possibly incur an expensive startup after; (2) Here we ignore ramping for simplicity. Storage units reserved in a given hour must also ensure they have a sufficient state of charge at the start of that hour to provide the service if activated, hence impacting charging / discharging decisions in previous hours, see Appendix 5 for more details.

Under co-optimisation, step 1 and step 2 of the sequential model are no longer necessary and the AS and energy markets are cleared simultaneously

- In the co-optimised model, there is **no need for all units to formulate AS bids** based on estimated opportunity costs. This means that step 1 and step 2 do not exist in the co-optimised model.
- Instead, the DA AS markets and the DA WS market are jointly cleared **("fully co-optimised").** In other words, the **demand for the different AS products are formulated as constraints** in the DA WS clearing, hence **indirectly impacting DA WS energy prices.**[1]



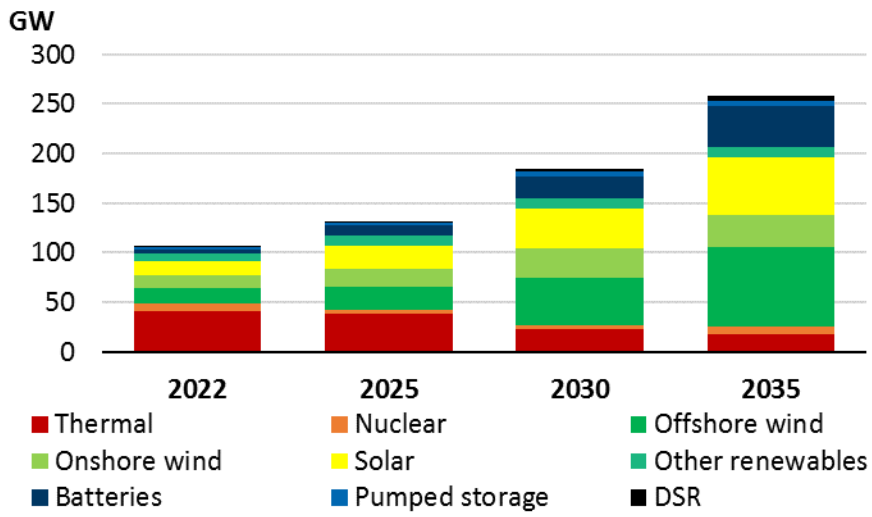
Notes: (1) This implies that scarcity in WS markets is reflected in AS markets and vice versa, while under sequential markets scarcity in WS markets is reflected in AS bids but scarcity in AS markets is not necessarily reflected in WS markets.

Modelling input assumptions

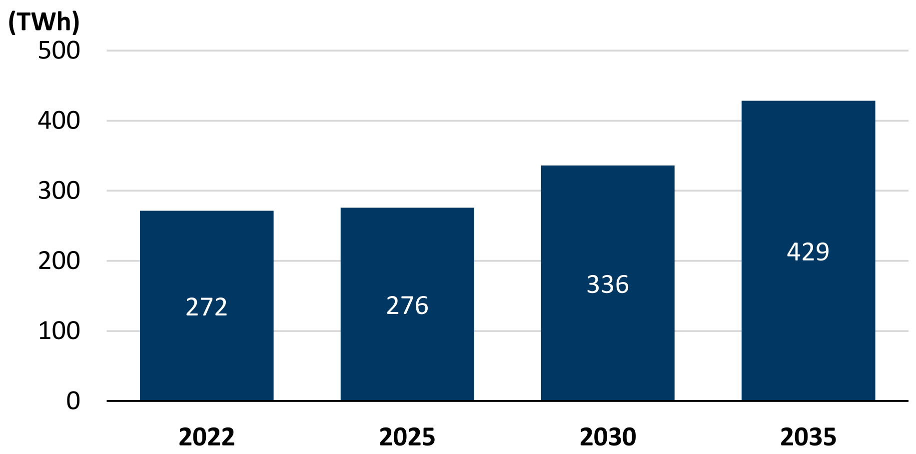
We use the FES 2022 Leading the Way scenario, with rapid deployment of renewable generation capacity

- In agreement with NG ESO, we have based the analysis in this report on the **FES 2022 Leading the Way (LtW) scenario**. We do not expect the choice of scenario to be a key driver of observed results and would **expect qualitatively similar outcomes under other FES scenarios.**
- The LtW scenario forecasts **significant growth in RES and storage capacity**, with **wind** capacity forecast to grow from **28 GW in 2022 to 113 GW by 2035**, and **battery** capacity to grow from **3 GW in 2022 to 42 GW in 2035.**
- In parallel, the volume of **thermal capacity** on the system progressively **decreases** from 41 GW in 2022 to 18 GW in 2035.
- For the 2022 historical assessment, we use **historical capacity data** from **FES 2023** and generation data from **Elexon** and **ESO**.
- Electricity **demand** is derived from **Elexon** for the **historical assessment**, and **FES 2022 LtW** for the **forward assessment**.
- **Electrification** of heating and transport causes **annual electricity demand to increase significantly** under Leading the Way, with a **high penetration of price-responsive** smart charging of electric vehicles.
- Total annual electricity demand is **276 TWh** in **2025, 336 TWh** in **2030,** and **429 TWh** in **2035**.
- For the **forward-looking** assessment, **interconnector** deployment follows the **LtW** scenario, with **over 20 GW of capacity by 2035.**
- In the **historical** assessment, **hourly interconnector flows** are **fixed** to match **historically-observed outcomes**.

Electricity generation capacity by technology based on FES 2022 LtW scenario (GW)[1]



Annual power demand, 2022-35 (TWh)



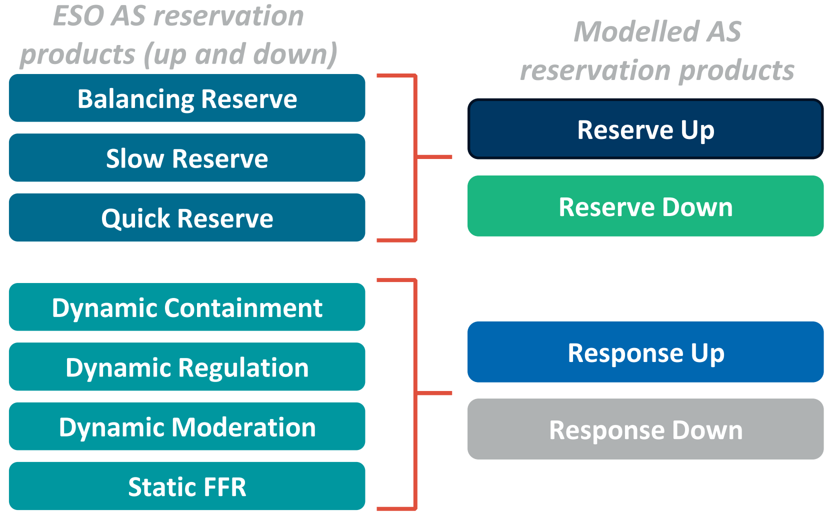
Note: (1) Thermal includes: CCGT, CCS, gas turbine, hydrogen, oil turbine, coal and CHP units; other renewables includes biomass, CCS biomass, hydro and waste.

We have aggregated ESO's AS products into four products – Up/Down Reserve and Response – with reserve demand increasing in line with wind deployment

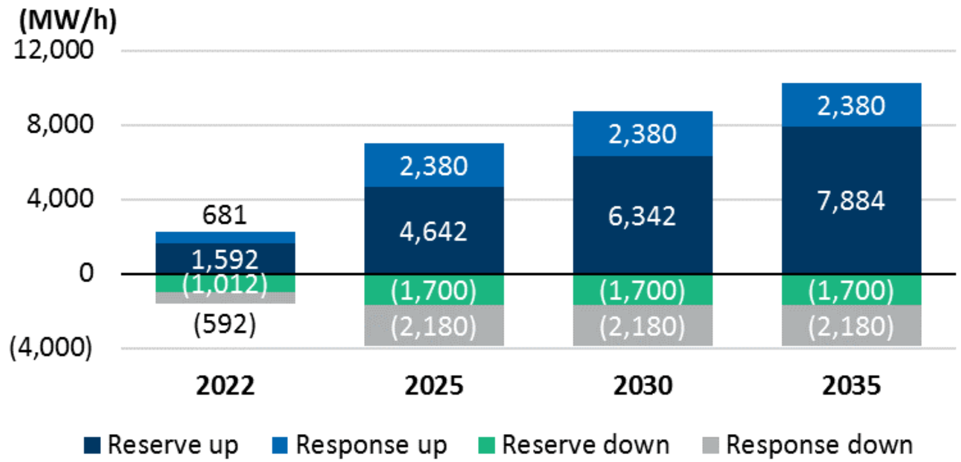
- The **demand** for each of the future **AS products** has been calculated based on **NG ESO inputs.**
- All **AS products are aggregated into reserve or response**, and each of those is modelled in two groups (up and down), respectively.
- For the **historical assessment**, hourly AS demand is relatively **variable across the year**, in line with **market conditions in 2022**. See **Appendix 8** for further detail.
- **Response Up (2,380 MW/h) and Response Down (2,180 MW/h)** requirements are **constant from 2025** onwards, both **within and between** years (as shown in the chart in bottom right). The requirements are constant from 2025 as Response requirements as dimensioned based on the N-1 contingency which does not change over the forward-looking modelling horizon.
- As well **Reserve Down (1,700 MW/h) requirements** are considered constant from 2025 onwards.
- However, because **Balancing Reserve Up requirements** are a function of the absolute hourly wind output, the average **Reserve Up demand increases across the modelling horizon**, in line with the projected **increase in GB wind capacity.**
- This is reflected in the increase in average Reserve Up requirement from **4,642 MW/h** in **2025** to **7,884 MW/h** in **2035.**

In the next slide we provide a breakdown of how the demand for ESO's AS reservation products are reflected in the four modelled products and illustrate how the demand for Reserve Up varies across hours in 2030.

Aggregation of ESO AS products for forward-looking assessment[1,2]



Average reserve requirements, 2022-35 (MW/h)[3]

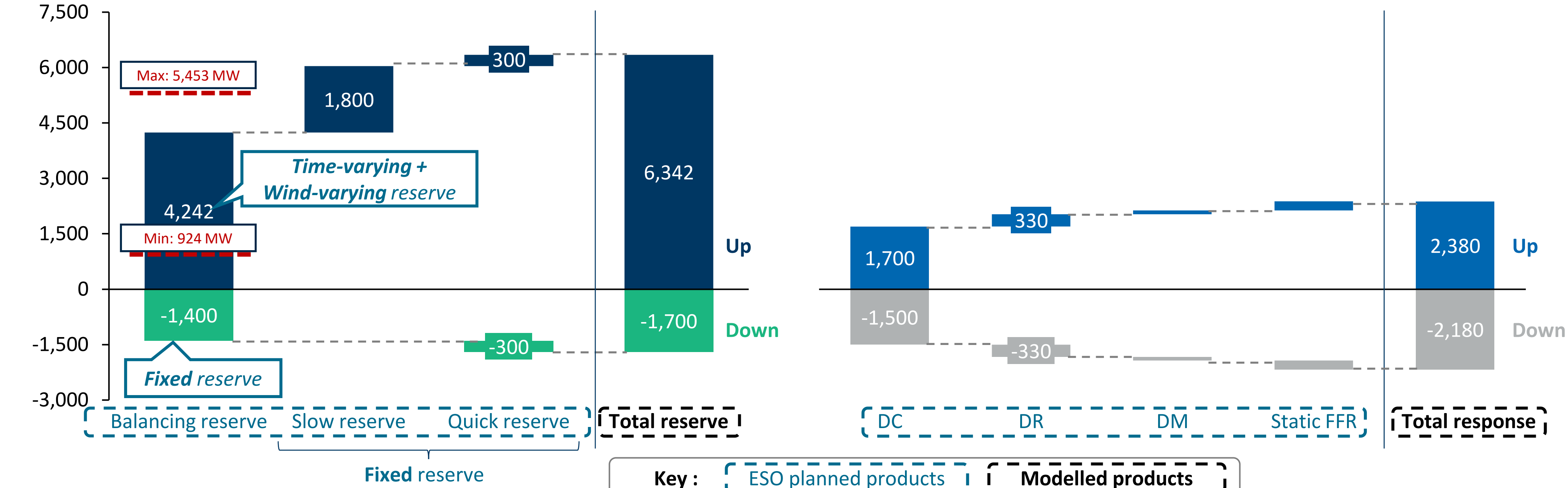


Notes: (1) The activation time of the different ESO AS reservation products differs. For model tractability, we group the different AS products into response (upwards/downwards) with a uniform activation speed and reserve (upwards/downwards) with a uniform activation time (see also Appendix 8); (2) We include Static FFR in both Response Up and Response Down for symmetry, as agreed with ESO. This does not imply that ESO intends to procure the Static FFR upwards and downwards in future; (3) For 2022 we used cleared volumes in AS markets, as explained further in Appendix 8.

There are large variations in Reserve Up demand within and between years due to changes in wind output

- The requirement for **Reserve Up varies within and between years**, driven by demand for **Balancing Reserve Up**, which includes time-varying and wind-varying components:
  i. The **time-varying reserve** component varies with the **hour, day** (i.e. working day vs non-working day) and **season** to reflect general uncertainties (which are assumed constant across the forward-looking modelling period); and
  ii. The **wind-varying reserve** component increases with the hourly **volume of wind generation** across GB, to account for wind forecasting uncertainty.[1] GB wind capacity increases from 42 GW in 2025 to 113 GW in 2035 (see slide 37), hence accounting for the increase in average Reserve Up requirements from 4.6 GW/h in 2025 to 7.9 GW/h in 2035, as set out on the previous slide.
- In the chart below, we set out the modelled demand for the Reserve and Response Up/Down products for the 2030 modelling year. Within the year, demand for Balancing Reserve Up fluctuates between 0.9 GW/h and 5.5 GW/h, with an average of 4.2 GW/h.

Reserve demand, up and down, average 2030, MW/h[2] | **Response demand, up and down, average 2030, MW/h[2]**



Notes: (1) In the modelling we have used the estimated wind generation, post BM for each hour – the same estimated volume is used in the sequential and co-optimised model. Details on the components of balancing reserve (both in terms of time-varying and wind-varying component) are provided in Appendix 9; (2) Based on sources provided by ESO and/or agreed with ESO.

AS participation reflects different technical characteristics and incentives facing different generation technologies

DA AS product technical characteristics[1]

	Reserve	Response
Activation time	10 min	1 sec
Product length	1 hr	1 hr
Duration activation	1 hr	15 min

Maximum participation in DA AS markets by technology (% of capacity)

Resource	Reserve[1]	Response[1]
Short-duration (1h) batteries	25%	75%
Long-duration (4h) batteries	50%	90%
Demand-side response	18%	N/A
Thermal[2]	90%	N/A
Pumped storage[3]	50%	N/A
Run-of-river hydro	100%	N/A
Wind	Reserve Down only	N/A
Solar	N/A	N/A

- We make various assumptions regarding AS product characteristics and participating technologies. These assumptions are made **purely for modelling purposes** and are not necessarily reflective of current or future eligibility to provide AS.
- In our modelling, the **product length for all DA AS products is 1 hour**. This is a simplification of the various length blocks that ESO's AS products are currently procured across.
- Response and reserve products have **different required activation times,** which restricts participation to certain technologies.
- **Response** in our modelling is assumed to **only** be provided by **short-duration (1h) and long-duration (4h) batteries,** as we assume a required activation time of 1 second.[4]
- We **further restrict response provision** in each hour to a **percentage of a battery's installed capacity** (see table in bottom left). This ensures that each battery has **sufficient energy in storage** to provide AS if activated (see Appendix 7 for further detail). We apply the **same restriction** in both **the sequential and co-optimised model for consistency.**
- We assume that **reserve** can be provided by a **broader asset base**, with **thermal, storage-based, wind, hydro and demand-side technologies all being eligible**, with varying levels of restriction applied (outlined further in Appendix 10).[5]
- Importantly, **wind** and **solar** units are assumed to offer all available generation into the **DA WS market** and therefore are **restricted from providing upward reserve.** Run-of-river hydro can offer up to its installed capacity in reserve up.

Notes: (1) Based on sources provided by ESO and/or agreed with ESO; (2) Here we include biomass, biomass CCS and waste; (3) In 2022, we set pumped storage max participation to 20% of capacity to better align the participation of pump storage in the modelling with observed participation in AS markets; (4) ESO's dynamic response services have historically been provided exclusively by batteries (ESO Market Roadmap 2022, p17); (5) These technologies have historically dominated reserve markets (i.e. MFR, FFR and STOR, ESO Market Roadmap 2022, p17, 24). See Appendix 8 for mapping of historical AS to modelled AS in 2022.

Modelling results

FTI CONSULTING

Across the historical and forward-looking assessments, co-optimisation lowers both system costs and consumer costs in all modelled years

	Key findings	Impact across WS and AS markets: System cost saving[1]	Impact across WS and AS markets: Consumer cost saving[2]
Historical assessment (2022)	▪ **Co-optimisation reduces AS prices by c.60%** and **WS prices by c.3%** relative to Sequential markets. Consumer cost savings are primarily driven by the reductions in WS prices, given the greater market size. ▪ **Consumer costs are reduced by over £1.4bn**. The **high WS prices** seen in **2022 due to the energy crisis** are reflected in the modelling of both market designs, and **improved allocation** of **price-setting marginal units** can **substantially impact** the **WS market**. ▪ **System cost savings** are an order of magnitude **lower**. **Improved scheduling of marginal units** reduces running costs, but generally **only for units at the margin** (e.g. baseload generators will be unaffected).	↓ **£178m**	↓ **£1,423m (3.1%)**

Forward-looking assessment (2025)	• In the **forward-looking** assessment, **Co-optimisation also reduces system and consumer costs** in **all modelled years**.
	• In **2030-35**, Co-optimisation enables **the large fleet of batteries** to be **scheduled across WS and AS** in a **coordinated manner**. Such improved scheduling **reduces WS prices** and in turn generates **significant savings for consumers** (though limited system cost savings).
	• **2025 is an outlier year** where the modelled **AS** markets are **very tight,** and the benefits of Co-optimisation are seen primarily in **system cost savings** (i.e. efficiency gains).
	    • Under Sequential, **generators make costly mistakes in forecasting uncertain WS prices** and incur **inefficient costs** (e.g. start-up), but also **self-cannibalise** the resulting **WS prices** which increases system costs relative to Co-optimisation.
	    • **Tight AS markets are better reflected in WS prices under Co-optimisation,** resulting in slightly higher WS prices (but lower AS prices) under Co-optimisation than under Sequential markets.
	    • The price signals in 2025 encourage entry into AS, hence we see a 'dynamic efficiency' effect.
		↓ **£831m**	↓ **£60m (0.2%)**

Forward-looking assessment (2030)		↓ **£189m**	↓ **£704m (4.1%)**
Forward-looking assessment (2035)		↓ **£137m**	↓ **£884m (5.6%)**

£3.4bn system cost saving (PV, 2025-35)

£4.9bn total consumer cost saving (PV, 2025-35)

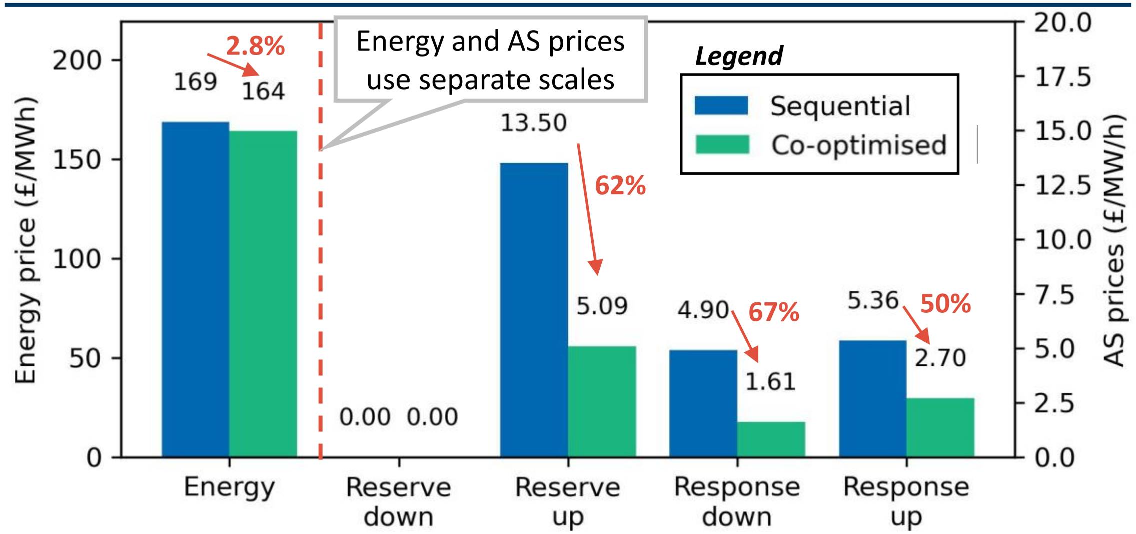
Notes: (1) We do not include % reductions for the system cost saving as this would require a complex valuation of the system costs (in the connected countries) associated with imports and exports to and from GB. The absolute system cost savings are comparable between Sequential and Co-optimised designs, as IC flows are broadly similar, and do not distort the evaluation; (2) We calculate consumer cost savings based on reductions in WS energy and AS prices. Accounting for offsetting changes in subsidy payments for generators on Contract for Difference schemes leads to a reduction in annual consumer cost savings of 7% in 2022, 11% in 2025, 63% in 2030 and 48% in 2035.

Historical assessment DA market: sequential vs co-optimised (2022)

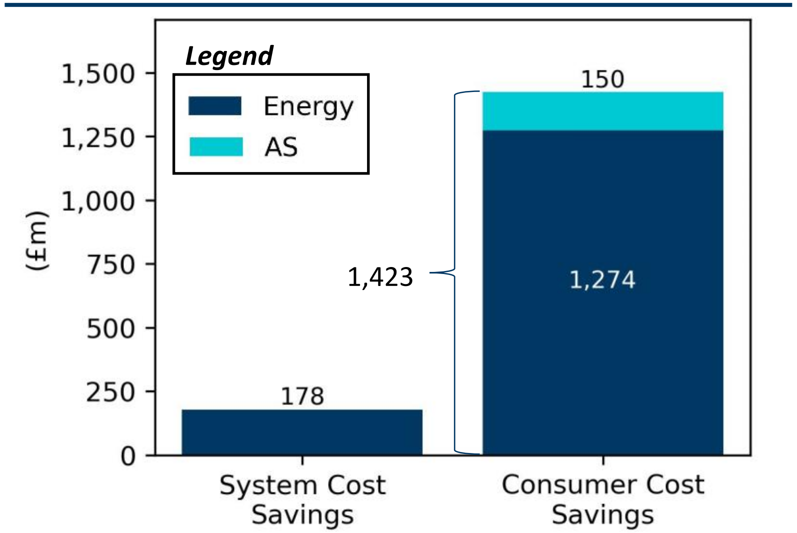
FTI CONSULTING

We estimate that DA co-optimisation of energy and AS in 2022 would have yielded consumer savings of £1.4bn and a £0.2bn reduction in system costs

Average energy prices (£/MWh) and AS reservation prices[1] (£/MW/h) 2022 – Sequential vs Co-optimised



System cost and consumer cost savings (£m) 2022 – Sequential vs Co-optimised

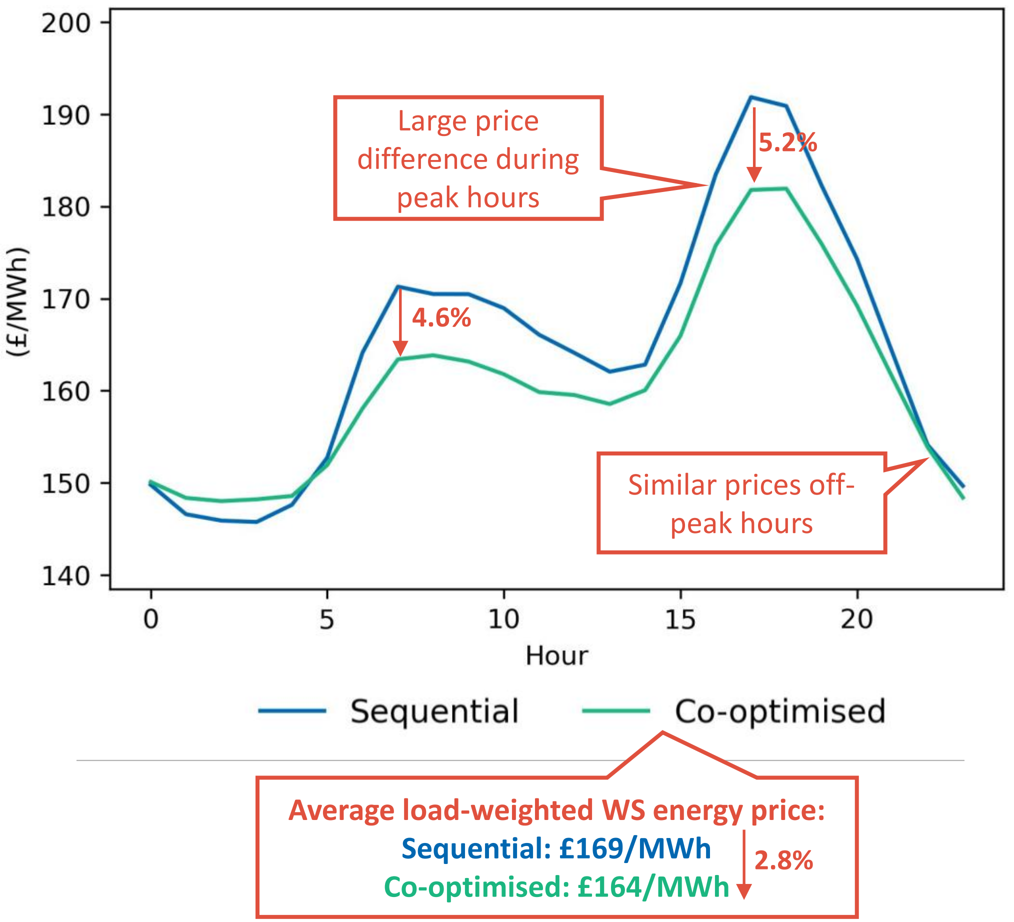


- We estimate that the **DA co-optimisation of energy and AS** would have generated **total consumer cost savings of £1.4bn in 2022[2,3] (a 3.1% reduction),** in the form of reduced DA WS energy and DA AS reservation consumer costs, relative to a sequential market design (shown in top right diagram).
- Even though AS reservation prices would have reduced significantly more (50-67% reduction) than WS energy prices (2.8% reduction) under co-optimisation (shown in the top left diagram), most **consumer cost savings would occur in the WS energy market (£1,274m in WS vs £150m in AS).[4]** This is due to the much higher energy demand (about 9x) and the absolute price levels of WS energy prices (in £/MWh) vs AS reservation prices (in £/MW/h).
- Consumer savings in the WS market arise because of a more efficient allocation of resources between AS and energy under co-optimisation. As such, **lower-cost resources can be utilised in the WS energy market** (compared to a sequential design) while the same AS reservation requirements are met.
- **More efficient scheduling would have let to a reduction in system costs of £178m (a 1.0% reduction)**. **System costs savings are an order of magnitude lower than consumer savings** because co-optimisation impacts mostly the allocation of units operating near the system margin. Difference in the resource allocation operating at the margin of the system can have a substantial impact on WS energy prices (set by the marginal unit) but less so on total system costs as the volume of units impacted by differences in the procurement of AS is low (e.g. baseload units in the energy market are not affected).
- AS prices and, consequently AS consumer costs, would have been significantly lower under co-optimisation because **market participants no longer would have had to rely on estimated opportunity costs** for being reserved in the AS market.

Notes: (1) See Appendix 11 for a comparison of modelled prices and historical prices; (2) Consumer cost savings are affected by the energy crisis, and therefore higher than in other modelled years; (3) Consumer cost savings do not account for subsidy payments, as explained on slide 42, footnote 2; (4) Uplifts for thermal units, representing £47m of additional AS costs under co-optimisation, are included in the Reserve Up average price and AS costs savings from co-optimisation.

Co-optimisation would have reduced wholesale energy prices most in peak hours, by up to 5% on average across 2022

Average daily load-weighted WS price pattern[1] (£/MWh) 2022 – Sequential vs Co-optimised

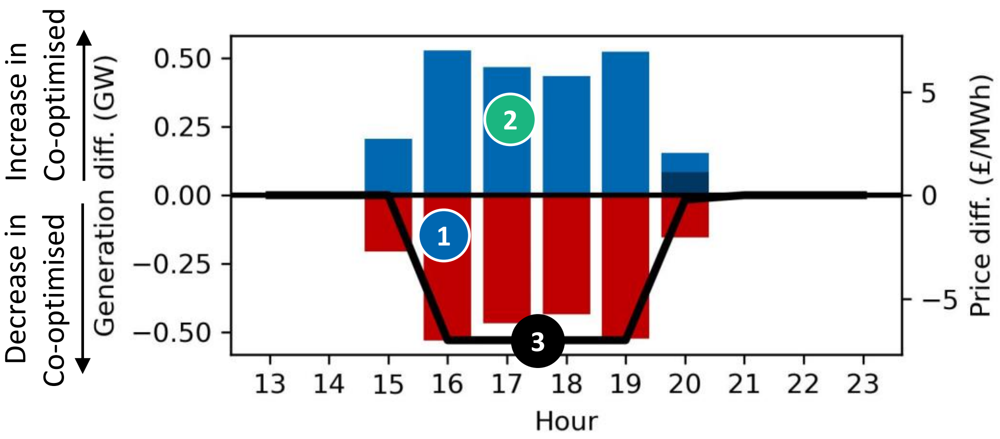


- **Co-optimisation** of energy (via the WS market) and AS would have **substantially lowered WS energy prices during peak hours.**
- As shown in the figure on the left, co-optimisation would have led to a **4.6% average WS price reduction during morning peaks** and **5.2% average reduction during evening peaks** relative to the sequential counterfactual.
  - Under a **sequential** market design, generators have imperfect information about the DA WS energy prices when bidding into the DA AS markets.
  - This imperfect information can lead some inframarginal units to underestimate WS energy prices at DA stage and (inefficiently) commit to providing reserves instead of generating the next day, thereby requiring higher marginal cost units to switch on to generate to meet WS demand and increasing WS energy prices.[2]
  - This WS energy price impact is most pronounced during price peaks because demand is highest (and most generators will already be generating) so typically a high-cost peaker unit needs to be started up to generate or a battery with a high opportunity cost needs to discharge.
  - However, under **co-optimisation** there are no (imperfect) opportunity cost estimations required because the co-optimisation algorithm has visibility of the costs and technical characteristics of all units which results in a least-cost WS energy schedule that complies with the AS requirements.
  - During off-peak hours, average energy prices are very similar in the sequential and co-optimised models because the inefficient allocation of resources between AS provision and energy scheduling has a less profound energy price impact (i.e. there are enough low-cost generators available that are not scheduled)
- In addition to reducing WS energy prices, co-optimisation also would have **reduced price volatility**, as seen in the reduced peak prices.

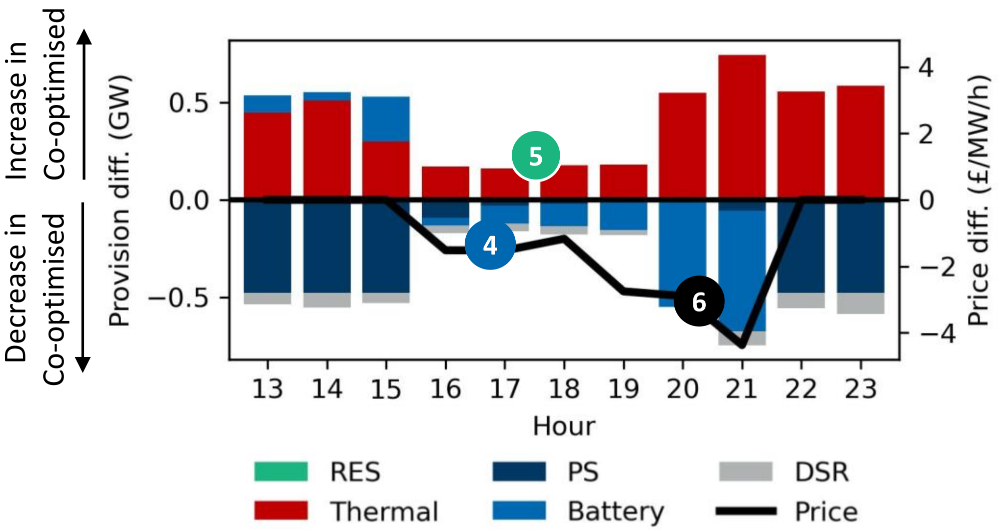
Notes: (1) Average prices do not include uplifts in the WS market (representing about 0.5% of all WS energy costs), see Appendix 12 for more detail; (2) See slide 25 of the qualitative report, published in parallel, for the mechanics behind this behaviour.

More efficient scheduling of AS would have enabled co-optimisation to lower WS energy prices

Difference in the WS energy schedule (GW) and price (£/MWh) – Co-optimised less Sequential, 13:00 – 23:00 3rd Sept 2022



Difference in Reserve Up AS provision (MW) and price (£/MW/h) – Co-optimised less Sequential, 13:00 – 23:00 3rd Sept 2022



Note: (1) Although there will still be imperfections (including imbalances and constraint issues), which will be adjusted for in RT.

On this sample day, the diagrams opposite show how more efficient scheduling under Co-optimisation would have facilitated significant **WS energy price reductions relative to the Sequential model**.

- The lowest-cost of meeting WS demand on this sample day is by **maximising the scheduling battery injection during peak hours in the WS market** (4-7pm during this day), and hence limiting the batteries in Reserve Up during those hours. This is what happens under co-optimisation.
- Instead, under the Sequential design, due to imperfect forecasts of the WS energy price, **more batteries are being cleared in the Reserve Up market during peak hours.**

WS energy market:

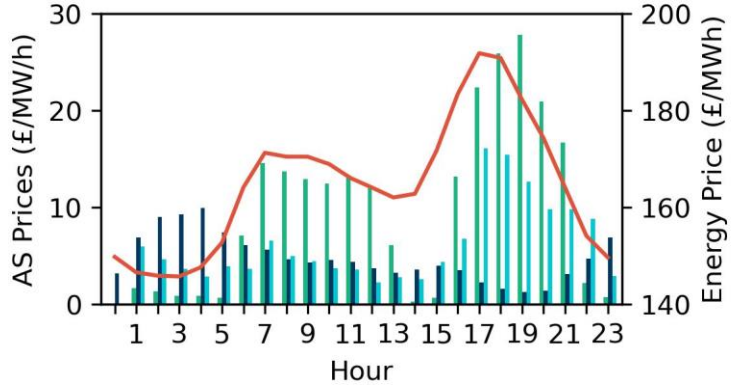
1. Under the Sequential design, more thermal units need to be scheduled in WS, which raises WS price relative to Co-optimised;
2. By contrast, under Co-optimisation, more battery injection is scheduled in WS between 4-7pm replacing thermal that is scheduled under Sequential.
3. **As a result, from 4-7pm, the WS energy price are c. £7/MWh higher in the Sequential model.**

Reserve up AS market:

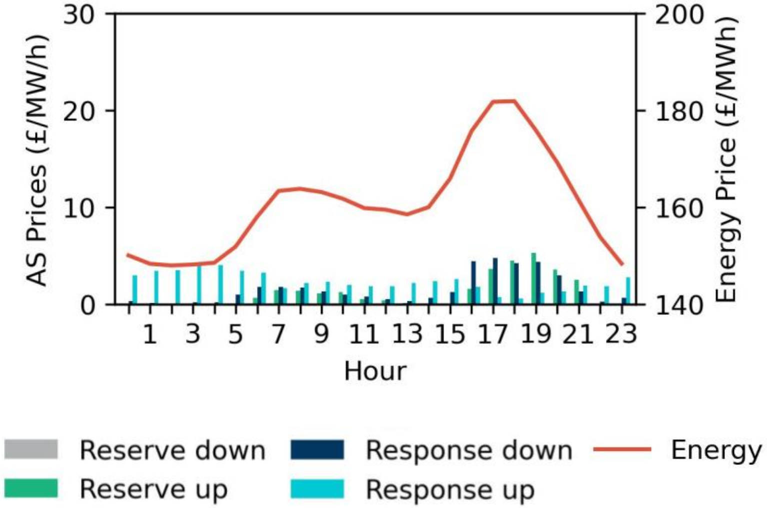
4. Under Sequential design, imperfect WS price forecasts lead to higher volume of batteries being cleared in the Reserve Up market rather than thermal generation.
5. By contrast, in the Co-optimised model there is no imperfect forecasting[1]; hence thermal generation is reserved instead, and valuable battery injection capacity is freed up to be scheduled in the WS.
6. **Reserve Up prices between 7-9pm are substantially higher under the Sequential design:** batteries expected peak prices to occur before (which also happened) and priced in opportunity costs for keeping energy in storage. This pressure on the AS price is less pronounced under Co-optimised due to overall better scheduling of batteries across the day.

AS prices would have been lower and less volatile under co-optimisation due to more efficient resource allocation between the DA energy and AS markets

Avg hourly AS prices (£/MW/h) 2022 – Sequential



Avg hourly AS prices[1] (£/MW/h) 2022 – Co-optimised



- **AS prices would have been significantly lower under co-optimisation.** The main reasons are:
  - Co-optimisation enables **lower-cost reservation of thermal units' capacity** compared to the Sequential model as in the latter commitment decisions are uncertain when units bid for AS provision.
  - Co-optimisation enables **batteries to be scheduled more efficiently**, as the central algorithm has an overview of the entire battery fleet, e.g. battery discharging can be spread across the peak period which as an indirect effect also leads to less spiky and overall lower AS prices.
- Under the **sequential model:**
  - prices for **Reserve Up** and **Response Up** are highest during and after WS peak hours:
    - **thermal units**, active in **Reserves Up**, and **batteries**, active in both upward AS services, have a high opportunity cost during peak hours as they must forgo generating/ discharging during those hours.
    - **batteries** also have a high opportunity cost just after peak hours as being reserved in those hours implies foregoing discharging in peak hours (i.e. to keep energy stored).
    - prices for **Response Down** are highest in off-peak hours: batteries must forgo charging in those hours when being reserved downwards. **Response Down** prices also show an uptick before peak hours due to the missed-out revenues from not being able to be fully charged.
- Under the **co-optimised model:**
  - **Reserve Up** prices are highest during the peak hours, reflecting mostly the opportunity costs of **thermal units**.
  - **Response Up** prices are highest early in the day which reflects not all **batteries** yet being fully charged after discharging the night before.
  - **Response Down** prices are highest before the peak and (even) during the peak due to part of the battery fleet being relatively full as it is valuable (from a system perspective) to discharge during or just after the peak.
- **Reserve Down** prices are always zero under both models, as many generators can provide Reserve Down at zero opportunity cost.

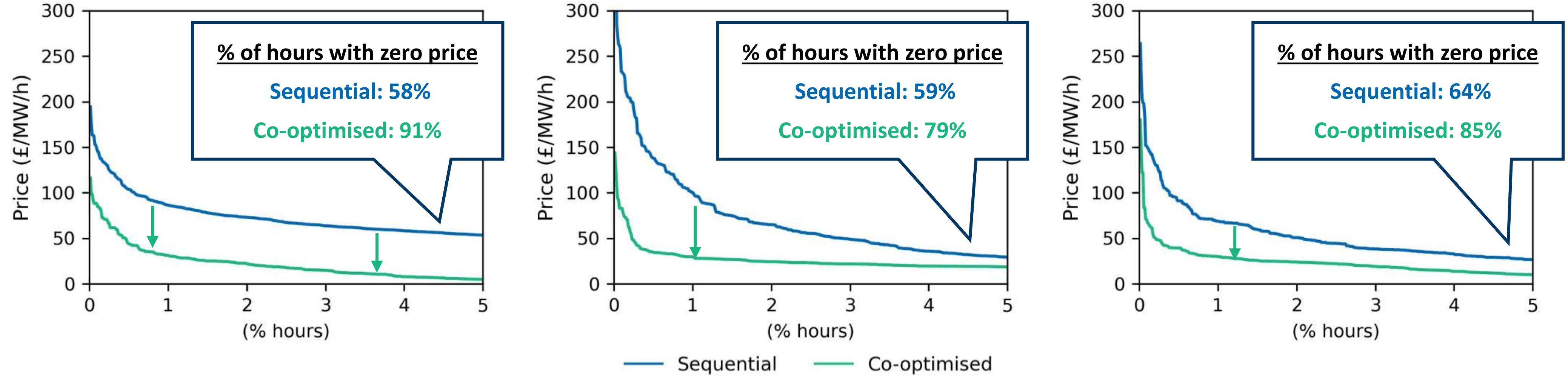
Note: (1) Uplifts represent c.67% of revenues in the Reserve Up market and are not internalised in the prices shown on this chart as they cannot be allocated to a specific hour.

Co-optimisation would have "flattened" AS price distributions: reducing AS peak prices and increasing the frequency AS markets clear at a price of zero

Hourly Reserve Up prices in descending order (£/MW/h) 2022 – Co-optimised vs Sequential

Hourly Response Up prices in descending order (£/MW/h) 2022 – Co-optimised vs Sequential

Hourly Response Down prices in descending order (£/MW/h) 2022 – Co-optimised vs Sequential

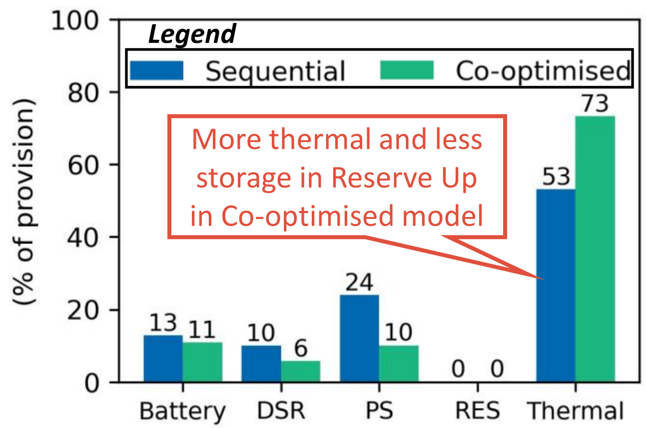


- **Co-optimisation would have reduced peak AS prices across AS markets**, relative to a **Sequential market design**. For example, in 2022, the peak **Reserve Up** price would have fallen from c. £200/MW/h in the Sequential model to c. £120/MW/h in the co-optimised model.[1]
  - Co-optimisation prevents AS prices from being driven upwards by **imperfect opportunity cost estimations.**
  - Due to the "central coordination" under **co-optimisation** resources are allocated more efficiently across markets and across the different hours of the day with the aim to minimise total system costs – under the sequential model, **individual market participants lack visibility of the actions of other market participants** and tend to all submit high bids/offers in the AS market during a few concentrated hours.
- Co-optimisation **would have increased the frequency of AS markets clearing at a price of zero** (from c. 60% of hours in each market in sequential to c. 80-90% in co-optimised). This is driven by the following difference between the two designs:
  - Under **co-optimisation**, hours when certain resources have **no opportunity costs can be directly identified without uncertainty**.
  - Under the **sequential model**, different resources have **slightly different DA energy price forecasts** implying that the hours in which resources estimate to face no opportunity costs (e.g. a battery after the peak hours providing response down) will be different (e.g. two batteries could expect the peak period to end during a different hour). As such, in **most hours at least a subset of resources will be offering non-zero AS bids** (while their actual opportunity cost would be zero under perfect information). Consequently, the **likelihood AS markets can fulfil their entire demand with zero-priced AS bids is lower** resulting in an **increased frequency in non-zero AS prices being observed**.

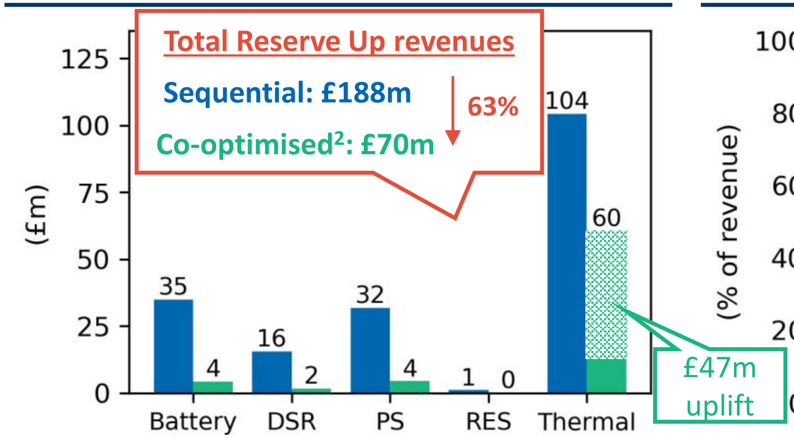
Note: (1) Reserve down has a price of zero in every hour, hence we do not show a chart for it above.

Under co-optimisation more thermal generation and less pump storage would have cleared in the AS Reserve Up market compared to the sequential model

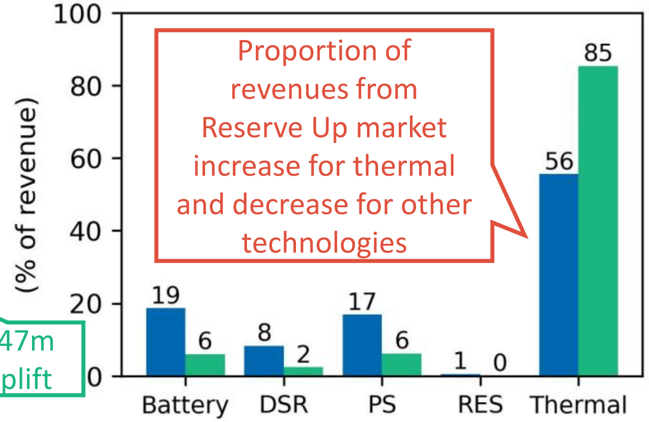
Reserve Up provision (%) 2022 – Sequential vs Co-optimised



Reserve Up revenues (£m) 2022 – Sequential vs Co-optimised



Reserve Up revenues (% of total) 2022 – Sequential vs Co-optimised



- The left chart shows that under Co-optimisation more thermal would have been reserved for Reserve Up.[1,2] Co-optimisation enables the combined value of a thermal unit in both AS and WS markets to be efficiently reflected in market outcomes. Under a Sequential design commitment decisions are uncertain when having to bid into the Reserve Up market and mistakes in the assumed WS commitment decision can be costly. Under **Co-optimisation,** improved co-ordination means that **more thermal units are committed to provide Reserve Up which eventually leads to lower total scheduling costs**.
- Greater provision of Reserve Up by thermal in Co-optimised model (compared to Sequential) would have meant that **less Reserve Up were provided by pump storage and by batteries and DSR**.
- The middle chart shows that under both models, **overall revenues earned under Co-optimisation would have been significantly reduced** relative to the Sequential model due to the reduced AS prices.
- **Thermal generators earn the largest share of revenues** from Reserve Up. In the Co-optimised model, an important share of the revenues from thermal units would have originated from uplift payments.[3] Under the Sequential model, **all costs of thermal generators** (e.g. starting-up earlier/later and running at MSL at a loss) **are priced into their AS bids** and hence reflected in AS prices, leading to also higher revenues for the other technologies.
- The right chart shows the proportion of revenues from Reserve Up market per technology class. The proportion of total revenues earned by thermal generation in the Reserve up market would have decreased under Co-optimisation while it would have decreased for the other technology classes.

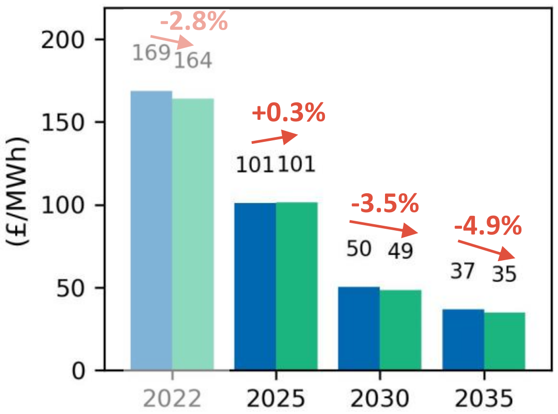
Notes: (1) Increased thermal reservations (and hence lower scheduling of thermal for WS) under Co-optimisation would lead to carbon emissions savings at DA stage, relative to the Sequential model. Total carbon savings would also need to account for activations in RT and for the emissions impact of changes in net imports across market designs, but this is not considered in this report; (2) We only present results for Reserve Up since Response services were solely provided by 1h batteries in 2022 and Reserve down prices are always zero (in both models), hence there are more resources available that can provide Reserve down for free than the Reserve down requirement; (3) Uplifts are required as the AS price (reflecting the increase in system cost when procuring one additional unit of reserve/response) might not be sufficient for all thermal units to recover all their operating costs.

Forward-looking results DA national market: sequential vs co-optimised (2025-2035)

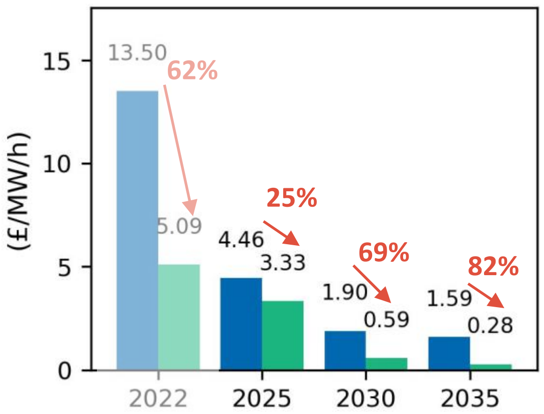
FTI CONSULTING

Co-optimisation substantially reduces AS reservation prices in all modelled years, and, more importantly, reduces WS energy prices in 2030 and 2035

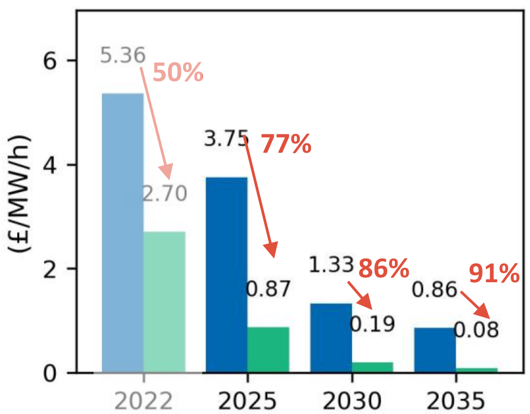
Average WS energy prices (£/MWh)
All years – Sequential vs Co-optimised



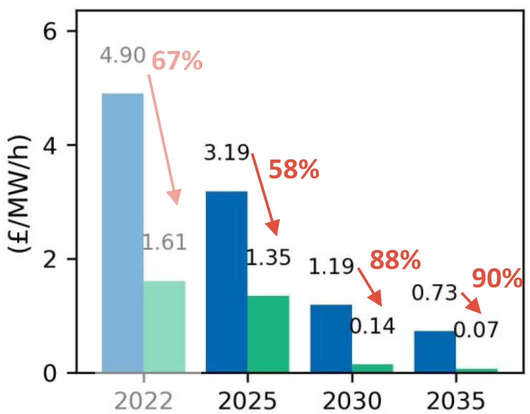
Average Reserve Up prices[1] (£/MW/h)
All years – Sequential vs Co-optimised



Average Response Up prices (£/MW/h)
All years – Sequential vs Co-optimised



Average Response Down prices (£/MW/h)
All years – Sequential vs Co-optimised



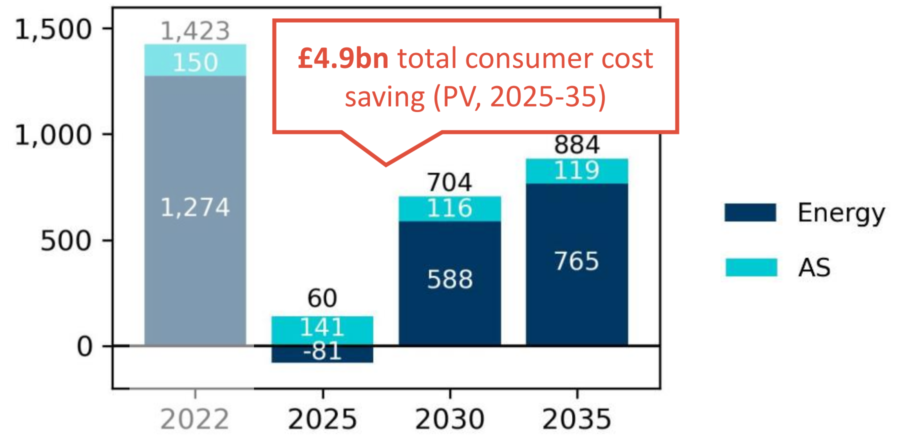
Sequential   Co-optimised

- Under both models **average WS energy prices decrease over the modelling horizon** as shown in the top left chart. The major driver of the downward trend in WS energy prices is that while electricity demand grows over the modelling horizon it is outpaced by investment in low marginal cost RES (see slide 37).[2]
- Even though **procured AS volumes increase across the modelling horizon** (see slide 38), **AS prices fall under both models over time,** driven by a steep increase in the volume of batteries in the system.
- **Under Co-optimisation WS prices reduce** in 2030 (-3.5%) and even more so in 2035 (-4.9%) while in 2025 **WS prices slightly increase (+0.3%).**
- The reason for the slightly counter-intuitive 2025 result is that, as explained in slide 26, WS prices are generally expected to increase under a Sequential market.[3] This is the case in 2022, 2030 and 2035, but does not always hold:
  1. **High AS prices are reflected in WS energy prices under Co-optimisation** which can lead to higher WS prices relative to under Sequential markets (where such feedback loop is not in place).
  2. **Inefficient scheduling can lead to having a lower cost marginal unit scheduled in the WS under Sequential markets** which leads to lower WS relative to under Co-optimisation. This can occur when high marginal cost generators under-estimate their opportunity costs as shown in slides 54-55.
- **AS reservation prices decrease in all years and over all products under Co-optimisation** with the largest decreases in 2030 and 2035 when the reliance on batteries in AS markets is greatest (see slide 57).

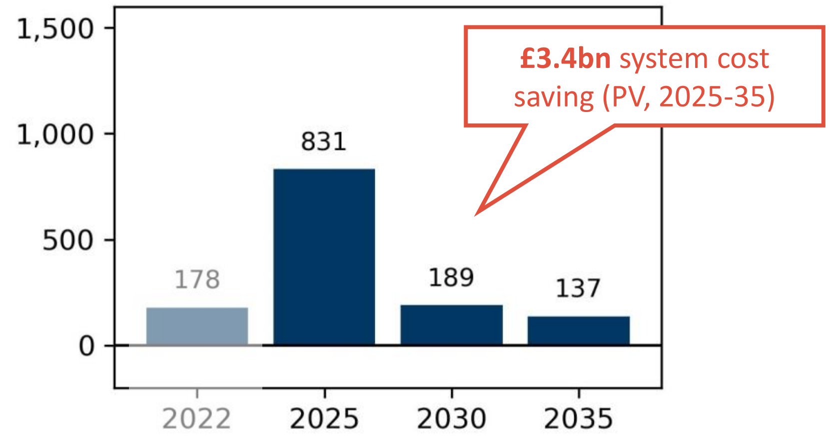
Notes: (1) Uplifts for thermal units are included in the Reserve Up average price; (2) In addition, assumed natural gas prices also decrease over the modelling horizon; (3) See also the worked-out example in slide 46.

We estimate that co-optimisation of energy and AS would generate consumer savings of £4.9bn from 2025-35, relative to a sequential market design

Consumer cost savings All years – Co-optimised vs Sequential (£m)



System cost savings All years – Co-optimised vs Sequential (£m)

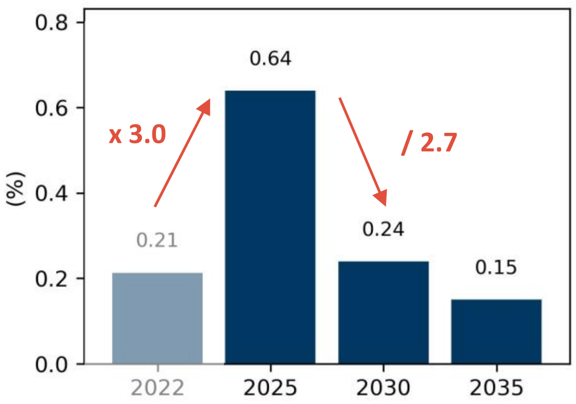


- Differences in WS energy and AS prices result in **total consumer cost savings[1] of £4.9bn from 2025 to 2035** (PV, left diagram) under DA Co-optimisation relative to a Sequential market design. These **cost savings are most significant in the WS energy market** because relatively modest decreases in WS prices result in energy costs savings that can be an order of magnitude larger than AS costs savings due to higher volume of demand in WS markets.
- Co-optimisation reduces total consumer costs in all modelled years and these cost **savings increase between 2025 and 2035** (reaching £884m in 2035, 5.6% of total WS and AS consumer costs). **Consumer savings are lowest in 2025** due to a small increase in WS prices under Co-optimisation.
- More efficient scheduling under Co-optimisation leads to **system cost savings of £3.4bn from 2025 to 2035** (PV, right diagram). The system cost saving is highest in 2025 (£831m). Limited consumer savings but high system costs savings imply **substantial producer gains in 2025 under Co-optimisation.** In other words, in 2025 co-optimisation provides stronger signals for new entry (i.e. current Sequential market design in 2025 underfunds).
- The **system cost savings are explained by an improved resource allocation between AS and WS markets**. The reason the system costs savings in 2025 are the largest is driven in part by the behaviour of thermal units and in part by AS markets being the tightest (see also next slide):
  - In 2025, uncertainty around the WS price forecast increases under a Sequential design as minor differences in the allocation of resources between the AS and WS markets can have a strong impact on the WS price (see slides 54-55).
  - This leads to some resources (e.g. thermal) making costly mistakes in 2025 by over-estimating the WS prices when they bid in the AS markets, which leads them to under-price themselves in Reserve Up. By subsequently bidding their MSL in the WS, they 'self-cannibalise' the WS prices, which makes their original AS bid uneconomic. **Lack of feedback loop from WS to AS drives economically inconsistent outcomes (and efficiency losses) under the Sequential design**. In turn, Co-optimisation addresses this inefficiency and hence helps reduce system costs to a significant extent.[2]

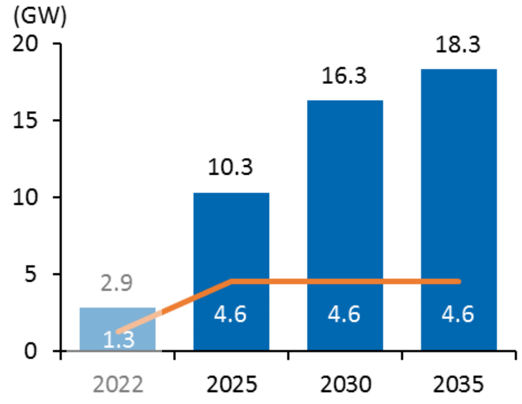
Notes: (1) Consumer cost savings do not account for subsidy payments, as explained on slide 42, footnote 2; (2) If, for example, there were more batteries able to provide AS, this would make WS energy prices less dependent on AS market outcomes, and fewer thermal resources would make mistakes in their WS energy price forecasts.

AS markets are tightest in 2025, as reserve requirements increase significantly, and total battery capacity is still relatively limited compared to later years

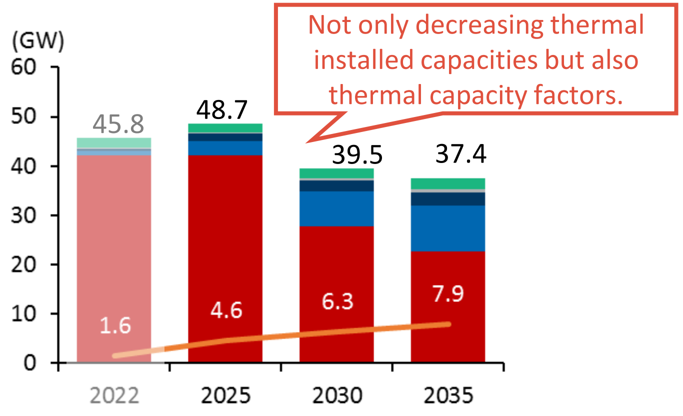
Total AS reservation market costs over WS energy market cost-to-load (%) All years – Co-optimised



Average Total Response requirement and eligible battery capacity[1] (GW) All years



Average Reserve Up requirement and available capacity by technology[1] (GW) All years

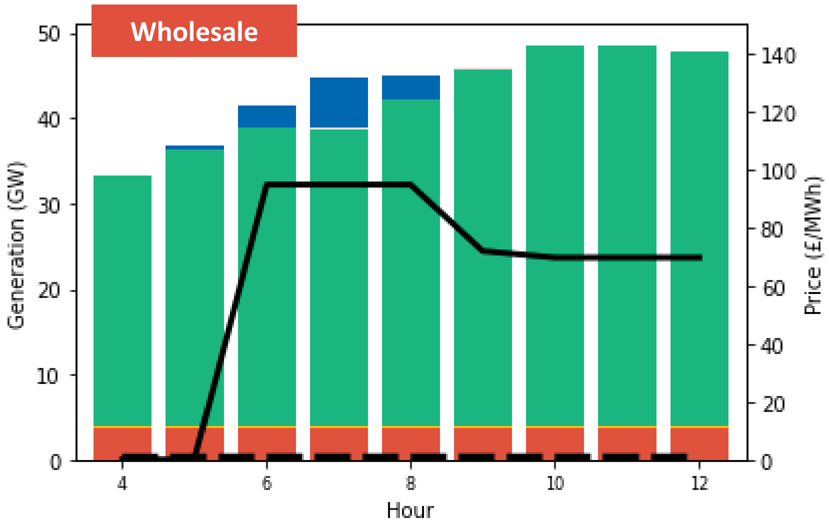


- **Relative AS market reservation costs peak in 2025** at 0.64% of total WS energy costs (left diagram) under Co-optimisation, **c. 3x higher than other years**.
- On the **demand side**, total Response requirements more than treble from 1.3 GW/h in 2022 to 4.6 GW/h in 2025 (middle diagram), whilst the average Reserve Up requirement also roughly trebles from 1.6 GW/h in 2022 to 4.6 GW/h in 2025 (right diagram). Total Response requirements remain constant from 2025 to 2035, as the largest possible lost load remains constant, whilst the average Reserve Up requirement increases from 4.6 GW/h in 2025 to 7.9 GW/h in 2035 due to increasing wind generation.
- On the **supply-side**, the installed capacity of batteries eligible to provide Response[1] increases throughout the modelling period from 2.9 GW in 2022 to 18.3 GW in 2035. However, thermal capacity for Reserve Up decreases over the same period. In addition, thermal capacity load factors in the WS market decrease over the modelling horizon (from 37% in 2022, 29% in 2025, 23% in 2030 to 20% in 2035) which has implications for their provision of AS: unlike more flexible technologies (e.g. batteries), providing Reserve Up by an (otherwise) out-of-merit thermal unit requires start-up costs to be incurred.
- The overall **supply-demand balance** in Response markets is very similar in 2022 and 2025 (total response requirement is c. 45% of eligible battery capacity in both years), but batteries also play a larger role in the Reserve Up market in 2025 (see slide 58). In 2030 and 2035 battery capacity increases substantially while the Response requirements are the same as in 2025; Reserve Up requirements increase but more batteries are able to be active in Reserve Up markets. **These dynamics make 2025 the tightest year for AS.**
- Tight AS markets in 2025 reduce the consumer WS energy cost savings from Co-optimisation, relative to other years. **Co-optimisation enables high AS prices to feedback into WS prices. Imperfect information** can lead in such context to lower WS prices as explained in slide 26 and Appendix 3.

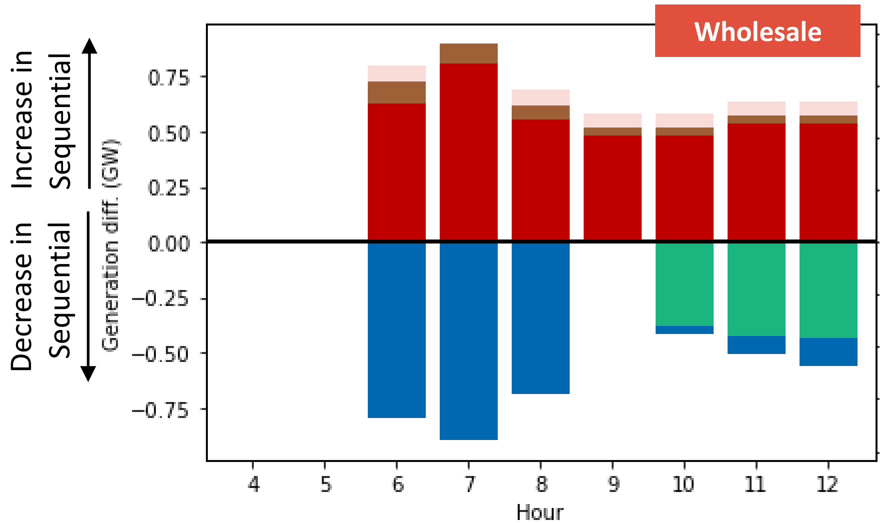
Note: (1) Eligible battery capacity = installed capacity of all batteries able to AS (i.e. only electrochemical batteries, no CAES, LAES, domestic or V2G batteries). Available capacity = total installed capacity, de-rated by AS participation factor (see slide 40). The RES category contains run-of-river hydro. Solar and wind are not assumed to provide Reserve Up.

With sequential procurement, in some hours imperfect forecasting leads to lower prices for consumers, due to a misallocation of units between markets…

Co-optimised WS schedule (GW), with co-optimised and seq wholesale price (£/MWh) - 25th Oct 2025



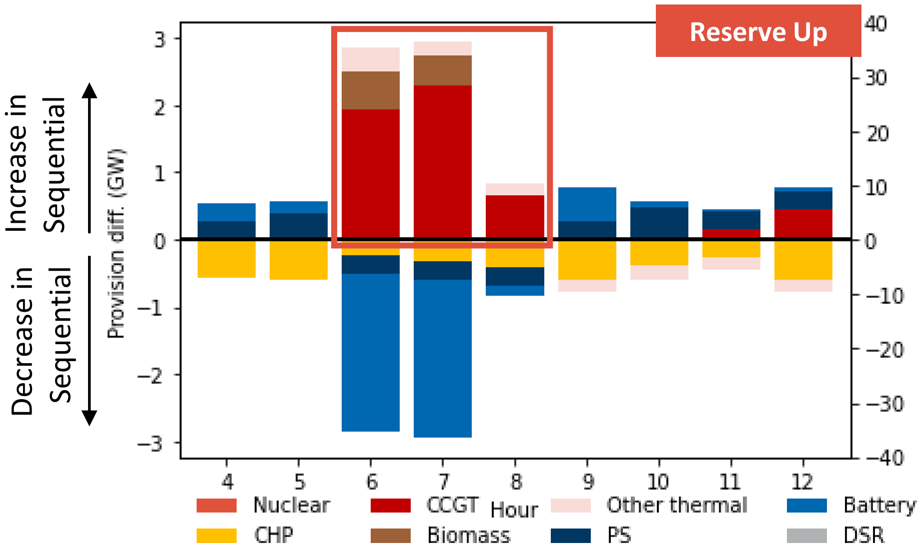
Difference in WS schedule (GW) – Sequential less Co-optimised, 25th Oct 2025



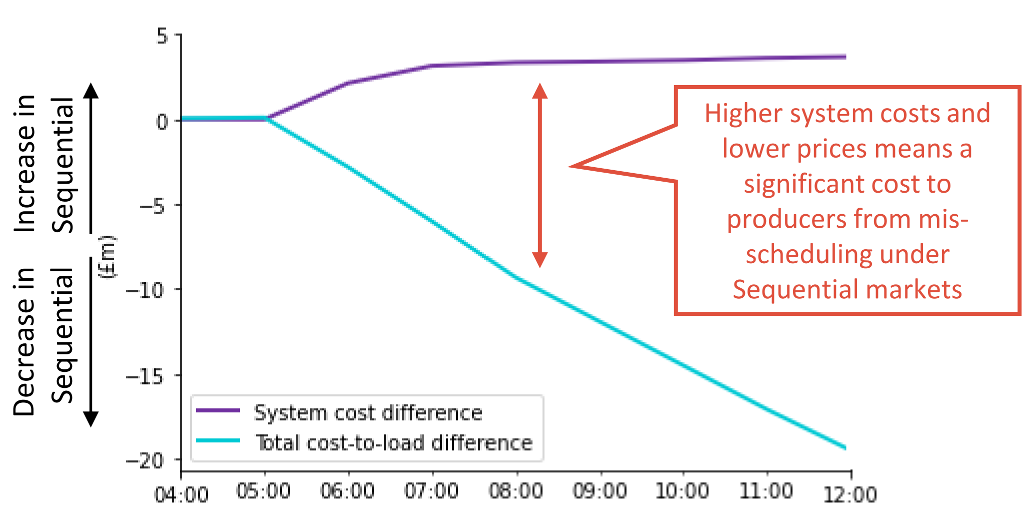
- **Imperfect bidding** under **Sequential markets** can sometimes cause WS markets to clear at **lower prices**, despite **higher overall system costs.**
- In the **top-left chart, for a high-RES day in 2025**, **under co-optimisation, RES** and **nuclear** can initially **fully meet the WM demand**, setting the **WS price** at **near-zero**. However, **from 6am onwards**, while there is almost enough RES to meet demand, **batteries are also required**, pushing the wholesale price higher (the **value of stored energy** on the day is **high** because (i) **AS markets are tight**; and (ii) **evening peak prices are high**).
- By contrast, **in sequential** the **wholesale price stays at near-zero** over the same period.
- The **top-right chart** shows the **difference in generation** between the **Sequential** (positive) and **Co-optimised** (negative) models. Under Sequential markets there is additional CCGT generation (driven by a costly mis-forecasting of wholesale prices by CCGTs– see next slide), which displaces the price-setting **battery injection** in the morning peak and causes a reduction in **renewable generation mid-morning**.
- Importantly, there is **still some battery injection in Sequential** in the morning peak, despite near-zero wholesale prices. These batteries have **cleared in the downward response AS markets** and must retain **sufficient spare capacity to provide the service**. As a result, **units with too much energy in storage** to comply with their downward AS commitments are **willing to generate at very low prices** and **do so in the morning peak, setting the wholesale price** at near-zero (but slightly higher than RES units would).

…with AS market commitments requiring units to bid very low prices in the WS market and depressing clearing prices, despite higher system costs

Difference in Reserve Up AS provision (MW) – Sequential less Co-optimised, 25th Oct 2025



Difference in system cost and cost-to-load (£m) – Sequential less Co-optimised, 25th Oct 2025

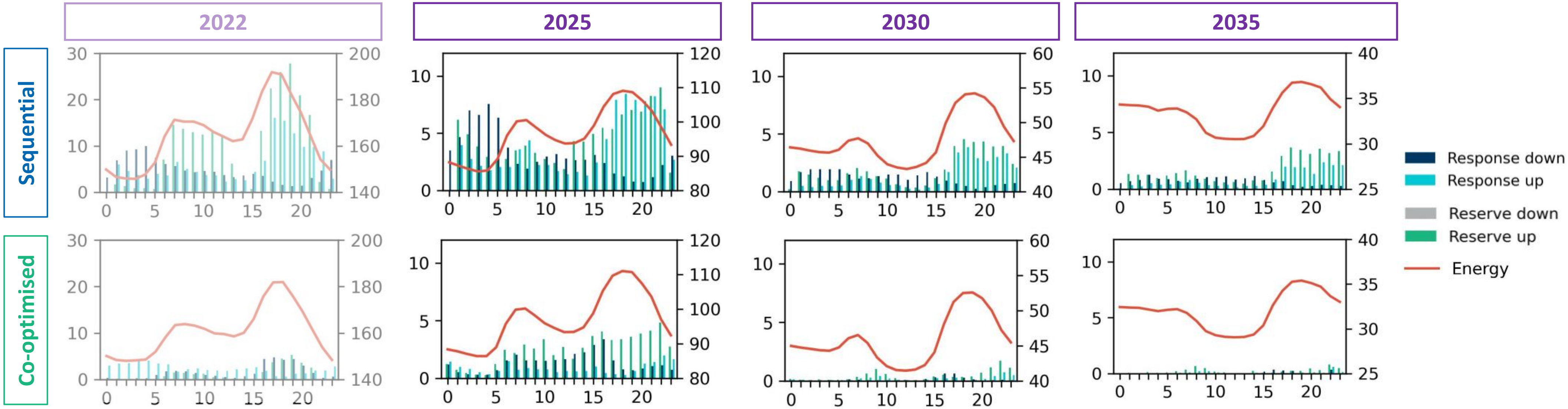


- The increased CCGT generation under Sequential markets is driven by a **costly mis-forecasting of WS prices by CCGTs**. Anticipating rising prices in the morning peak (incorrectly so – these price rises do not materialise under the Sequential design), the CCGTs **expect to already be brought online for WS provision** and so are **willing to offer into AS markets at a relatively low price (i.e. the cost of anticipated start-up).**
- The impact of this can be seen in the **Reserve Up market**, shown in the **top-left chart**. Under Sequential markets, in the morning peak **2.3 GW more CCGT capacity** clears in **Reserve Up** than under Co-optimised, mostly displacing batteries. Under Co-optimisation**, no CCGTs are used for AS or in the WS schedule** in the highlighted hours (meaning no CCGT running costs are incurred).
- Because they successfully clear in the Sequential AS market, the **CCGTs are then required to generate at least at MSL** in the **wholesale market** to meet their **AS commitments**. For **generation up to MSL-level**, the CCGTs become **price-takers** and **crowd out battery and renewable generation.**[1]
- An inefficient **allocation of assets across markets under Sequential markets results in higher system costs** (right-hand chart). Consumers benefit through **lower WS prices**, to the detriment of producers who: (i) face **higher scheduling costs** than under Co-optimisation; and (ii) receive a **lower WS price.**[2]
- This dynamic is observed more in **2025**, when most **battery capacity is required in AS markets**, particularly Response. Under **Co-optimisation**, this **scarcity** (and the value of batteries) is **reflected in both WS and AS prices**. In **Sequential**, the **crowding out** of batteries by **higher cost units lowers WS prices.**

Notes: (1) While no thermal is reserved between 9-10am, some thermal remains online in the WS market as this is a lower cost solution then to shut down and start up later in the day; (2) In the presented modelling we use 'static' AS bidding methodology. Units do not adjust their AS bidding in response to the strategy's profitability on previous days. One would expect thermal units to learn from this experience, increasing AS offer prices on future high-RES days.

AS prices follow similar daily patterns in both models, although prices are significantly lower in the Co-optimised model and fall over time

Avg hourly AS prices (£/MW/h, left axis) and avg hourly WS price (£/MWh, right axis) All years – Sequential and Co-optimised



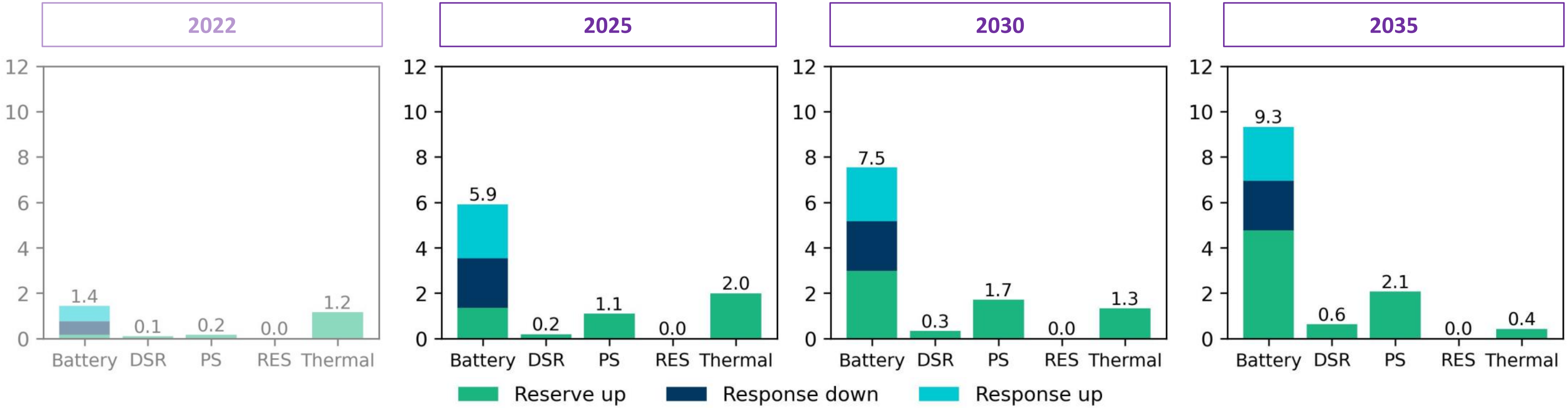
Note: Hour of day along x-axis.

- Average AS prices are **lower in Co-optimised** than Sequential and **fall over time** as the installed battery capacity increases relative to the AS requirements (see slide 53). Across all modelled years, we see the general pattern that:
  - Prices of **upward AS** are **highest during WS peak hours under a Sequential market**, as due to a lack of central coordination, market participants estimate their opportunity costs to be highest during WS peak hours. **Under Co-optimisation**, especially in the later years, the bottleneck to providing AS is not the power output of batteries but rather state of charge management, i.e. **upward AS prices are highest just after the WS peak hours** as batteries need to ensure enough remaining state of charge to provide AS.
  - **Response down** prices are **highest when WS prices are low under the sequential model;** batteries face an opportunity cost of not being able to recharge during these hours. **Under Co-optimisation downwards AS prices are lowest before the WS price peak**, again reflecting the scarcity in battery state of charge that leads to an opportunity cost for batteries when providing AS.
- In the later years, there are **many zero-priced hours for AS under Co-optimisation.** Increased battery capacity combined with a AS and WS schedule that is centrally coordinated means there is sufficient capacity of batteries that can provide AS with no opportunity cost to meet the AS requirements.[1]

Note: (1) This is, in part, a reflection of the exogenous modelling assumptions regarding the deployment of batteries and other assets across GB. In practice, the deployment of batteries might, in part, depend on the expected AS revenues (alongside other sources of revenues). If those are expected to be relatively low, the battery capacity projections in this modelling could be different. We have not sought to evaluate this dynamic in our assessment.

Batteries play an increasingly important role in AS markets over the modelling horizon

All AS (excl. Reserve down) average provision by technology (GW/h) All years – Co-optimised

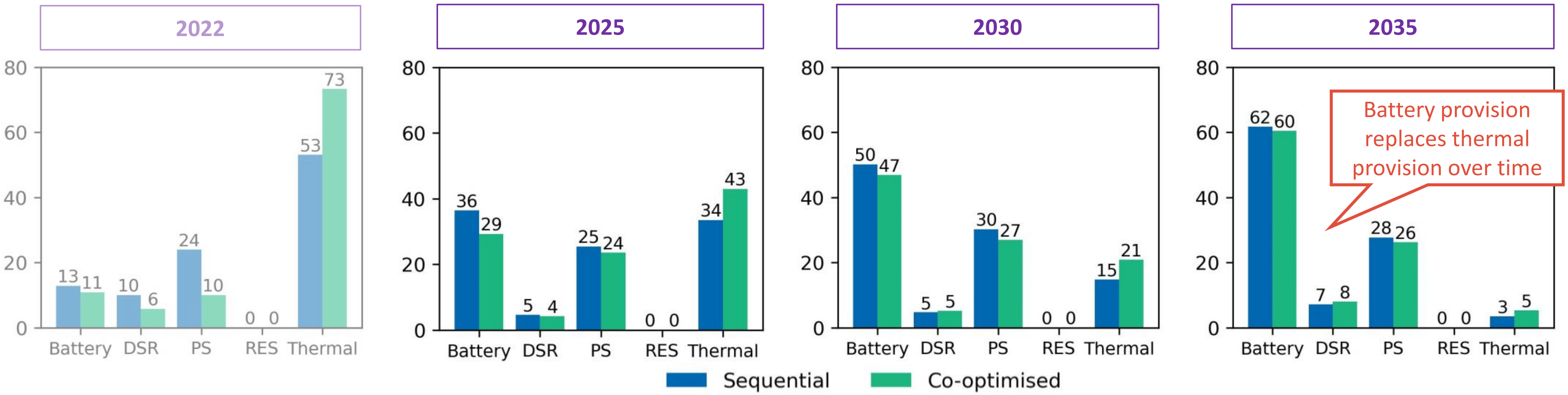


- **Battery provision of all AS** (excl. Reserve down) **increases over the modelling period** from an average of 1.4 GW/h in 2022 (50% of total) to 9.3 GW/h in 2035 (75% of total).
  - Due to the one second activation[1] required to provide Response, batteries are the only technology able to provide this service, hence batteries play a large role in AS markets in all modelled years.
  - Total response requirements then increase from 1.3 GW/h in 2022 to 4.6 GW/h in 2025 (see slide 53), causing battery AS provision to increase significantly.
  - After 2025, response requirements do not change, but batteries provide an increasingly large portion of the growing Reserve Up market, displacing thermal provision.
- **Thermal** Reserve Up provision increases from 1.2 GW/h in 2022 to 2.0 GW/h in 2025 as Reserve Up requirements increase. However, thermal AS provision then decreases in later modelled years as thermal units are increasingly decommissioned or displaced in WS energy markets by RES and storage technologies, meaning thermal units are not online to provide Reserve.
- **DSR and pumped storage** Reserve Up provision increases over time as while we keep the degree to which the available DSR and pump storage capacity increases, but even by 2035 these technologies provide only 22% of total AS.

Note: (1) This is our modelling assumption. In practice, required activation times are shorter than 1 second.

Improved scheduling under Co-optimisation enables thermal units to provide a larger share of Reserve Up, freeing up storage assets to participate in WS

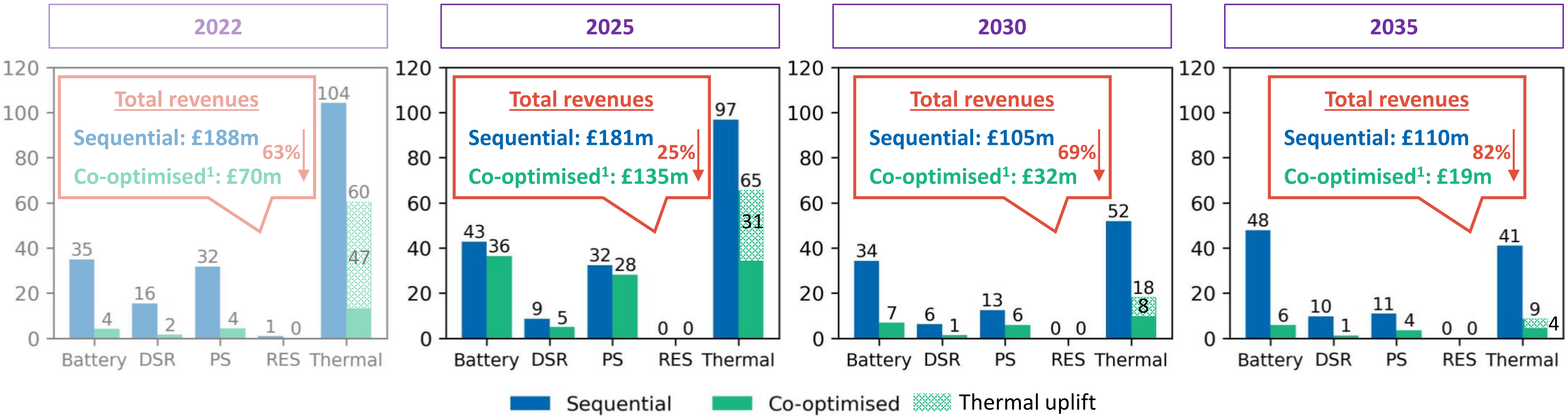
Reserve Up provision by technology (% of total provision) All years – Sequential vs Co-optimised



- We focus here on the relative composition of the Reserve Up between Sequential markets and Co-optimisation as Response is only provided by batteries.
- In both modelled scenarios, **thermal units are increasingly displaced over time by batteries in providing Reserve Up** as battery capacity increases and thermal units are decommissioned. Also, the capacity factor of the remaining thermal units decreases over the years, making them often a more costly technology to provide Reserve Up (due to the additional start-up costs they would incur).
- The relative proportion of the participation of PS and DSR is broadly constant over the modelling horizon. The installed capacity of both technologies increases over the modelling horizon (see slide 37) but the total Reserve Up requirement increases nearly proportionally (see slide 38) while the maximum participation in Reserve Up per MW installed is kept constant (see slide 40).
- Across all modelled years, **storage provides a slightly lower share of Reserve Up in Co-optimised than in Sequential (and vice-versa for thermal)**.
  - The co-optimisation algorithm is better at scheduling thermal units since it can jointly optimise unit dispatch across WS energy and AS markets, thus removing the price and dispatch uncertainty that thermal units internalise in their Reserve Up bids in the Sequential model. As shown in slides 54-55, it is also possible that in some hours more thermal is reserved under Sequential markets. That reserved thermal under the Sequential markets would not have been scheduled under a least-cost schedule and leads to important system cost increases under Sequential markets relative to Co-optimisation.
  - Increased reservation of (efficiently committed) thermal under Co-optimisation in turn reduces the need for batteries and pumped storage to provide Reserve Up, thus facilitating these units to participate more freely in the WS energy market, hence often lowering WS energy prices.

Total Reserve Up revenues are always lower under Co-optimisation, with the smallest differences in revenues compared with Sequential Markets in 2025

Reserve Up revenues by technology (£m) All years – Sequential vs Co-optimised



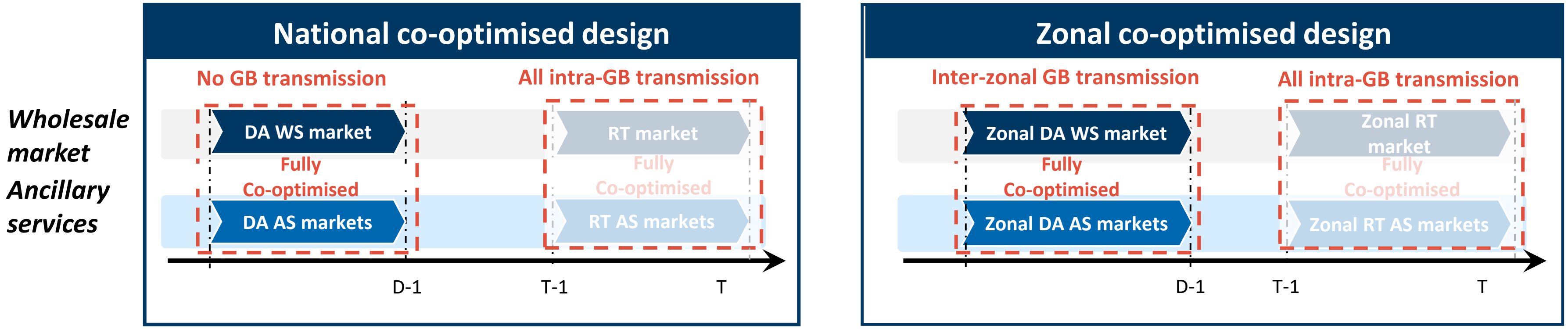
- **Reserve Up revenues earned under Co-optimisation are significantly lower (by 25-82%)** than in the Sequential model due to the reduced AS prices.
- **Total Reserve Up revenues also decrease over the course of the modelling period** in both models (e.g. from £70m in 2022 Co-optimised to £19m in 2035), as storage units replace thermal units as the dominant provider of Reserve Up (see previous slide).
- Nonetheless, **thermal units still earn the highest Reserve Up revenues of all technologies**, except under Sequential markets in 2035, for two key reasons:
  1. Thermal units face higher opportunity costs of providing Reserve Up than storage units. As a result, thermal units typically provide Reserve Up when Reserve Up **prices are high**, thereby earning a large share of Reserve Up revenues.
  2. In the **Co-optimised model**, thermal units earn a significant share of their revenues (c. 50%) from **uplift payments**, which compensate these units for losses made in the WS energy market in days when providing reserves.[1] Under the Sequential model, **all costs of thermal generators** (e.g. starting-up earlier/later and running at MSL at a loss) **are priced into their AS bids** and hence reflected in higher Reserve Up prices, also leading to higher revenues for the other technologies.

Note: (1) Uplifts are required as the AS price (reflecting the increase in system cost when incurring one additional unit of reserve/response) might not be sufficient for all thermal units to recover all their operating costs such as start-up costs.

Section 5: Day-ahead co-optimisation under zonal wholesale pricing

Our zonal assessment provides insights in how AS prices can vary across GB and elaborates on benefits of combining zonal pricing with co-optimisation

- Within its Review of Electricity Market Arrangements (**REMA**)[1], DESNZ is considering adopting locational wholesale pricing in GB and has expressed support for "the continued examination of zonal pricing". A move to locational pricing has also been supported by ESO[2] and examined in detail by Ofgem[3] and FTI as its advisors.[4]
- The **objective of our zonal assessment is different** from our welfare analysis of Co-optimisation vs Sequential markets under a national market design.
  - We do not compare a **zonal sequential vs a zonal co-optimised market.**
  - Instead, we compare **national co-optimised** DA WS and DA AS markets relative to **zonal co-optimised** DA WS and DA AS market outcomes, as illustrated below. We focus, in particular how AS prices can vary across GB under a zonal design, and we elaborate qualitatively on the benefits of combining zonal pricing with co-optimisation.



  - The quantitative zonal assessment provides an illustrative, rather than an exhaustive, welfare analysis of DA zonal vs national co-optimised WS and AS markets due to choice of **conservative modelling assumptions under the zonal market design** (no reduction in total reserve requirements relative to national design, and no cross-zonal procurement of reserves).[5]
- In this section we discuss:
  1. Our methodology for co-optimisation under a zonal market design;
  2. Key modelling assumptions;
  3. Results of the quantitative assessment; and
  4. Qualitative discussion of further (unmodelled) impacts of zonal pricing and the benefits of combining zonal design with co-optimisation.

Notes: (1) DESNZ (March 2024) "Review of Electricity Market Arrangements: Options Assessment" (link); (2) ESO "Net Zero Market Reform" (link); (3) See Ofgem "Assessment of Locational Wholesale Pricing for GB" (link); (4) See FTI report for Ofgem: "Assessment of Locational Wholesale Electricity Market Design Options in GB" (link); (5) Zonal reserve requirements are calibrated to be equal to national reserve requirements which is not reflective for how zonal reserve requirements would be determined (which is out of scope of this work). Also, no cross-zonal procurement of reserves is assumed to limit computational complexity while cross-zonal procurement is expected to be highly beneficial to reduce zonal AS prices.

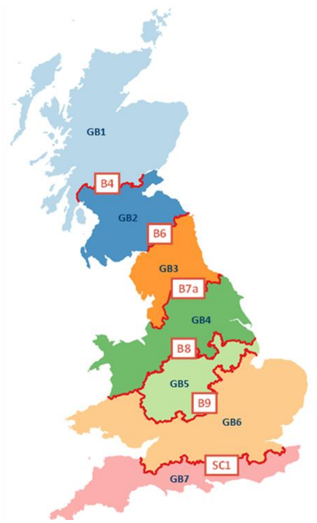
Implementation of the methodology

FTI CONSULTING

With zonal pricing, both wholesale and reserve markets clear with distinct prices for each zone, reflecting local demand and supply conditions

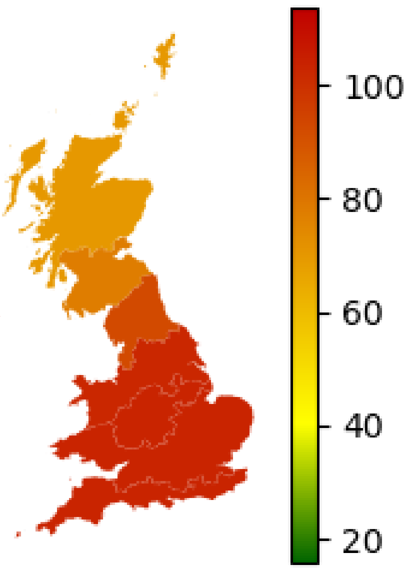
- In this section, we examine Co-optimisation under a zonal WS electricity market design. We use a **seven-zone model**, which was originally derived for our work on assessing the options for locational design with Ofgem.[1] We continued to use the set up with the seven zones, but we recognise that different arrangements may be possible.
- Under a zonal market design, **locational pricing is introduced to both wholesale energy and reserve markets** by considering inter-zonal transmission constraints at the day-ahead stage.
- **Response markets remain national** under the zonal design. In line with how ESO operates the network in practice, the model assumes sufficient transmission capacity on all intra-GB lines to ensure frequency response can be provided by assets located anywhere on the GB network.
- Co-optimisation under zonal pricing still seeks to minimise the total GB system cost of meeting demand for energy and AS (as it did under the national design in previous section), but the consideration of inter-zonal transmission constraints at day-ahead means that **energy and reserve markets** can clear with **zonally-distinct prices** that are more reflective of the cost of meeting demand in each zone.

Seven-zone model



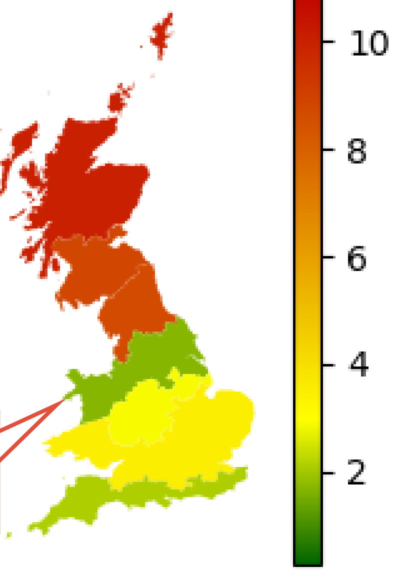
Zonal average wholesale prices - 2025

Co-optimisation clears with zonally-distinct prices for the WS energy market…



Zonal AS average prices (Reserve Up) - 2025

…and also with zonally-distinct reserve prices



Notes: (1) See FTI report for Ofgem: "Assessment of Locational Wholesale Electricity Market Design Options in GB" (link).

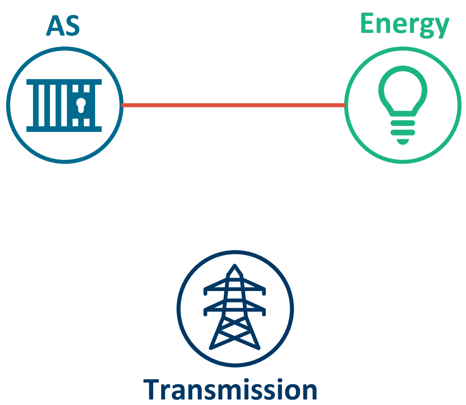
FTI CONSULTING

Co-optimisation under zonal pricing can take various forms; our quantitative assessment models zonal wholesale pricing with 'within-zone' AS procurement

Possible degrees of co-optimisation with locational pricing

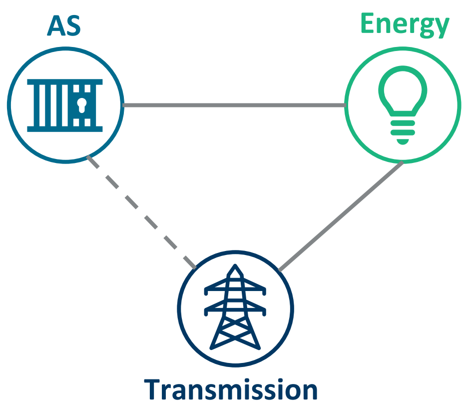
Modelled in this report

Co-optimisation national



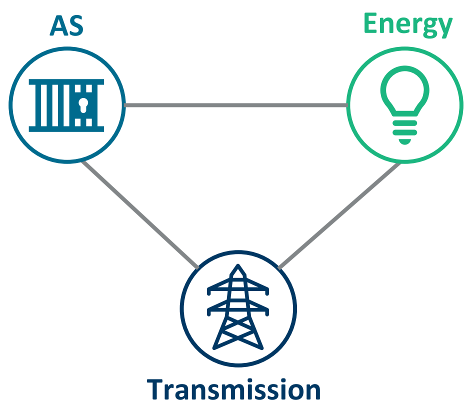
Modelled in this report

Co-optimisation zonal - partial (only intra-zonal reserve procurement)



Not modelled in this report

Co-optimisation zonal - full (cross-zonal procurement of reserves)



The **forward-looking assessment** explored the co-optimisation of energy and AS at day-ahead under the **current national wholesale market design**, identifying the optimal outcome across both markets while **ignoring intra-GB transmission constraints**.

We explore this approach to co-optimisation under zonal design in the following slides

- **Zonal wholesale pricing** enables a further degree of co-optimisation to be introduced between **energy, AS and transmission assets**.
- The extent to which **transmission is added to the co-optimisation** depends on whether **only energy, or in addition AS**, can be traded **across zonal transmission boundaries**.

In our **quantitative assessment**, we take the **conservative view** that **energy** can be traded across inter-zonal transmission boundaries up to the available transmission capacity, but **each zone's reserve requirement must be met by assets located within the zone.**[1]

We also provide a **qualitative discussion (slide 81)** of the **additional benefit** that could be realised with co-optimisation under zonal pricing via **efficient allocation of inter-zonal transmission capacity across the WS and reserve markets to allow for cross-zonal reserve procurement.**

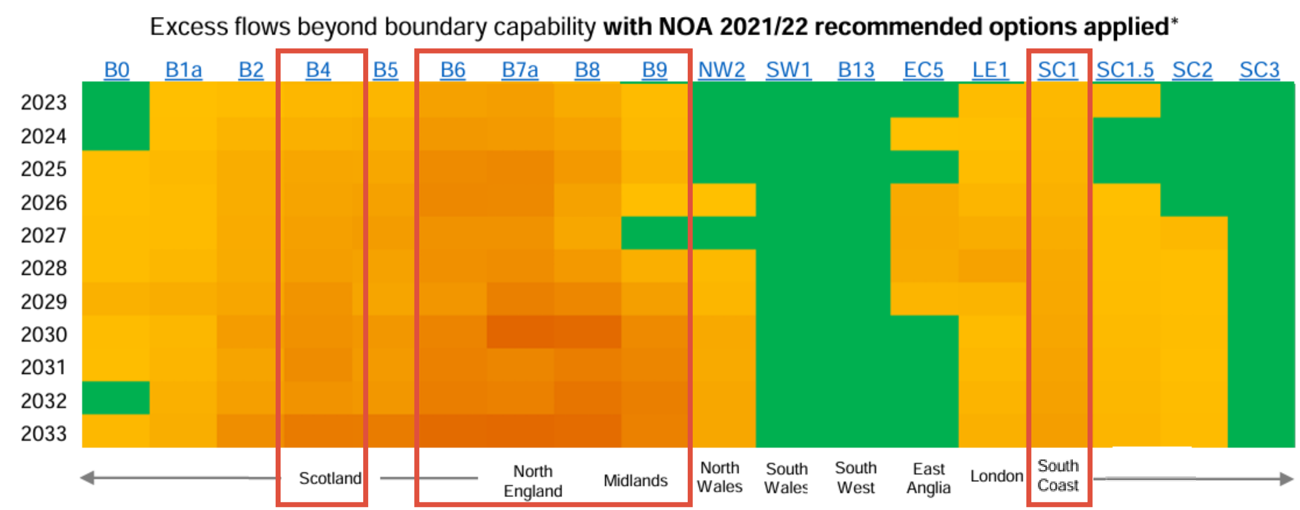
Note: (1) Response is nationally procured assuming safety margins on transmission lines allow it to be delivered from anywhere on the GB grid.

Modelling assumptions

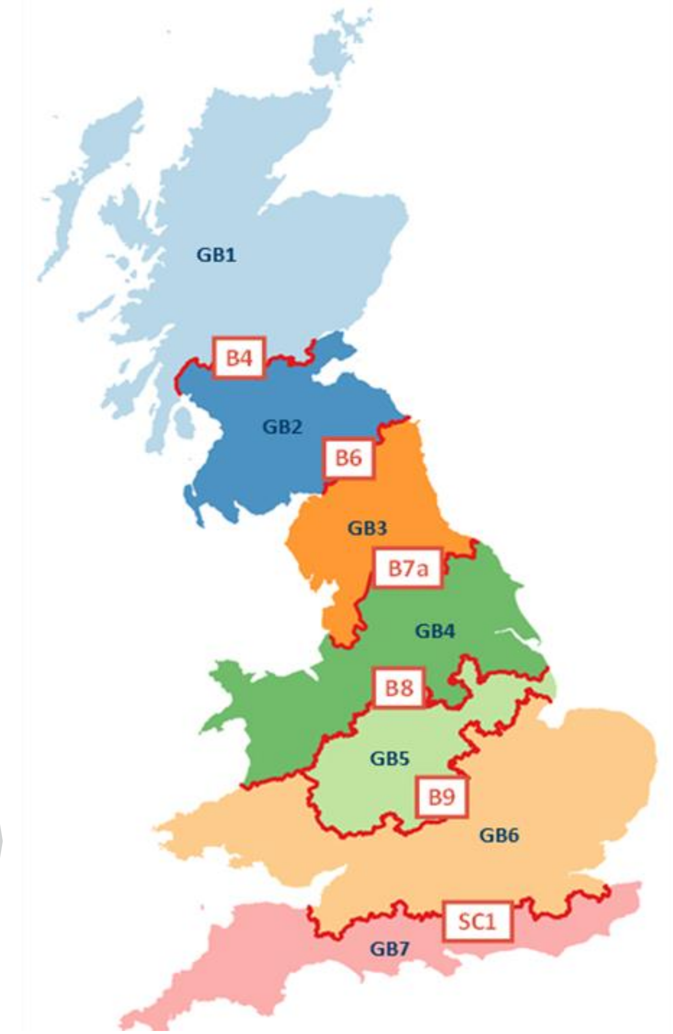
We split the GB market into seven zones, delineated by the most constrained boundaries, with transmission data from HND and NOA 2021/22 Refresh

- The **seven GB zones** used in our analysis, which we originally derived for our previous work on locational design with Ofgem,[1] are separated by the following boundaries:
  - B4: SSEN – SP transmission border
  - B6: SP – NGET transmission border
  - B7A: Upper North of England
  - B8: North of England to Midlands
  - B9: Midlands to South of England
  - SC1: South Coast
- As shown in the chart below highlighting congestion across key transmission boundaries, we focused on these **zonal boundaries because they are the most constrained**, based on ESO forecasts and modelling set out in the Electricity Ten Year Statement ("**ETYS**") 2022.[2]
- Transmission capacity between zones evolves across the modelling period, following the HND and NOA 2021/22 Refresh assumptions.

Most constrained transmission boundaries (2023-2033), ETYS 2022



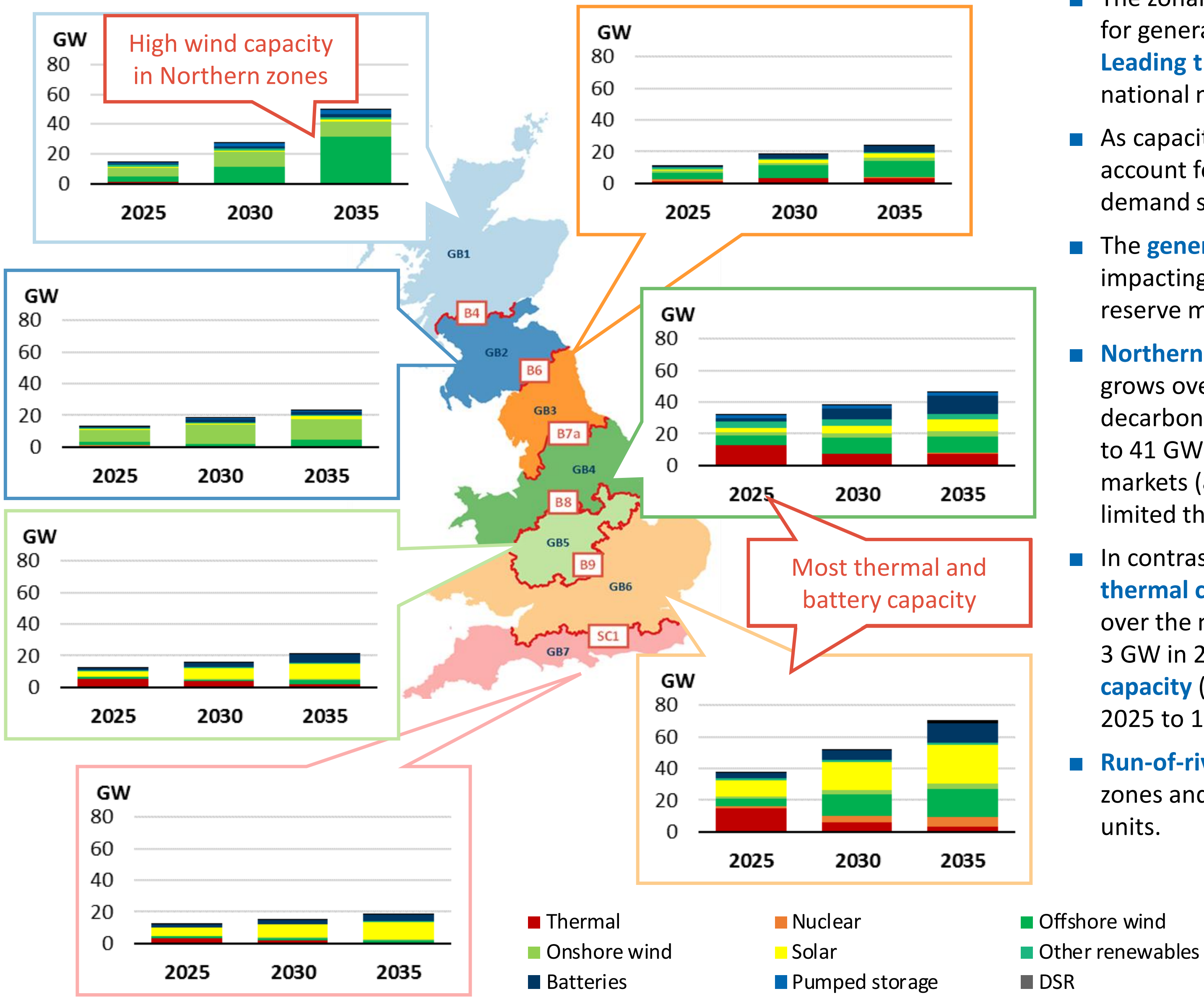
Seven-zone GB model



Notes: (1) See FTI report for Ofgem, Assessment of Locational Wholesale Electricity Market Design Options in GB (link); (2) See ESO, ETYS 2022 (link).

Generation capacity data from FES 2022 LtW leads to different generation mixes per zone which are –in aggregate– consistent with the National model

Electricity generation capacity by GB zone, FES 2022 LtW scenario, (GW)

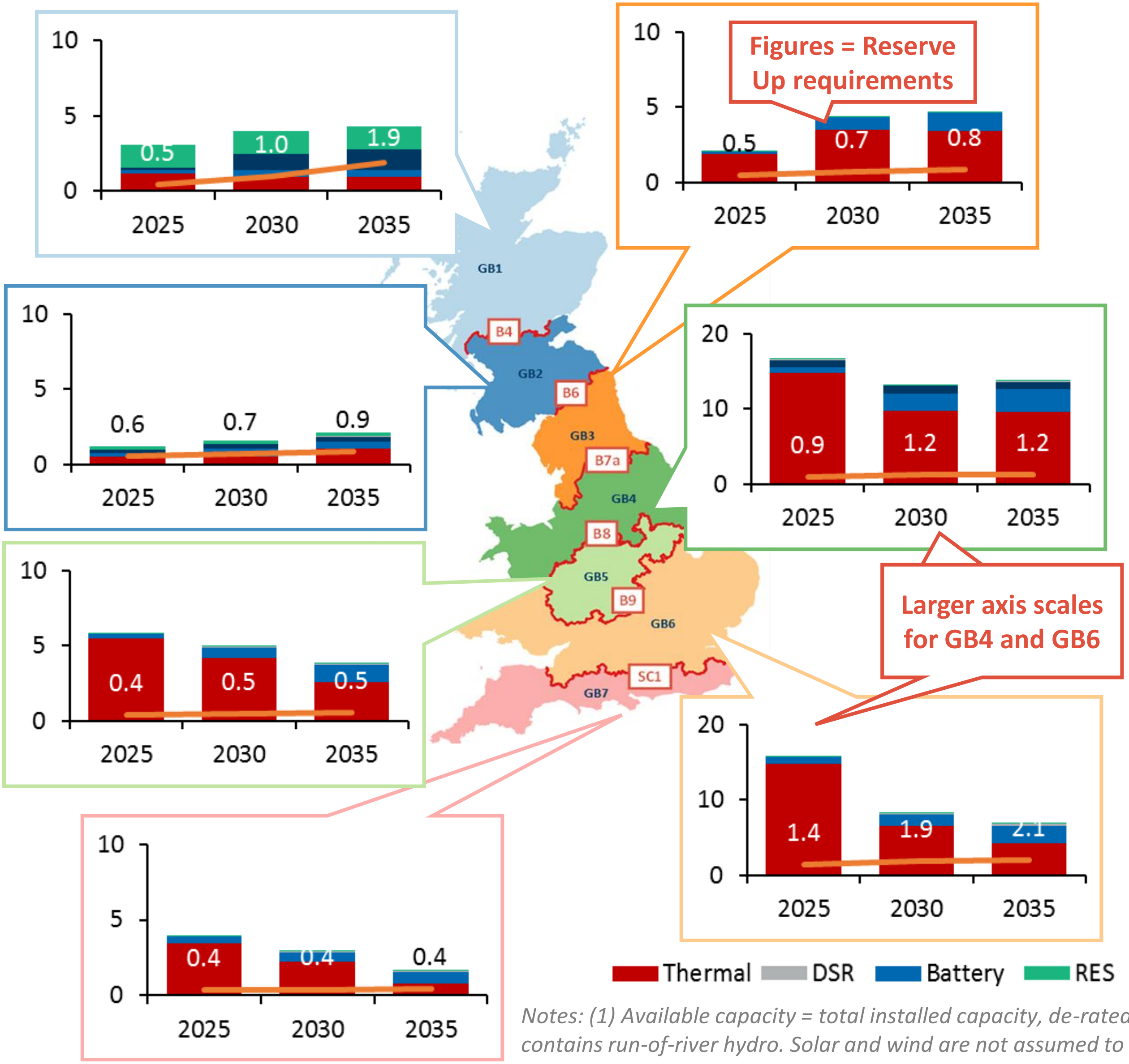


- The zonal model uses installed capacity and location data for generation units and batteries from the **FES 2022 Leading the Way** Scenario, thus aligning with then national model.
- As capacity roll-out is fixed to FES, the model does not account for any optimisation of generator, storage or demand siting in response to locational pricing.[1]
- The **generation mix varies significantly across zones**, thus impacting the capacity available to bid into each zonal reserve market.
- **Northern zones have high shares of wind capacity**, which grows over the modelling period as the system decarbonises (e.g. increasing in GB1 from 10 GW in 2025 to 41 GW in 2035). Wind capacity dominates these markets (82% of installed capacity in GB1 in 2035), with limited thermal and battery capacity.
- In contrast, the **Southern zones have higher installed thermal capacities**, although thermal capacity decreases over the modelling period (from 15 GW in GB6 in 2025 to 3 GW in 2035), being **replaced by RES and battery capacity** (battery capacity increases from 3 GW in GB6 in 2025 to 12 GW in 2035).
- **Run-of-river hydro** capacity is concentrated in Northern zones and only GB1, GB2 and GB4 have **pumped storage** units.

Note: (1) This leads us to underestimate benefits of a zonal market design as we would expect some re-siting under zonal pricing, to better match local supply and demand.

Demand and supply balance: there is significant variation across zones in Reserve Up requirements and the technology mix able to provide Reserve Up

Average Reserve Up requirement and available capacity by zone and technology[1] (GW)



- **Reserve Up requirements** (orange line and figures in charts) are dependent upon both WS energy demand and expected wind output[2], and hence vary significantly across zones.
  - Reserve requirements are consistently highest in GB6, as this zone has the largest WS energy demand.
  - The Reserve Up requirements grows rapidly in GB1 (from 0.5 GW/h in 2025 to 1.9 GW/h in 2035) due to the rapid expansion in wind capacity in GB1, causing extra reserve to be required to cover wind forecast uncertainty.
  - In contrast, Reserve requirements are lowest in GB5 and GB7 (0.4-0.5 GW/h across all years), as these zones have both low WS energy demand and low installed wind capacity.
- The **available capacity[1] and mix of technologies** able to provide Reserve Up also varies significantly across zones.
  - Across Southern zones, thermal units make up most of the available capacity, although battery capacity increases over the modelling period. GB4 and GB6 have the largest installed capacities available to provide Reserve Up (note larger axis scales for these zones in charts).
  - Only GB1, GB2 and GB4 have pumped storage units, whilst GB1 and GB2 also have the largest available RES capacity (i.e. run-of-river hydro).

Notes: (1) Available capacity = total installed capacity, de-rated by Reserve Up participation factor (see slide 40). The RES category contains run-of-river hydro. Solar and wind are not assumed to provide Reserve Up; (2) See following slides for methodology to calculate zonal Reserve Up requirements.

The zonal reserve demand is allocated relative to load and wind output within each zone and equals the national reserve demand in aggregate

- Total reserve required across GB is composed of **fixed, time-varying and wind-varying demand**.[1] The three components and how they are derived for the zonal setup are set out below:

Regional allocation of 'fixed' Reserve Up/Down and 'time-varying' Reserve Up requirements (MW), 7am 11th April 2025

GB1 (Load proportion = 3%)
Reserve Up: 68 MW
Reserve Down: 44 MW

GB2 (Load proportion = 6%)
Reserve Up: 162 MW
Reserve Down: 104 MW

GB3 (Load proportion = 8%)
Reserve Up: 212 MW
Reserve Down: 136 MW

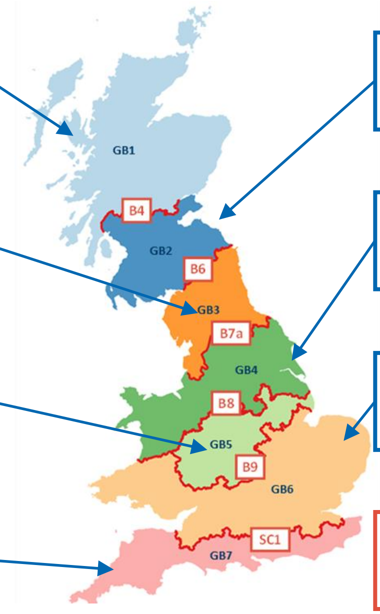
GB4 (Load proportion = 20%)
Reserve Up: 535 MW
Reserve Down: 343 MW

GB5 (Load proportion = 13%)
Reserve Up: 347 MW
Reserve Down: 223 MW

GB6 (Load proportion = 39%)
Reserve Up: 1,040 MW
Reserve Down: 667 MW

GB7 (Load proportion = 11%)
Reserve Up: 286 MW
Reserve Down: 183 MW

Market total
Reserve Up: 2,650 MW
Reserve Down: 1,700 MW



- In our zonal modelling, we take the **conservative assumption** that total **GB reserve demand** in each hour will be **equal** to that procured under a **national** market design.
- For the **'fixed' and 'time-varying' reserve** requirements, where demand does not vary based on forecasted wind output, national reserve demand is allocated across zones **proportional** to the **zone's hourly demand for electricity**.

Regional allocation of 'wind-varying' Reserve Up requirements (MW), 7am 11th April 2025

GB1
Reserve Up: 498 MW

GB2
Reserve Up: 402 MW

GB3
Reserve Up: 132 MW

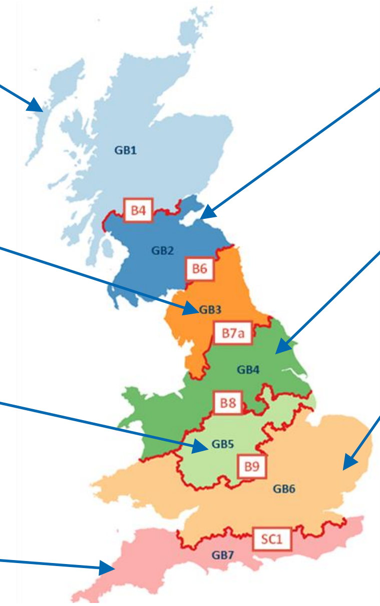
GB4
Reserve Up: 234 MW

GB5
Reserve Up: 71 MW

GB6
Reserve Up: 215 MW

GB7
Reserve Up: 59 MW

Market total
Reserve Up: 1,611 MW



- As with the National model, the **'wind-varying'** component of zonal Reserve Up is calculated based on **forecasted post-constraint wind output**.[2]
- We perform **hourly** 'wind-varying' Reserve Up forecasts for **each zone individually**, based on **local weather conditions**. However, in hours where **wind output varies** substantially **between zones**, this leads to different total Reserve Up demand between national and zonal, as the wind-varying reserve requirement scales sub-linearly with the forecasted wind output.
- We **maintain a consistent GB Reserve Up demand between national and zonal** by allocating any **difference in Reserve Up demand** across GB zones, as outlined in the following slides.

Notes: (1) We aggregate three reserve services into our Reserve Up service, the largest of which being Balancing Reserve Up, which is the only service with wind-varying and time-varying components; Reserve down is fixed. (see slide 39); (2) See Appendix 13 for explanation and a worked example.

In the zonal model, forecasting 'wind-varying' reserve demand individually for each zone would cause total GB reserve demand to be lower than in national…

- As with the national model, we estimate the hourly demand for wind-varying Reserve Up as a function of post-constraint wind output…
- …but for the zonal model we perform a separate forecast for each zone.
- When hourly wind conditions vary between zones, this can lead to a difference in total GB reserve demand compared to national…
- …as in the first example to the right, where, without further adjustment, zonal would have 1/3 less reserve demand.[1]

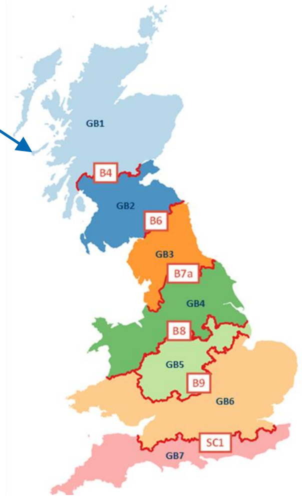
7am 11th April 2025, High GB1 wind, low wind rest of GB

GB1

Wind output: 4,995 MW
Capacity factor: 52%

Rest of GB

Wind output: 2,759 MW
Capacity factor: 9%



	GB1	Rest of GB	GB total (zonal)	GB total (national)
Wind output	4,995 MW	2,759 MW	7,754 MW	7,754 MW
Wind capacity factor	52%	9%	19%	19%
'Wind-varying' reserve demand	484 MW	585 MW	1,069 MW	1,611 MW
'Wind-varying' reserve as proportion of wind output	10%	21%	14%	21%

Without additional adjustment, the amount of **'wind-varying' reserve** procured across GB would **differ significantly** between the **national** and **zonal** models in hours when wind conditions vary between GB zones…

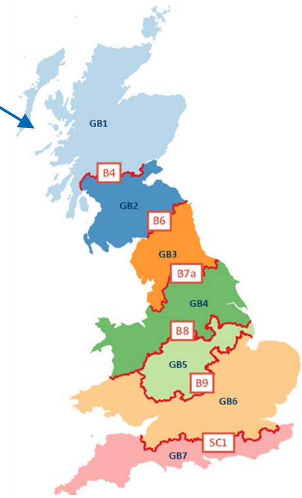
6pm 8th March 2025, high GB1 wind, high wind rest of GB

GB1

Wind output: 4,995 MW
Capacity factor: 52%

Rest of GB

Wind output: 25,606 MW
Capacity factor: 80%



	GB1	Rest of GB	GB total (zonal)	GB total (national)
Wind output	4,995 MW	25,606 MW	30,601 MW	30,601 MW
Wind capacity factor	52%	80%	74%	74%
'Wind-varying' reserve demand	484 MW	1,610 MW	2,093 MW	2,180 MW
'Wind-varying' reserve as proportion of wind output	10%	6%	7%	7%

…although the issue is **less prevalent** in hours when wind conditions are **similar across zones**.

Note: (1) See Appendix 9 where a table with the mapping of forecasted wind output on Balancing Reserve demand is provided.

…so to be consistent in our treatment of national and zonal, we assign any 'shortfall' from national across zones relative to their proportion of GB load

- In hours where there is less demand forecasted under the zone-by-zone approach than under national, we first calculate the shortfall (i.e. the difference in demand)…
- …and then allocate the shortfall across zones relative to their share of GB load in the relevant hour.
- This ensures that the modelling of national and zonal have equal demand for Reserve Up…
- …while directing the demand for 'wind-varying' reserve to the zones with the most wind in each hour.

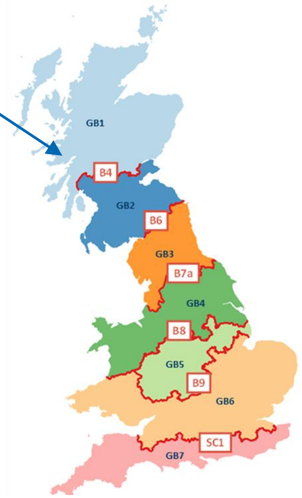
7am 11th April 2025, high GB1 wind, low wind rest of GB

GB1

Wind output: 4,995 MW
Capacity factor: 52%

Rest of GB

Wind output: 2,759 MW
Capacity factor: 9%



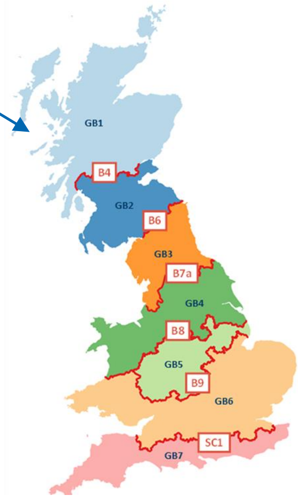
6pm 8th March 2025, high GB1 wind, high wind rest of GB

GB1 (19%)

Wind output: 4,995 MW
Capacity factor: 52%

Rest of GB

Wind output: 25,606 MW
Capacity factor: 80%

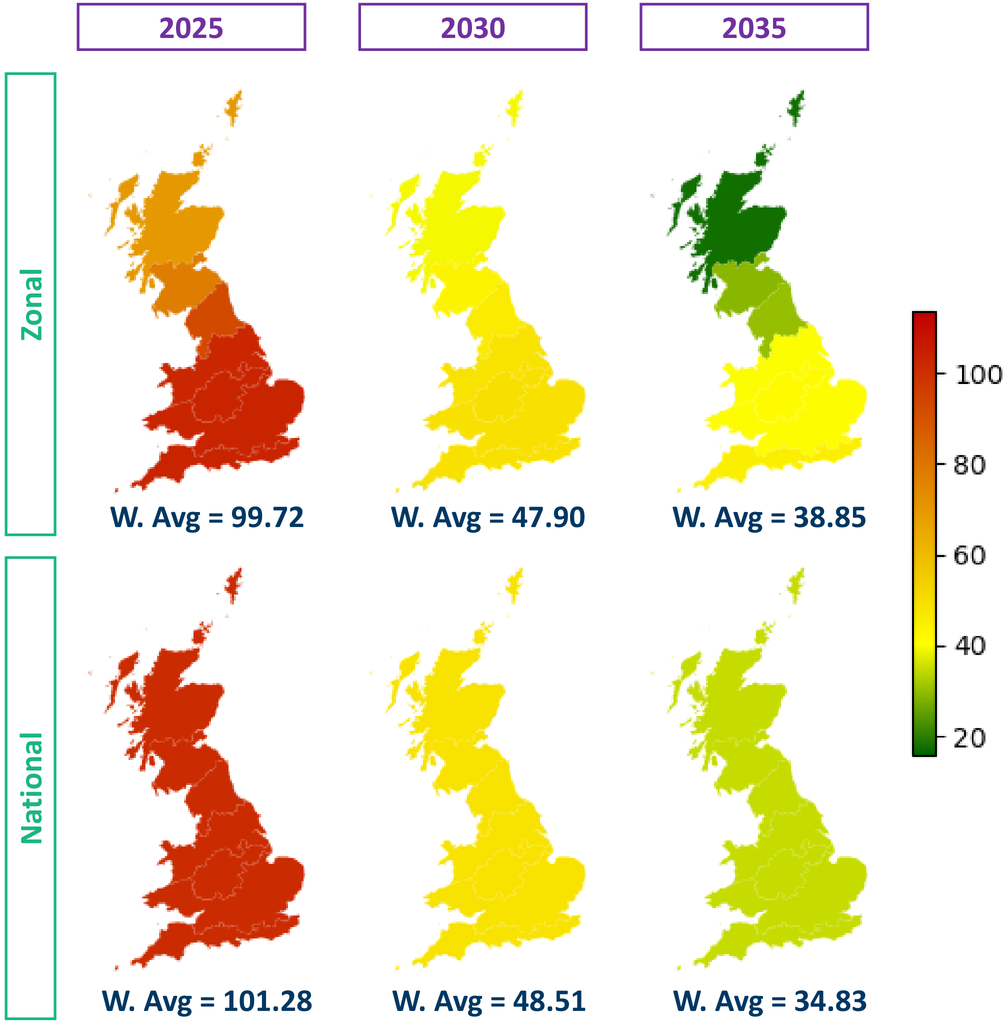


	7am 11th April 2025: GB1	Rest of GB	GB total (zonal)	6pm 8th March 2025: GB1	Rest of GB	GB total (zonal)
Wind-reserve demand (national approach)	---	---	1,611 MW			2,180 MW
Wind-reserve demand (zonal approach)	484 MW	585 MW	1,069 MW	484 MW	1,610 MW	2,093 MW
Shortfall	---	---	542 MW	---	---	87 MW
Load proportion	3%	97%	---	3%	97%	---
Shortfall allocation	16 MW	526 MW	---	3 MW	84 MW	---
Total zonal wind-varying reserve	500 MW	1,111 MW	1,611 MW	487 MW	1,694 MW	2,180 MW

Modelling results

Similarly to the National model, WS energy prices fall over time in the zonal model, with North of GB experiencing the lowest WS energy prices

WS Energy weighted average prices by zone (£/MWh)

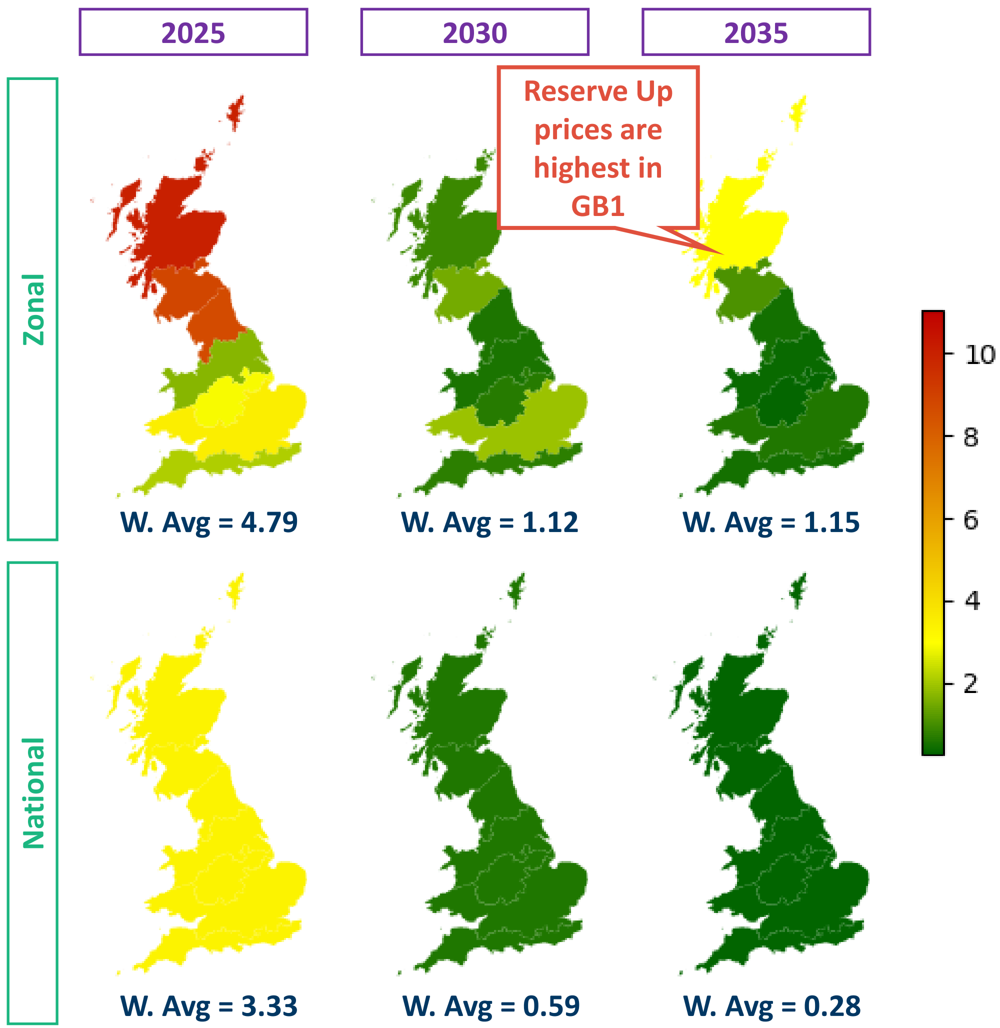


- Similarly to the National model, **WS energy prices fall over time** in the zonal model as high-marginal cost thermal units are increasingly displaced by low-marginal cost RES units.
- However, in the zonal model there is significant variation in WS energy prices across zones, with **North of GB experiencing the lowest prices**, due to high installed wind capacity (over 40 GW in GB1 alone by 2035) and comparatively low demand.
- In 2025, thermal units often set the price in the National model. In the zonal model, thermal units still set the price in Southern zones, leading to **very similar prices to the National model** (1-2% increase across GB4 – GB7), but wind is more likely to set the price in in the North of GB, thus leading to lower zonal prices.
- 2030 has the lowest variation in average prices across zones as, under HND and NOA 2021/22 Refresh assumptions, **transmission capacity is significantly built out** between 2025 and 2030, enabling significant volumes of power to flow from North to South. Consequently, average prices across zones in the zonal model are similar to the average price in the National model.
- In 2035, price variations widen across zones as the continued growth in wind capacity in the North of GB exceeds transmission development, hence exacerbating North-South transmission constraints. Unlike in 2025, RES units often set the price in the National model in 2035, hence zonal pricing not only reduces Northern GB prices (by 50% in GB1), but also slightly increases prices across Southern zones, where prices are more likely to be set by thermal units or exports to Europe.[1]

Note: (1) Zonal pricing also delivers consumer benefits in the form of reduced constraint costs and the creation of inter-zonal congestion revenues (further there can also be impacts on required CfD payments). We do not analyse these benefits in this assessment, however our previous work for Ofgem identified an aggregate consumer benefit of transitioning from a national WS energy market to a zonal market. See FTI report for Ofgem, Assessment of Locational Wholesale Electricity Market Design Options in GB (link).

In the zonal Reserve model, Reserve Up prices vary significantly across zones and are often highest in North of GB due to the high wind output

Reserve Up weighted average prices by zone (£/MW/h)

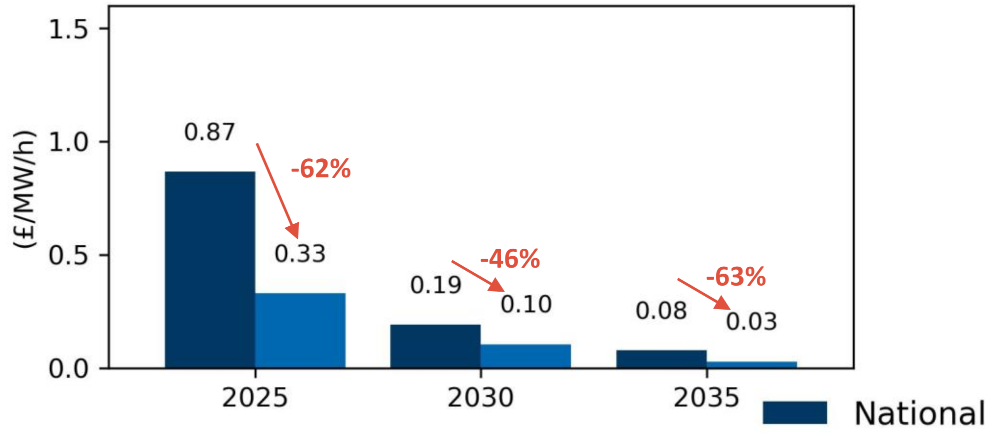


- Similarly to the National model, average **Reserve Up prices fall over the modelling horizon** as battery capacity increases, but this transition does not occur evenly across zones.
- Zonal Reserve up prices are driven by the location-specific supply and demand balance, causing **prices to vary significantly across zones**.
  - In general, zones where demand for reserve up is highest (i.e. areas with a high share of wind output) will have higher Reserve Up prices.
  - Similarly, areas where supply of Reserve Up is limited (e.g. areas with resources we assume are not able to provide Reserve Up such as wind) will also have higher Reserve Up prices.
- These two factors combine particularly strongly in the most Northern zone (GB1). High wind output causes the Reserve Up requirement to increase from 0.5 GW/h in 2025 to 1.9 GW/h in 2035, and as wind is the dominant generation source in GB1 (see slide 67), there is limited storage and thermal capacity to meet the large reserve requirements. As a result, **GB1 has the highest Reserve Up prices**, averaging £10/MW/h in 2025.
- In contrast, **Reserve Up prices are lowest in GB4** (North England and North Wales) across the modelling horizon, as low wind generation leads to relatively low reserve requirements, whilst pumped storage units and baseload thermal capacity are large sources of low-priced reserve (see slide 67).
- We also see the first instances of **non-zero Reserve down prices under Zonal pricing** (not shown in the diagrams).[2] Wind is the predominant technology to provide Reserve Down under National pricing. However, when having zonal Reserve down markets **batteries** are increasingly required to provide Reserve down in Southern zones with low wind generation. Dispatched generators (above their MSL) can provide Reserve down for free, but batteries may face **opportunity costs**, thereby increasing Reserve down prices.

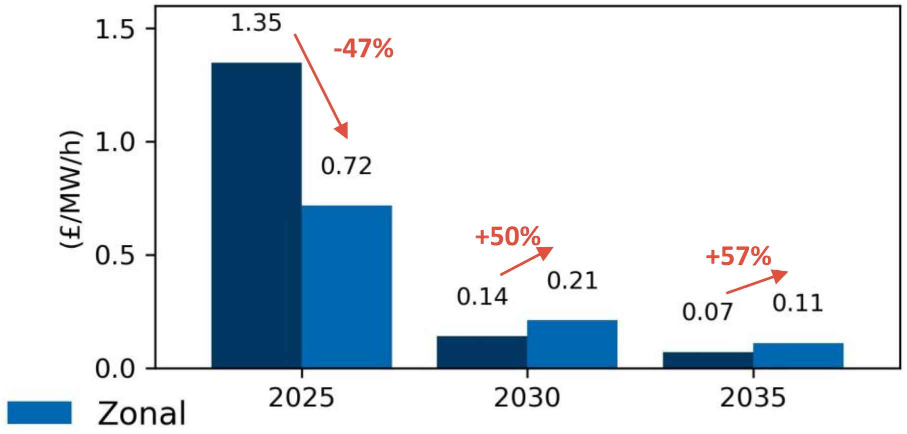
Notes: (1) See slide 68 for a detailed breakdown across zones of reserve requirements and capacity available to provide reserve; (2) Nonetheless, average zonal Reserve down prices are still very low (£0.01/MW/h or below across all zones in all years).

Zonal pricing leads to fewer batteries in Reserve Up and more in Reserve Down, typically reducing prices in Response Up and increasing them in Response Down

Average Response Up prices (£/MW/h) 2025/30/35 – National vs Zonal



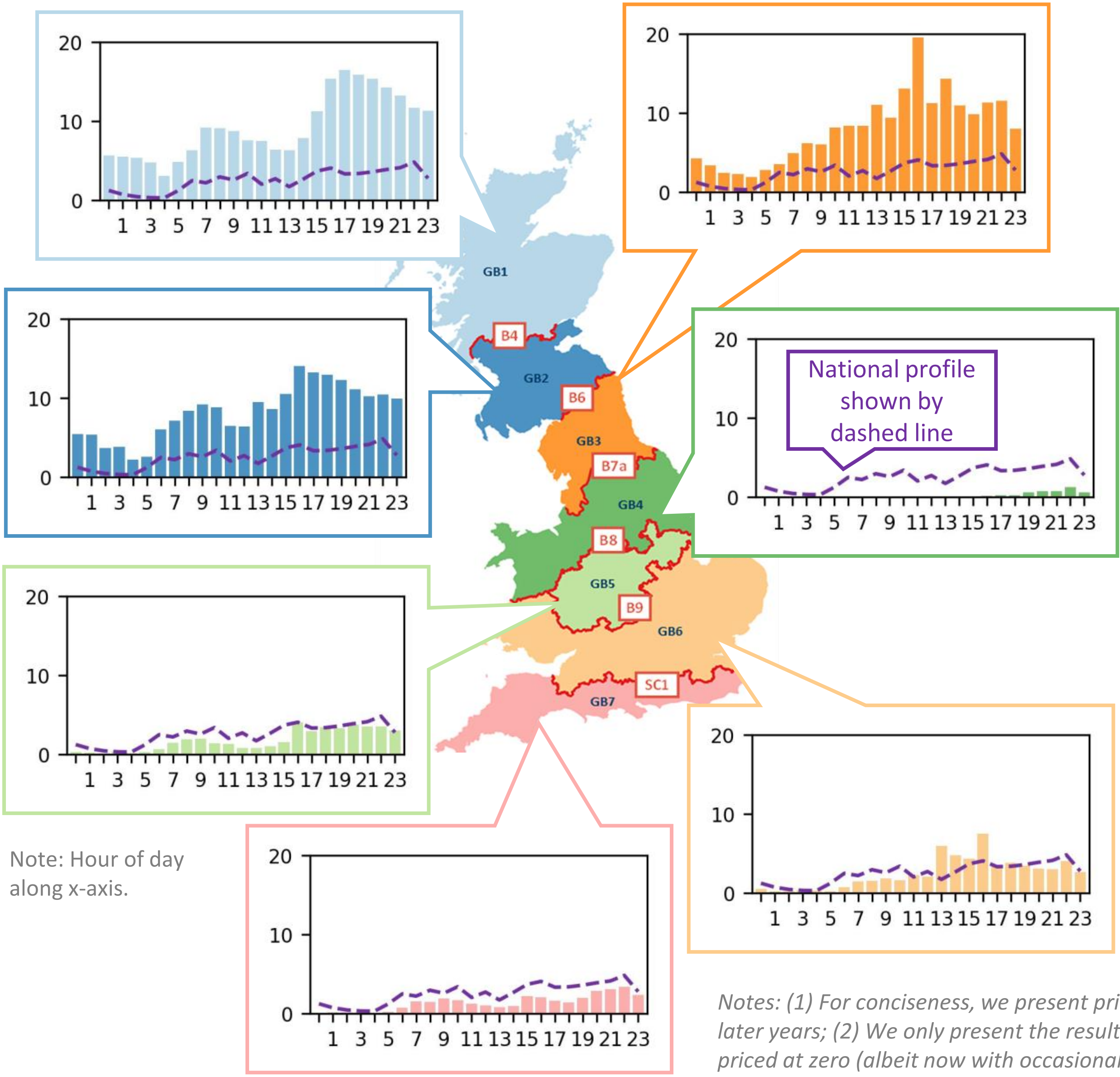
Average Response Down prices (£/MW/h) 2025/30/35 – National vs Zonal



- **Response is procured nationally** in both models, but the **introduction of zonal Reserve markets impacts Response prices through battery reservations**.
- **Response Up prices fall by 46-63%** in the zonal model as total battery provision of Reserve Up falls in the zonal Reserve markets (see slide 77), hence freeing up battery capacity to provide response.
  - Under National pricing, batteries across the country are considered equally valuable in providing Reserve Up, but in the zonal model Reserve cannot be procured across zonal borders, hence batteries located in zones with high capacities of batteries installed are more likely to be procured in Reserve Up under National pricing (as they can be used to fulfil the National Reserve Up requirement) than under Zonal pricing (to fulfil the Zonal requirement).
  - For instance, GB4 has the largest battery capacity of any zone in 2030 (see slide 68), as well as significant pumped storage and thermal capacity, causing GB4 to have the lowest Reserve Up prices of all zones. Demand for Reserve Up in GB4 (under a zonal model) is met to a greater extent by pumped storage and thermal capacity, freeing up GB4 batteries to provide Response Up at a low opportunity cost.
- Battery reservations in zonal Reserve markets impact Response down prices through **two channels**:
  1. Reduced battery provision of Reserve Up allows more batteries to provide Response services putting downward pressure on both Response prices.
  2. Zonal Reserve markets lead to increased battery provision of Reserve down (by over 150% in 2035) because the Southern zones sometimes have insufficient wind and thermal generation to meet their Reserve down requirements. Increased battery participation in Reserve down reduces the available fleet of batteries with a low state of charge to provide Response down, thereby increasing Response down prices.
- The Reserve Up market is tight in 2025 (see previous slide) causing the **first effect to dominate** and hence **Response Down prices fall by 47% in 2025** in the zonal model. However, total battery capacity is significantly higher in the later years (increased from 2.9 GW in 2025 to 22.6 GW in 2035), hence reducing tightness in Reserve Up, so the **second effect dominates** and **Response Down prices increase by 50-57% in 2030 and 2035.**

Zonal Reserve Up price patterns generally follow that seen under National pricing, although with some variability between the zones

Avg hourly Reserve Up price by zone (£/MW/h) - 2025



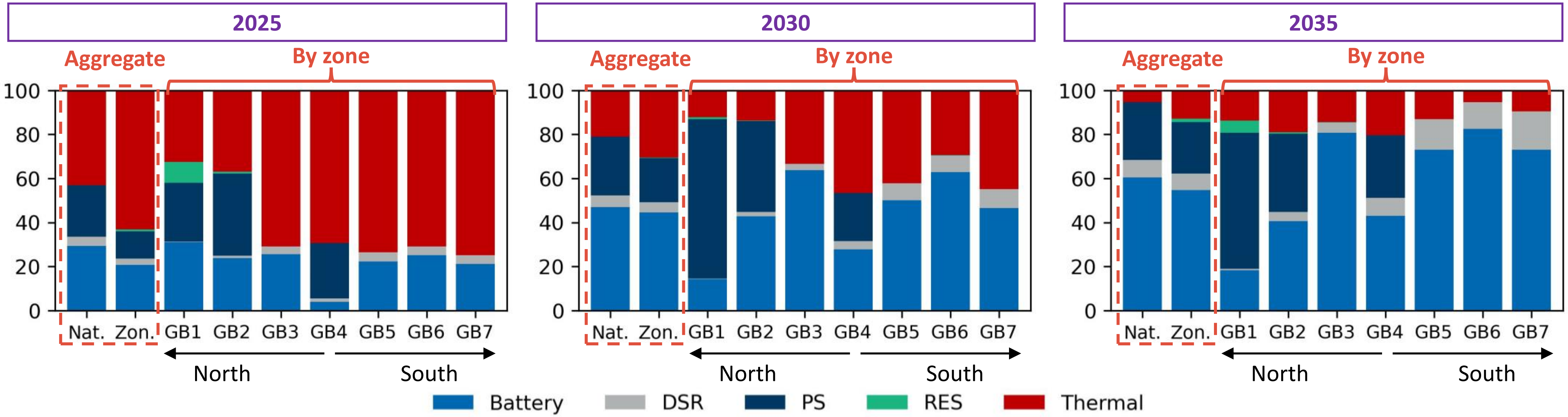
Note: Hour of day along x-axis.

- **Reserve Up prices are highest in northern zones** due to high Reserve requirements and low capacity of batteries and thermal units to provide Reserve Up cheaply (see slide 68).
- **Zonal Reserve** Up prices generally **follow** the pattern seen in under **National pricing**, with higher Reserve Up prices later in the day, around the evening energy price peak.
- The key difference, however, is that under National pricing Reserve Up prices peak after the evening energy price peak (i.e. around 10pm), but in many zones in the zonal model (particularly in the North), Reserve Up prices peak in the mid-afternoon (around 4-5pm).
  - Batteries provide a larger share of Reserve Up in the National model (see slide 77), hence Reserve Up prices peak after the evening WS energy price peak, just after many batteries have discharged, when the fleet of batteries having enough battery charge left to provide Reserve Up is small.
  - Battery capacity is concentrated in the South, hence Southern regions have low Reserve Up prices, with similar profiles to the National model.
  - However, thermal units set the high Reserve Up prices in Northern zones, hence Reserve Up prices are typically higher in the late afternoon/ early evening when thermal units face higher opportunity costs in the WS market.

Notes: (1) For conciseness, we present price profiles only for 2025, but similar patterns can be observed in the later years; (2) We only present the results for Reserve Up results. Reserve Down continues to be predominantly priced at zero (albeit now with occasional non-zero prices), and Response services are provided nationally.

The zonal model better captures that some batteries are not well-located to provide Reserve Up, therefore it must rely in some zones on other alternatives

Reserve Up average provision by technology and zone (% of total provision)

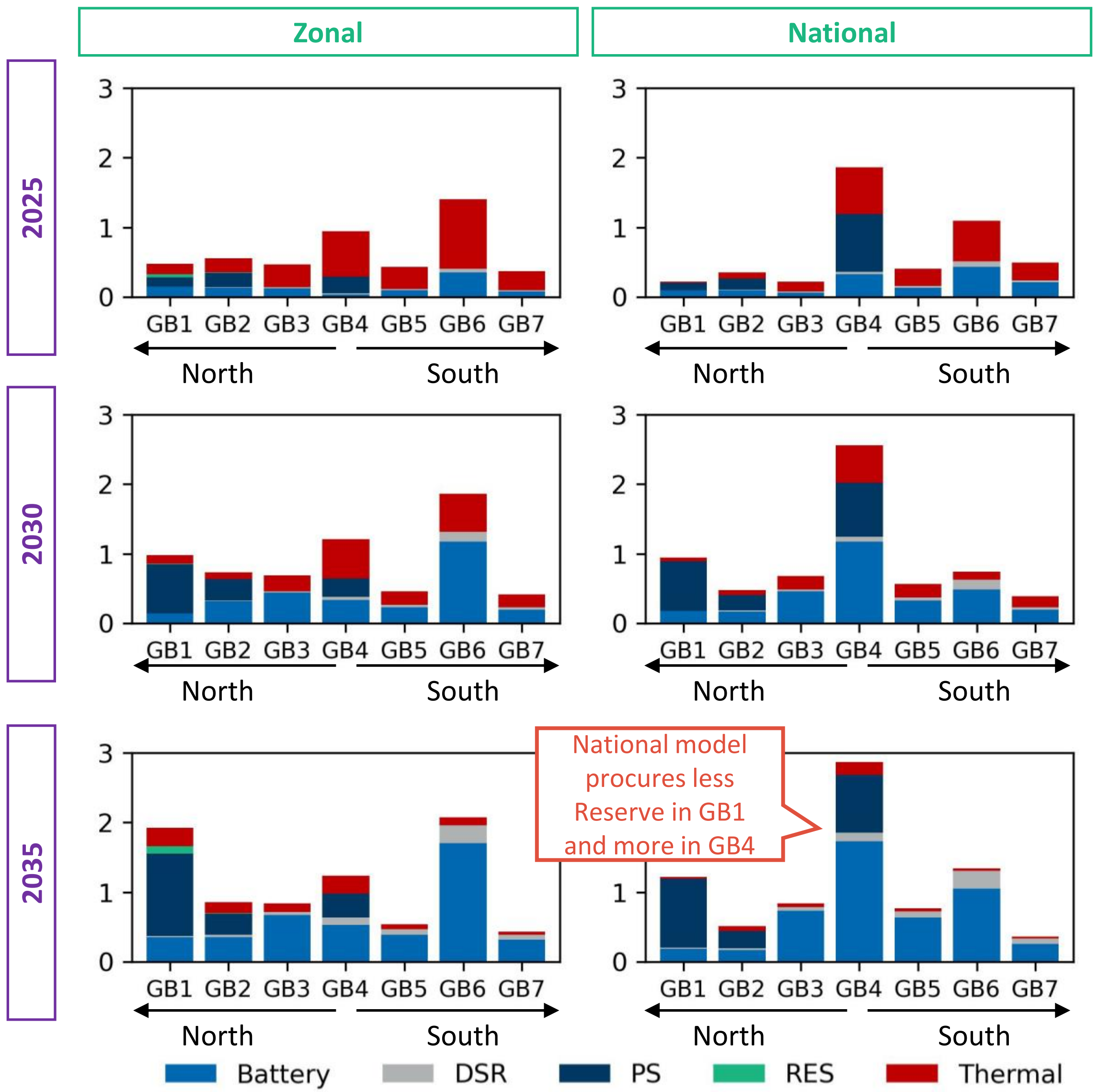


- Across the modelling horizon and all zones, **thermal units provide a decreasing share of Reserve Up** while the share of Reserve Up provided by batteries increases (see left two bars of charts above). This is explained by the changing generation mix (see slide 67).
- On aggregate, **batteries provide a lower share of Reserve Up under zonal** (21% zonal aggregate vs 29% National in 2025) and **thermal units provide a higher share** (63% zonal aggregate vs 43% National in 2025). Because the model setup does not allow for cross-zonal reserve procurement, zones with limited battery capacity or few running thermal with spare capacity (e.g. CHP) must rely on more expensive technologies to meet Reserve requirements. In zones with a high supply of low-cost options to provide Reserve Up, more batteries (relative to under National pricing) can be allocated to Response markets leading to lower Response prices (see slide 75).
- The provision of Reserve Up by other technologies **varies significantly across zones** , depending upon the demand and supply of Reserve in each zone.[1]
  - **Pumped storage units** are a relatively low-cost source of Reserve Up provision, hence pumped storage units provide a large share of Reserve Up (1.2 GW/h average in GB1 in 2035) in the zones where they are present (i.e. GB1, GB2 and GB4).
  - The opportunity cost of Reserve Up provision by **run-of-river (RoR) hydro** is typically high (i.e. the WS price as the marginal cost is zero).[2] RoR hydro provides some Reserve Up in GB1 (0.26 GW/h average by 2035) due to high Reserve requirements in GB1 and limited alternatives, but lower capacities of RoR hydro installed (see slide 68), higher WS prices and lower Reserve Up prices all cause RoR hydro provision to be insignificant in other zones.
  - **DSR** Reserve Up provision is highest in Southern zones (i.e. GB5-7) as higher energy demand in these zones leads to higher available DSR capacity.

Notes: (1) See slide 68 for a detailed breakdown across zones of reserve requirements and capacity available to provide reserve; (2) This implies that when these units provide Reserve Up, the Reserve Up price will be greater than or equal to the WS energy price.

The national model does not restrict reserve procurement based on location, while the zonal model leads to a more geographically balanced procurement

Reserve Up average provision by technology and zone (GW/h)



- The **National model** does not consider location when procuring Reserve up at DA stage, hence reserve Is procured from the **cheapest source** across the country. However, in the **zonal model** we apportion the national reserve requirement across zones in line with zonal demand and wind output (see slides 69-71), so that Reserve is procured **where it is most needed.**
  - For example, in 2035 under National pricing **c. 60% less Reserve Up is procured in GB1 than GB4**, as GB4 has a high capacity of pumped storage and batteries units available to provide reserve cheaply.
  - However, the **reserve requirement under Zonal pricing is c. 50% higher in GB1 than GB4**, as more reserve is required in GB1 to cover high wind output. The zonal model therefore must procure more expensive alternatives in Reserve in GB1, hence explaining the higher prices and overall increased reliance on thermal in the zonal model.
- In the **National model, reserve resources are not necessarily procured where they are most valuable** and where they can be used to meet reserve needs. For example, under National pricing batteries provide high volumes of Reserve Up in the south, but it is uncertainty whether these batteries are of use in RT if there would be low wind output in the North of GB and the transmission grid is congested.[1]
- To allow for comparability, we use the same total reserve requirement in both the national and zonal models. In practice, however, **better geographical targeting** in a zonal model would **facilitate reducing the total reserve requirements**, as set out in the following section.

Note: (1) In practice, ESO's DA AS procurement occurs in National markets (and hence locational aspects are not considered), but RT AS top ups are conducted through the BM, implicitly accounting for network congestion.

Next steps

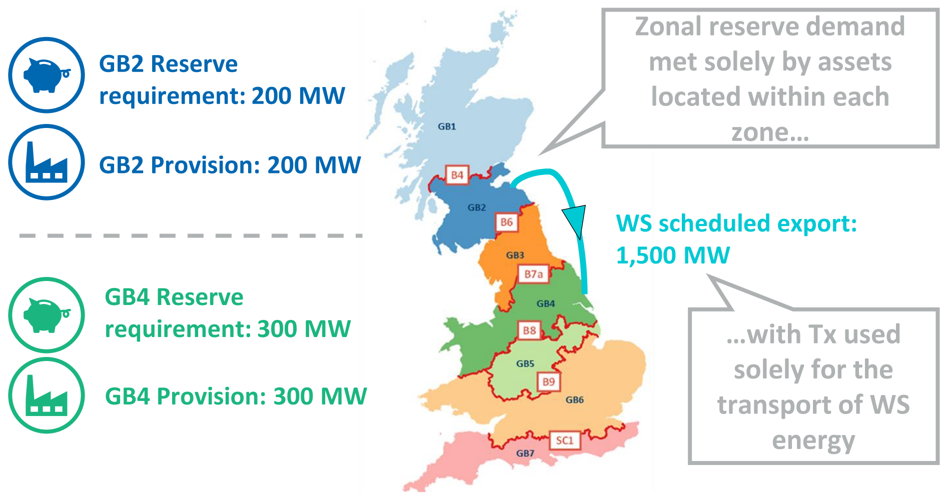
Zonal Reserve markets can provide further benefits through lower Reserve requirements, cross-zonal Reserve procurement and dynamic efficiency gains

	1 Lower cost due to reduced reserve requirements under Zonal Reserve markets	2 Lower cost due to access to cheaper reserves in other zones via the cross-zonal procurement of Reserves	3 Dynamic efficiency via improved price signals under Zonal Reserve markets
National vs Zonal Reserve markets	Benefit of **Zonal Reserve markets**[1]	Highly recommended under **Zonal Reserve markets**	Benefit of **Zonal Reserve markets**
Co-optimisation vs Sequential markets	**Independent** of Co-optimisation or Sequential AS and WS markets	Significantly more efficient allocation of inter-zonal Tx capacity under **Co-optimisation** (see next slide)	Improved price signals under **Co-optimisation** (see Section 4)
What we have modelled	The **same reserve requirements** under National and Zonal markets	**No cross-zonal procurement of Reserves**[2]	**No re-siting** of generation or storage
Why the modelling choice is conservative	Zonal Reserve markets allow for **better geographical targeting in the procurement of reserves** reducing the need to over-procure Reserves to compensate for the likelihood of procuring undeliverable Reserves	Reserve requirements in one zone cannot be covered by units in another zone even though there might be **available inter-zonal transmission capacity** (or it could be cost efficient to allocate transmission capacity for Reserves rather than WS scheduling)	While the most important investment incentives originate from the WS market, **zonal Reserve prices can better inform siting decisions** (i.e. more investment in flexible assets in zones with high Reserve prices)
What can be modelled in the future	Calibration of Zonal-specific Reserve requirements	Cross-zonal Reserve procurement	Allowing for re-siting based on zonal WS and Reserve prices
Expected benefits	Lower AS costs and lower WS prices (due to reduced pressure of the Reserve market on the WS market)	More competition in Reserve markets, reduced AS costs and hence potentially lower WS prices	Overall reduction in system costs; more investment in flexible assets in locations where they are most needed

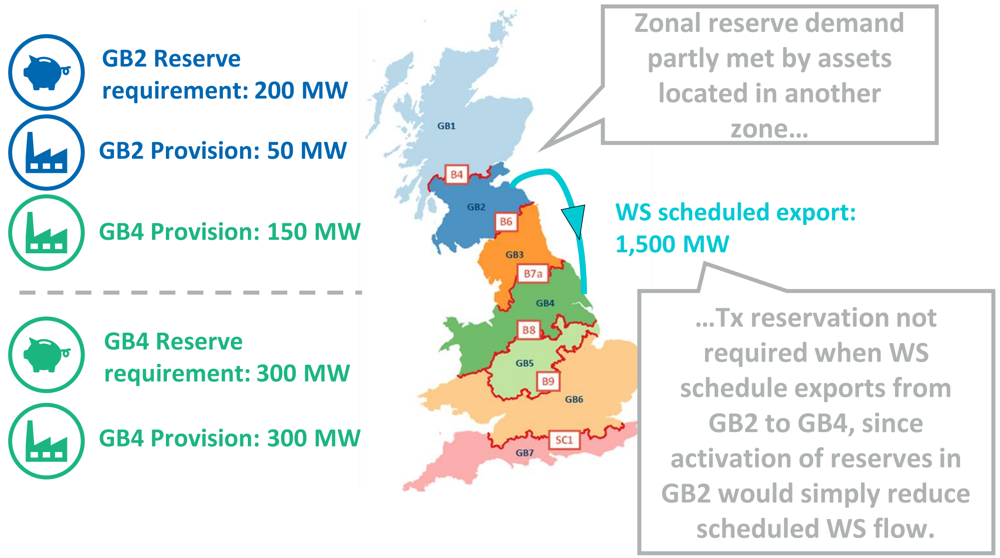
Notes: (1) Zonal Reserve markets also exist under a National WS market design but are rare. Typically, these concern reservations for redispatch purpose, e.g. the German Network Reserve; (2) We have not modelled cross-zonal procurement of Reserves for reasons of computational complexity. In practice, cross-zonal Reserve procurement is in place in the Nordics under Sequential markets (for aFRR – see link) or in US ISOs under Co-optimisation (i.e. the co-optimisation of WS scheduling, AS markets requirement and transmission capacity).

The efficient cross-border procurement of Reserves, reducing AS costs under zonal Reserve market, is conditional upon having Co-optimisation in place

Illustrative clearing of Zonal Reserve Up market, within-zone reserve procurement



Illustrative clearing of Zonal Reserve Up market, cross-border reserve procurement



- The zonal model mandates that all reserve demand be met by assets sited within the relevant zone. For example, as shown in the top-left diagram, GB2 and GB4 each meet their total reserve demand with local assets. Tx between the zones (bootstrap) is used only for energy scheduling in the WS market.
- This might not be the least cost solution in case Reserves are expensive in GB2 while there are low-cost Reserves available in GB4 in excess of the GB4 Reserve requirement. For example, as shown in the bottom-left diagram, lower-cost GB4 generators can provide some reserve in GB2 AS markets, substituting for higher cost GB2 generators.
- However, **cross-border Reserve procurement is conditional upon available inter-zonal transmission capacity**…which depends on the WS schedule.
- In the example on the left, GB2 is exporting to GB4 occupying the entire inter-zonal capacity (1500 MW). No inter-zonal transmission capacity needs to be allocated to Reserves to make the reserved capacity in GB4 deliverable in GB2. Reserve can be provided by (i) GB4 generation increasing; and consequently (ii) GB2 would need to export less energy to GB4 helping to resolve the imbalance in GB2, while not resulting in a shortfall in GB4.
- If the flow on the bootstrap would have been reversed (export from GB4 to GB2), the export scheduled in the WS market for energy would have to be reduced by 150 MW to make the reserve procured in GB4 deliverable to GB2. Whether that would be optimal depends on the **value of using that 150 MW of Tx capacity to reduce AS costs across GB or to use that 150 MW to reduce energy WS scheduling costs.**
- **The trade-off between using inter-zonal capacity for (i) inter-zonal WS energy exchange, and (ii) cross-border Reserve procurement is <u>internalised under Co-optimisation</u>** as WS and AS are simultaneously cleared. In contrast, under **Sequential markets** when the AS markets clear, **<u>the relative value of transmission for both purposes needs to be estimated</u>**, inevitably **leading to inefficiencies** in the allocation of valuable transmission capacity.

Section 6: Real-time activation of AS for energy imbalances

FTI CONSULTING

We qualitatively and quantitatively assess the impact of Sequential vs Co-optimised market designs on real-time costs and find a negligible impact

- As shown in Appendix 3, the re-allocation of resources among markets can lead to different units being reserved under Sequential and Co-optimised procurement.
- For both market designs, we do not model any relationship between a unit's real-time cost of activation and its selection in DA AS markets.
- Also, under a national WS market design, the proximity of the reserved asset to the areas of the grid where imbalances or transmission constraints are most likely to arise is not considered in the procurement process.

In this section we assess whether the re-allocation of resources under co-optimisation has a systematic effect, either upwards or downwards, on the real-time costs incurred by ESO to address energy imbalances. To answer this, we:

1	**Examine the size of the market**	***The RT market is dominated by transmission constraint-driven actions and energy actions have historically accounted for a relatively small proportion of total balancing costs (c. 18% across FY2023/24), as explained on the following slide.***
2	**Explore the theoretical likelihood of a systematic impact**	***With a National WS market there is no locational signal at DA to schedule units to resolve thermal constraints or energy imbalances. We therefore find no theoretical reason why RT costs would be systematically different under Sequential or Co-optimised market designs, as explained on slide 85.***
3	**Quantitatively assess real-time costs**	***We model RT activations under both Sequential and Co-optimised market designs for a 3-week period and find a negligible difference in RT costs (see slide 90).[1]***

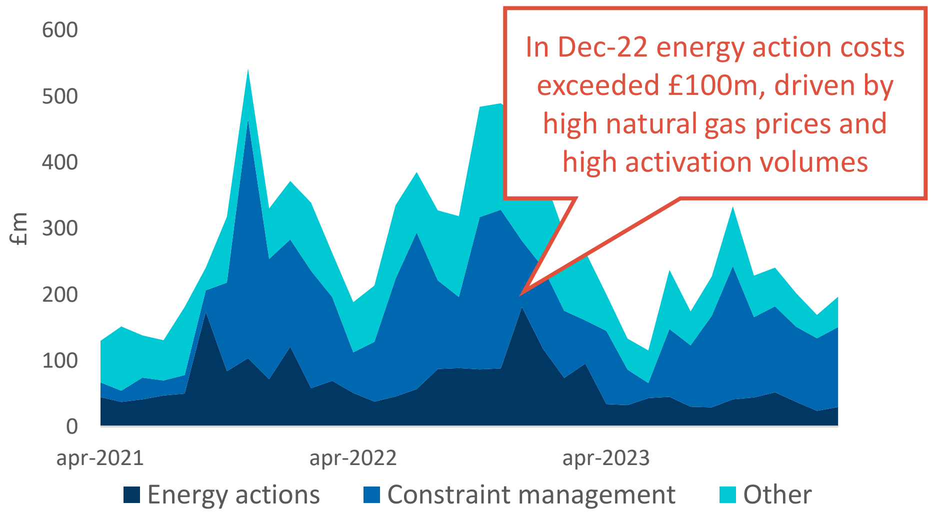
The RT modelling results show a RT cost difference between Sequential and Co-optimised market designs equal to 0.3% of DA consumer costs.

Note: (1) We have not tested a full year, and the modelling results are a reflection of the specific modelling assumptions used and/or weeks selected. Nonetheless, we have identified no theoretical or conceptual reason as to why there should be a systematic impact of co-optimisation on the real-time costs.

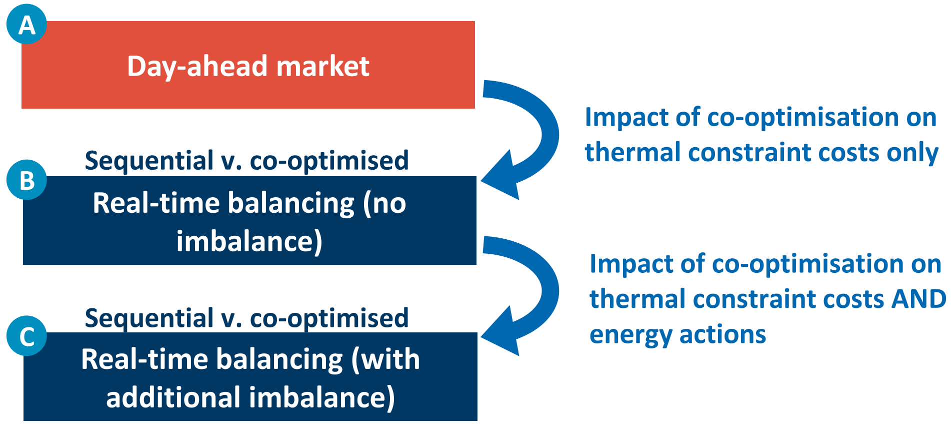
Balancing actions as a result of energy imbalances represent a small share of GB balancing costs; we test RT market outcomes with and without imbalances

- In GB, ESO ensures the system is balanced in RT which requires simultaneously addressing ("implicit co-optimisation") specific needs which can be broadly categorised into:
  - **Energy actions** – ensuring that supply and demand meet at an aggregate GB level.
  - **Constraint management ('Redispatch')** - responding to transmission grid constraints to avoid the overloading of network elements.
  - **Other** – ensuring that other grid security criteria are met (e.g. frequency, voltage, black start…).
- As outlined in slide 24, **energy actions** have historically accounted for a relatively **small proportion of total balancing costs** (c. 18% across FY2023/24). The chart below shows both the **volatility** of energy action costs (imbalances typically arise due to generation / demand forecast errors, which are inherently hard to predict) and the **increasing dominance of constraint management costs**. These costs are ultimately borne by consumers.
- To test the **impact of co-optimisation on the cost of energy actions**, we perform **two real-time model runs** for **each market design**:
  - First, **without an additional energy imbalance**, isolating **constraint management actions** to make the wholesale market schedule physically feasible.
  - Second, with an **additional energy imbalance** (see slide 88 for further detail), testing the **incremental impact of energy actions** on balancing costs.[1]
- We compare outcomes across **three weeks in 2030** to test whether co-optimisation systematically impacts balancing costs. Units successful in DA AS markets are required to be **available** to provide the contracted service in real-time (although **selection** in real-time **depends upon activation prices**).

Monthly balancing costs by category (£m), 2021-24, ESO



Overview of the real-time modelling scenarios



Note: (1) The RT dispatch run's objective function is to minimise the joint cost of actions to solve energy imbalances and constraint management.

DA reserve scheduling does not consider transmission constraints, so there is no theoretical reason for co-optimisation to systematically impact real-time costs

Figure 1: Simplified illustration of a transmission constraint

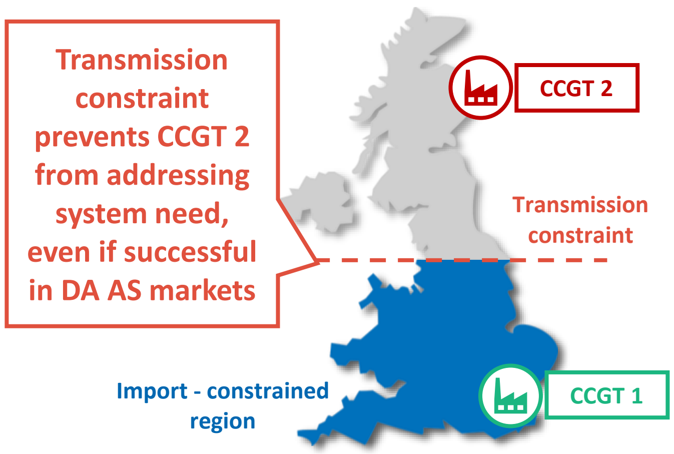
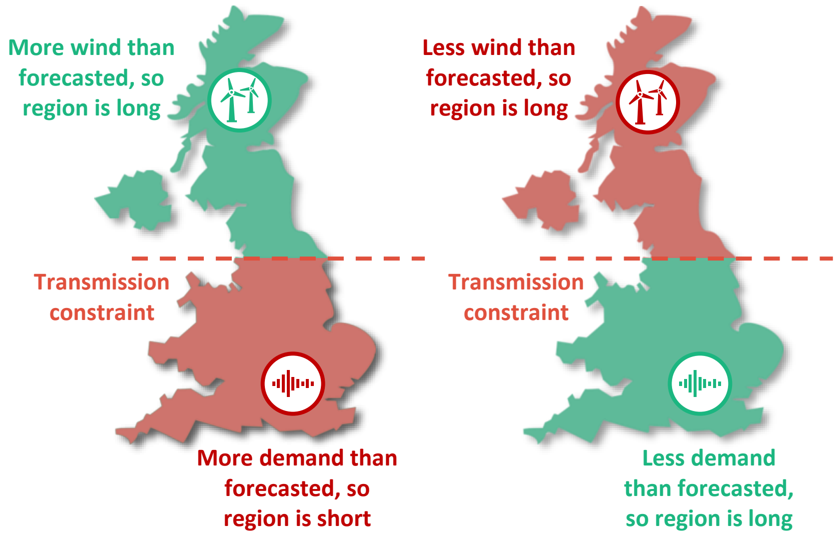


Figure 2: Simplified illustration of regional imbalances



- DA reserve procurement can help reduce the balancing costs incurred in addressing transmission constraints and/or regional imbalances from imperfect forecasts of demand and supply. However, this depends on **where the real-time system need is**, where the **successfully-cleared assets are sited** and the **availability of transmission** between the two.
- For example, in Figure 1, we assume only one of the two CCGTs can be successful in the DA reserve market, but both assets can bid, in RT, to address a system need. Assuming a RT system need arises for more generation (or less demand) in the South of GB, **only CCGT1 can address this RT need** due to a **transmission constraint.**
  - If **CCGT1** had been successful in the **DA Reserve Up** auction, it would be **running (at MSL)**, and hence able to provide a more **competitive RT utilisation offer** compared to CCGT2 (which is assumed to be **offline** and would need to incur **start-up costs**).
  - If **CCGT2** had been successful at **DA** to provide **Reserve Up**, it is likely to be running, but a **transmission constraint** means that **CCGT1 is still required to meet the RT system need** (at a **higher cost**, as in this case **CCGT1** needs to incur **start-up costs**).
  - **Co-optimisation,** by only considering **energy** and **AS** needs (but **not transmission constraints**) does not systematically select CCGT1 or CCGT2 (and neither does sequential design). Without incorporating transmission to the co-optimisation, the **availability of resources** to address RT system needs (and hence their cost) is **not linked** to whether **DA reserves are co-optimised with energy.**
- Similar observations apply to **inaccurate forecasting** of **wind** or **demand** which may cause parts of GB to be **'long'** or **'short'** (respectively too much or too little generation relative to demand, illustrated in Figure 2). For **regional imbalances**, we do not model any **connection** between **DA AS markets** and the potential **location of real-time system needs.** This is reflective of the **current design** of the **GB wholesale market** and **key AS markets.**[1]
- As a result, there is no *a priori* theoretical reason for any **systematic upward** or **downward** impact on balancing costs under either the **Sequential** or **Co-optimised models.**[2]

Notes: (1) In practice, ESO's DA AS procurement occurs in National markets (and hence locational aspects are not considered), but RT AS reserve top ups are conducted through the BM, implicitly accounting for network congestion; (2) As set out in Section 5, including transmission in the co-optimisation through zonal wholesale and AS markets could provide signals at DA to select assets closer to the forecasted distribution of imbalances. We do not test this in our real-time modelling.

Implementation of the methodology

FTI CONSULTING

We model a pay-as-bid real-time market, with bid/offer pricing linked to the opportunity cost of deviating from DA positions

Our BM pricing assumptions incorporate incurred/avoided physical costs, subsidy impacts and offer uplifts, and are applied to all BM actions, whether performed to address thermal constraints or energy imbalances

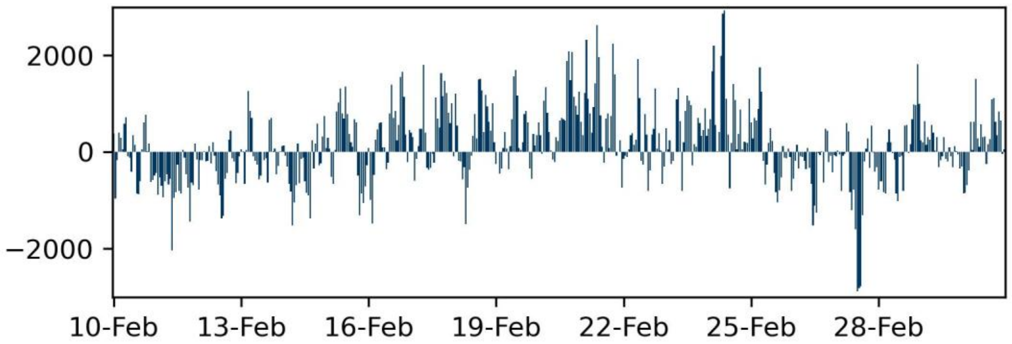
Technology	Cost to ESO Bid (constrained off)	Cost to ESO Offer (constrained on)	
Fossil fuel	- Fuel cost - carbon cost	Offer uplift	+ Fuel cost + carbon cost
Biomass	- Fuel cost	Offer uplift	+ Fuel cost
CCS Biomass	Carbon price – Fuel cost	Offer uplift	+ (Fuel cost – carbon price)
ROCs renewables	ROCs	(theoretical only so no price assumed)	
CfD renewables	CfD strike price – Wholesale price	(theoretical only so no price assumed)	
Merchant renewables	£0	Offer uplift	
Batteries	- Price Paid	Offer uplift	+ Price Received
Other Storage Technology	- Marginal Value	Marginal Value	
Hydrogen generation	- Marginal Value	Marginal Value	
Interconnector	Cost of reversing flow	Cost of reversing flow	

- In our **real-time modelling** (representing the BM), assets are **re-dispatched** to help **address thermal constraints** and (if applicable) **energy imbalances**.
- **Units successful in DA AS markets are online** and **available** to provide the **contracted capacity** in real-time. With **start costs already implicitly reflected in DA bids**, this **increases the units' likelihood of utilisation** in the BM, but only if they are located in areas where they can meet system need.
- The lefthand table summarises the **BM assumptions** used in our real-time modelling – the **cost to ESO** of **increasing** or **decreasing** the **output/load** of different assets to address system needs.
- For **traditional generators**, pricing primarily reflects the **physical costs** incurred or avoided by increasing or decreasing output. An **additional uplift** reflects **historically-observed pricing**.
- For **storage technologies**, pricing focuses on the **medium-term opportunity cost** of using (or gaining) an additional unit of energy.
- Our assessment of balancing costs focuses **purely on consumer impacts and does not reflect any transfers among cohorts** of GB stakeholders (e.g. a change in revenues for producers), for consistency with current stated ESO methodology.

We model an imbalance profile to analyse RT activations which includes wind and demand forecast errors (+/-), randomly distributed across nodes

- We **introduce imbalances to the model in RT to simulate wind and demand forecast errors**, enabling an analysis of how these imbalances are resolved under Sequential and Co-optimised markets.[1]
- We calculate imbalances on a node-by-node basis, as explained below. **Wind forecast errors** are **grouped by zone** (i.e. all wind generators in each zone make the same error), but **load errors are uncorrelated.**
- We then **compare the national imbalance to historical 2022 data** (assumed to hold constant across the modelling period) and redistribute the shortfall/surplus to align the total GB imbalance with historical data.

Imbalance profile, 10 February – 2 March 2030 (MW)[1]

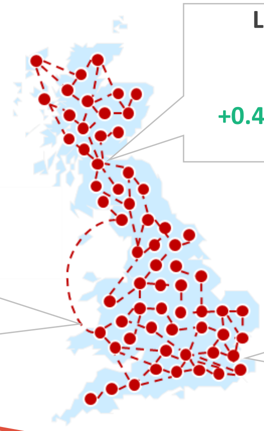


Nodal imbalance methodology

1. Develop **node-by-node forecast error** for **demand** and **wind**; normally distributed with 95% hours within **+/- 5%** of forecast
2. **Sum nodal imbalances** to find **GB aggregate** imbalance…
3. …and **compare** to **historical 2022** GB imbalance to find **net shortfall/surplus**[2,3]
4. **Redistribute shortfall/surplus** across nodes (**weighted by load**) to **align total GB imbalance** with historical data
5. Apply net node-by-node imbalance **across each node** in the **real-time model**

Note: this causes **imbalances** to generally be concentrated in areas with **higher demand** or **wind** output

Example nodal imbalance adjustment – 11am, 23 February 2030



Livingston East imbalance:
+1.2 MW (Load)
+3.7 MW (Wind)
+0.4 MW (Historical adjustment)
+5.3 MW (Net)

GB imbalance:
+847 MW (Historical)
+442 MW (Model)
+405 MW (Historical adj.)

Canterbury imbalance:
-4.1 MW (Load)
-4.4 MW (Wind)
+1.4 MW (Historical adj.)
-7.1 MW (Net)

Initial model imbalance is 405 MW lower than historical GB imbalance, requiring load-weighted adjustment of nodal imbalances.

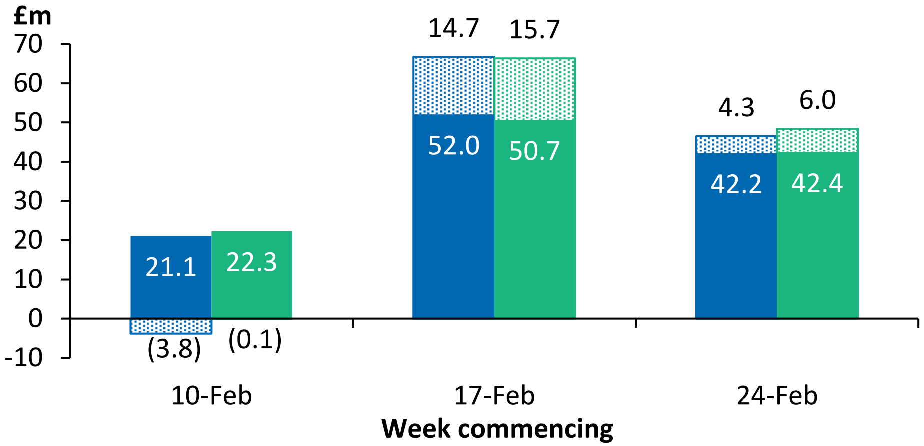
Notes: (1) In the absence of such forecast errors, the DA schedule and RT dispatch would differ only in terms of actions taken to respond to transmission constraints; (2) Modelled imbalance profile equals historical profile from Elexon (link); (3) Before calculating the shortfall/surplus, we match the historical and modelled imbalances to their closest matching hour, to ensure the net GB position in each modelled hour follows a similar pattern to the historically-observed outcome.

Modelling results

FTI CONSULTING

The difference in balancing costs between the two market designs is arbitrary and insignificant relative to the difference in day-ahead consumer costs

Weekly real-time balancing costs, Sequential and Co-optimised, without and with imbalance, 10 February – 2 March 2030 (£m)



Total consumer costs, Sequential and Co-optimised, without and with imbalance, 10 February – 2 March 2030 (£m)

	Day-ahead costs	Real-time balancing costs	
	WM + AS	Thermal constraints	Energy actions (imbalance)
Sequential	2,094	115.3 *(5.5% of DA costs)*	15.3 *(0.7% of DA costs)*
Co-optimised	1,998	115.4 *(5.7% of DA costs)*	21.7 *(1.1% of DA costs)*
Sequential less Co-optimised	96	(0.1)	(6.5) *(0.3% of DA costs)*

Difference in energy action costs is c.0.3% of total DA consumer costs

- The charts above compare total balancing costs under each market design without an imbalance profile (i.e. solely the cost of resolving thermal constraints) and with an additional imbalance (highlighting the incremental impact of addressing energy imbalances).
- **Under both market designs, balancing costs are dominated by thermal constraints.** The proportion of energy action costs to total balancing costs (12-16% in our results) broadly aligns with the 18% we observe in historical data (see slide 24).
- The cost impact of imbalance-driven energy actions varies depending on whether the GB system is net 'long' or 'short'. In the first modelled week, the imbalance generally causes GB to be net 'long', coincidentally reducing broader balancing costs by reducing the volume of constraint actions required.[1] In Weeks 2 and 3, GB is often net 'short' and additional generation must be activated in real-time on top of that required to address thermal constraints.
  - On thermal constraints: Across the modelled weeks, we find **no systematic bias towards sequential or co-optimised** – constraint management costs are marginally higher with co-optimisation in the first and third weeks, and marginally lower with co-optimisation in the second week.
  - On energy imbalances: energy imbalance costs across the weeks are £6.5m higher under co-optimisation (c.0.3% of total DA consumer costs), driven by differences in the size and location of assets cleared in DA AS markets relative to the nodal imbalance.[2] However, as explained in slide 85, there is no theoretical basis to expect real-time energy costs to be systematically different under Co-optimised relative to Sequential across a longer timeframe.

Notes: (1) The mis-forecasting of energy or demand does not, in isolation, reduce consumer costs, but can lead to negative cost outcomes if the resulting imbalance reduces the requirement for constraint management actions (e.g. greater wind output than expected at DA in the South may mean that less generation needs to be constrained on in RT). (2) The observed difference in total balancing costs between the two market designs forms 6.8 % of the consumer cost savings delivered by co-optimisation at DA, and 0.3% of total DA consumer costs. In essence, consumer cost impacts from co-optimisation are dominated by changes in DA market outcomes.

Appendices

List of appendices

Appendix 1 – Assessment of key modelling assumptions and limitations

We make four critical modelling choices in defining the counterfactual sequential scenario, which cause our estimates to be conservative in all modelled years

Modelling choice	Description of assumptions	Implications for estimated benefits of co-optimisation
1 Sequence of modelling DA AS and DA energy markets	**We run the AS reservation markets ahead of the DA WS energy market.** This is most representative of current arrangements[1] and implies bidding in the DA AS market based on opportunity costs from the DA energy market, which is arguably less complex for participants than if DA energy market was cleared ahead of DA AS markets.[2]	⬆ Historically, the time lag between the AS reservation and DA WS **energy market was typically longer.[3]** A longer time lag would introduce **greater forecast error** to the opportunity cost estimation under sequential design, thereby increasing the estimated benefits of co-optimisation. As such, our assumption is conservative.
2 Modelling of AS markets at DA time frame	The **DA AS market** only has **one market for response**, and **one for reserve**, rather than three response and three reserve markets.	⬆ Historically, pre-EAC, **greater number of markets** made it harder to **efficiently allocate resources** across them. As such, our assumption is conservative for historical modelling.
	DA AS products are **hourly.**	⬆⬇ Historically the product length for AS has been **multiple hours**. Estimating the opportunity costs across multiple hours is harder but **start-up costs** are easier to reflect in **multi-hour products**.
3 Co-optimisation of AS markets at DA time frame	We assume the **co-optimisation of the DA procurement of response and reserve products in the DA time frame ("Partly co-optimised")**. This assumption holds since the introduction of the EAC (2023) and is therefore appropriate for the forward-looking analysis. To limit model complexity, we also apply this assumption to the historical analysis.	⬆ **Co-optimisation of frequency and response** leads to a **more efficient counterfactual** than the **former arrangement**. As the overlap of resources competing in response and reserves markets is relatively small, the impact of co-optimising their procurement can be expected to be limited.
4 Modelling of RT markets	In the RT timeframe we assume **full co-optimisation of energy, redispatch and reserve markets.[4]**	⬆ Currently **the BM applies heuristics** to combine all available resources to solve **RT energy imbalances**, **redispatching needs** and **reserve system requirements ("implicit"** – see slide 16). These heuristics are likely less efficient than co-optimisation and, as such, our assumption is conservative.
Legend	⬆ Benefits of co-optimisation likely to be higher in practice than in our modelling	⬆⬇ Benefits of co-optimisation may be higher or lower in practice than in our modelling

Notes: (1) Although some aspects of the current GB design are different – e.g. frequency services bought post DA energy auction; as a simplification we do not model this; (2) This is because opportunity costs are easier to estimate when DA AS markets clear first, hence leading to lower information asymmetry; (3) One month ahead for STOR; (4) The activation of frequency response happens automatic and is not modelled here.

The historical model is not intended to fully replicate outcomes from 2022, but rather to reflect the prevailing market conditions in a conservative manner

Key assumptions and limitations

Our assumptions are generally expected to lead to a conservative estimate of the impacts of co-optimisation

	Applies to	Description of assumptions	Implications for estimated benefits of co-optimisation
Bidding methodologies	Counterfactual	We apply an opportunity cost estimation methodology that uses different estimations of the DA energy prices	↑↓ No clear impact of not perfectly replicating historical bidding behaviour
Consolidated AS markets	Both counterfactual and co-optimised runs	We consolidate the multiple AS markets into four – Response Up, Response Down, Reserve Up and Reserve Down	↑ Simplified market structure reduces information asymmetries in the counterfactual
Imperfections in markets behaviour	Both counterfactual and co-optimised runs	We do not consider imperfections such as gaming of BM, market power, and (potentially flawed) human decision-making	↑ Gaming in DA AS markets is harder in co-optimised markets and fewer human decisions need to be made
Perfect foresight	Both counterfactual and co-optimised runs	Perfect foresight in the RT timeframe over the entire day when activating resources to solve energy, reserve and redispatch issues. For example, batteries are expected to be utilised more often in RT compared to historical observations	↑↓ No clear impact of perfect foresight in the RT timeframe as it applies to both the sequential and co-optimised models
Representation of RT deviations	Both counterfactual and co-optimised runs	We model system imbalances in RT but do not attribute what specific units are deviating in RT from their DA market schedule	↑↓ No clear impact of the representation of RT deviations

Legend

↑ Benefits of co-optimisation likely to be higher in practice than in our modelling

↑↓ Benefits of co-optimisation may be higher or lower in practice than in our modelling

Appendix 2 – Calculation of total system costs

Total system costs equal the DA scheduling costs plus a correction derived by computing the difference between consumer costs and producer surplus

- The equations below provide additional detail on the calculation of the two main evaluation metrics of the merits of co-optimisation between energy and ancillary services we examine in this report: consumer costs and total system costs, building on the summary presented in the main body of this report.
- The equations also set out how the total system costs, consumer costs and producer surplus are interrelated.

$$\text{Consumer costs} = \underbrace{\text{DA load} \times \text{DA wholesale price} + \text{DA wholesale uplift}}_{\text{DA GB energy scheduling costs}} + \underbrace{\text{DA AS reserved volume} \times \text{DA AS reservation price} + \text{DA AS uplift (under co-optimised)}}_{\text{DA GB AS reservation costs}}$$

We assume half of the IC rents accrue to the GB-side of the IC owner

$$\text{Producer surplus} = \text{DA scheduled volume} \times \text{DA wholesale price} + \text{DA wholesale uplift} - \text{DA scheduling costs} + \text{DA AS reserved volume} \times \text{DA AS reservation price} + \text{DA Uplift (under co-optimised)} + \frac{1}{2}\,\text{DA IC rent}$$

$$\text{Total System Costs} = \text{Consumer costs} - \text{Producer surplus} = \text{DA scheduling costs} + \left(\text{DA load} - \text{DA scheduled volume}\right) \times \text{DA wholesale price} - \frac{1}{2}\,\text{DA IC rent}$$

Substituting:

$$\text{DA IC rent} = \left(\text{DA wholesale price} - \text{Overseas DA price}\right) \times \text{Import} + \left(\text{Overseas DA price} - \text{DA wholesale price}\right) \times \text{Export}$$

$$\text{DA scheduled volume} = \text{DA load} + \text{Export} - \text{Import}$$

Results into:

$$\text{Total System Costs} = \text{DA scheduling costs} + \boxed{\text{Average of GB and overseas DA wholesale price} \times \left(\text{Import} - \text{Export}\right)}$$

A system with lower/higher DA prices will export/import more, having a direct impact on the DA scheduling costs in GB. This term corrects for this effect.

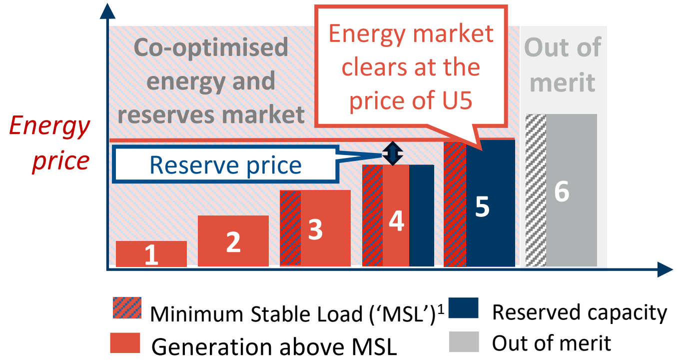
Note: (1) The formulas displayed here should be computed on an hourly basis – e.g. there will not be negative IC rents. For simplicity we disregard storage in this formulation. Storage revenues are included in the producer surplus in our results. Storage withdrawals/injections are not included in DA load but subtracted form/added to the DA scheduled volume.

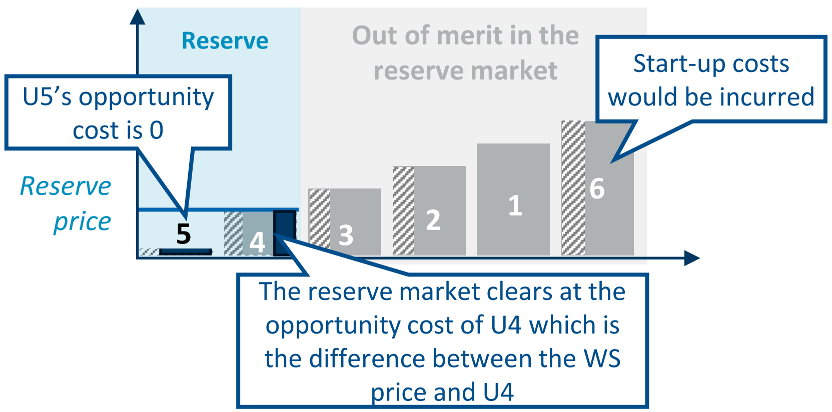
Appendix 3 – Price formation dynamics

By design, co-optimisation delivers an efficient schedule: prices in DA AS reservation markets reflect conditions in DA WS energy markets **and vice-versa**

- To illustrate the difference between co-optimised and sequential procurement of energy and ancillary services, we assume the market is served by **six generation units**.
- To meet the market's demand**, four units of energy** and **one unit of upward reserve** are needed.
- On this slide we describe the market outcome under Co-optimisation. The next three slides show how different bidding scenarios in AS markets under Sequential procurement can bring higher, equal or lower WS energy costs under Sequential markets. These scenarios are illustrative, not exhaustive.

Co-optimised Energy and Reserve





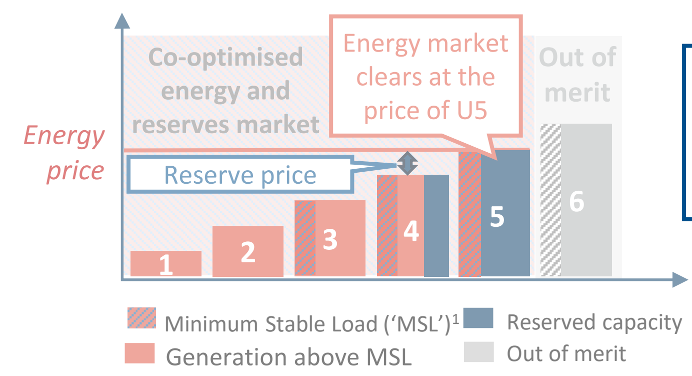
- The objective under **Co-optimisation is to minimise scheduling costs subject to AS requirements**.
- The least-cost schedule that satisfies AS requirements is the following:
  - U1, U2 and U3 are scheduled to operate in the WS market at their maximum capacity.
  - U4 is scheduled for 2/3 units and provides 1/3 unit of reserves.
  - U5 is scheduled at its MSL in the WS market, i.e. at 1/3 unit, and provides 2/3 reserve.
  - Not reserving U4 would require starting up U6 in addition to U5 to provide reserves which would be a more costly solution.
- The **prices in the DA WS and AS market are interlinked**:
  - The least-cost solution to satisfy a marginal increment in the WS energy market is to marginally increase the scheduled generation of U4.[2]
  - But to be able to schedule that marginal increase in generation, U4 would have to free up some reserved capacity.
  - The value of that freed-up reserved capacity equals the AS reserve price.
  - Hence the **WS energy price equals the SRMC of U4 plus the AS reservation price.**
  - **The AS reservation price is set by the opportunity costs of U4** which equals the difference between the WS price and the SRMC of U4.
  - An **equilibrium**, where no unit is losing out (relative to not running), occurs when the **WS energy price equals the SRMC of U5**, and hence the **AS reserve price equals the SRMC of U5 minus the SRMC of U4.**

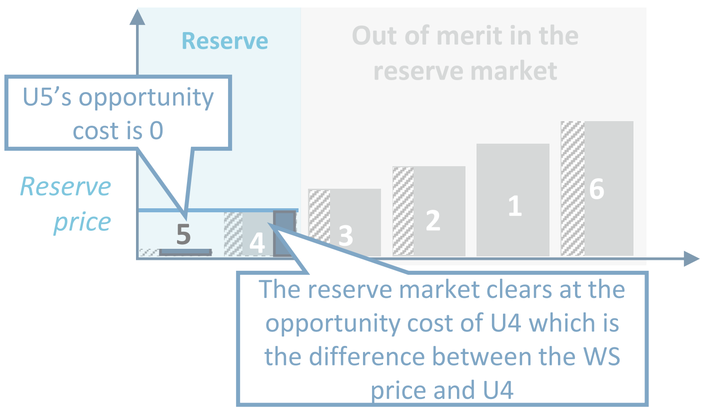
Notes: (1) We assume an MSL equal to 1/3 of the max. capacity; (2) Here we disregard discontinuities in the pricing – alternatively we can assume that the total demand equals slightly less than four units.

Due to inaccurate WS price forecasts, sequential markets can lead to inefficient WS scheduling and an increase in WS consumer costs

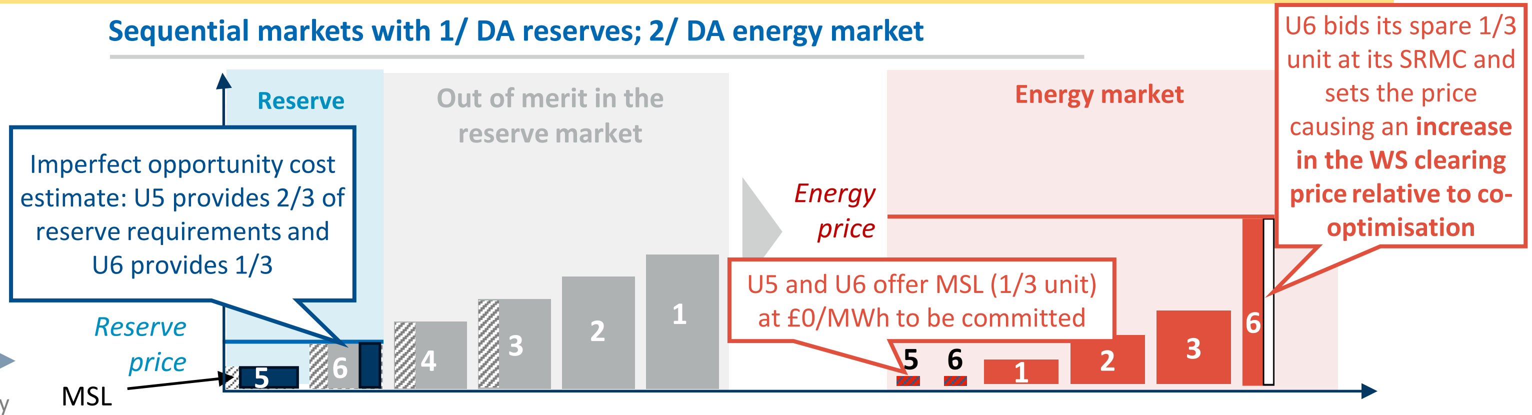
Here we illustrate how **imperfectly estimated opportunity costs under Sequential** markets can lead to **an inefficient schedule** and to a **WS energy price increase** (hence WS consumer savings from Co-optimisation).

Co-optimised Energy and Reserve





Sequential markets with 1/ DA reserves; 2/ DA energy market



DA AS market

- U6 and U4 anticipate the WS energy price to be higher than under Co-optimisation:
  - U6 (wrongly) assumes it will be in-merit in the WM, so its AS bid does not have to account for its start-up costs; its AS offer price decreases compared to its costs to be reserved under Co-optimisation.[1]
  - U4 increases its AS bid as its expected foregone WS revenues are (incorrectly) higher.
- Consequently, U6 gets selected in the reserve market rather than U4. **U6's AS bid sets the AS price**.

DA energy market

- U5 and U6 bid in their MSL at £0/MWh to be sure to be committed and able to provide reserves. U1, U2 and U3 bid in their marginal costs and are fully cleared in the WS market.
- The least cost solution to satisfy the demand is to accept the offer of U6 at its SRMC. U5 is fully reserved beyond its MSL and U4 would have to be started-up to be scheduled which would be more costly.
- The **WS energy price is set by the SRMC of U6.**

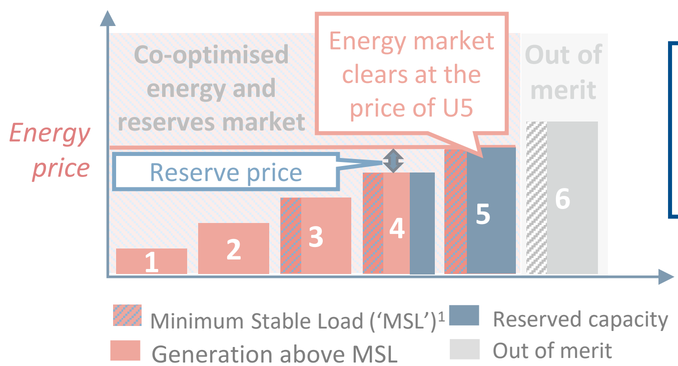
The resulting **WS energy price under Sequential Markets is higher than under Co-optimisation** due to an overestimate of the WS price (relative to the WS under Co-optimisation) by some market parties. The **inaccurate WS price forecasts eventually also materialise ("self-fulfilling prophecy")**. The dispatch schedule is costlier than under Co-optimisation. This implies that Co-optimisation **leads to consumer WS savings and overall efficiency gains.**

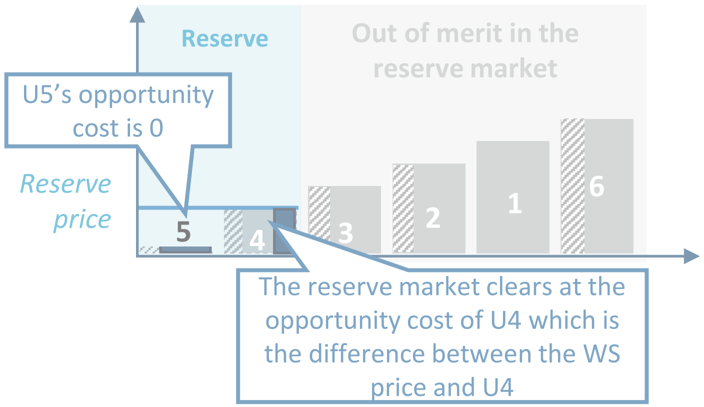
Note: (1) Not necessarily the entire start-up costs would be internalised in the reserve bid. If a unit would be in-merit before, the reserve bid would represent the costs of a delayed shutdown. Conversely, in case the unit would be online later in the day, the costs for starting up earlier would be reflected in the bid (i.e. the loss for running at MSL at low WS priced hours).

When some units make wrong WS price forecasts, sequential markets can lead to inefficient WS scheduling without an impact on WS consumer costs

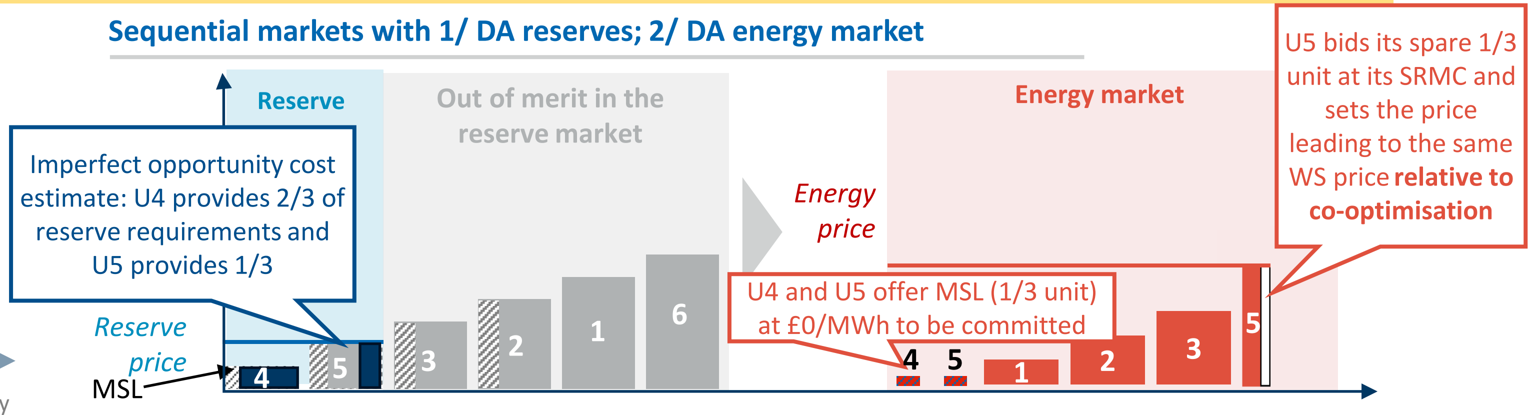
Here we illustrate how **imperfectly estimated opportunity costs under Sequential markets** can lead to **an inefficient schedule** (increasing system costs) and to **equal WS energy prices** (hence no WS consumer savings nor losses from Co-optimisation)

Co-optimised Energy and Reserve





Sequential markets with 1/ DA reserves; 2/ DA energy market



DA AS market

- U4 expects the WS energy price to be lower than in the Co-optimised model (and vice versa for U5):
  - U4's AS bid decreases compared to its opportunity costs to be reserved under Co-optimisation.
  - U5 increases its AS bid; expected foregone WS revenues are higher under a higher WS price.
- Similarly to the Co-optimised reservations, U4 and U5 are in reserve but in different proportions. The reserve price is set by U5's AS bid.

DA energy market

- U4 and U5 bid in their MSL at £0/MWh to be sure to be committed and able to provide reserves. U1, U2 and U3 bid in their marginal costs and are fully cleared in the WS market.
- The least cost solution is to also accept the offer of U5 at its SRMC. U4 is fully reserved beyond its MSL.
- The **WS energy price is set by the SRMC of U5 and hence equal as under Co-optimisation, but the schedule is more costly as more generation of U5 is scheduled rather than U4.**

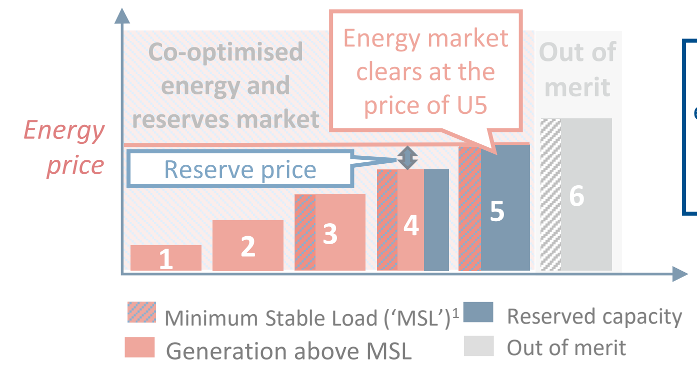
The **resulting WS energy price under Sequential Markets is the same as under Co-optimisation** while some units wrongly estimate their opportunity costs. The final dispatch schedule incurs higher system costs than under Co-optimisation, implying **efficiency savings and increased producer surplus under Co-optimisation.** There are **no consumer WS savings nor losses from Co-optimisation in this scenario.**[1]

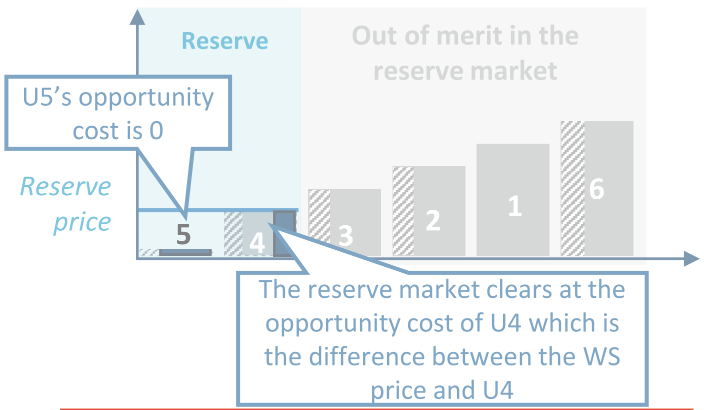
Note: (1) There may be a difference in the clearing AS reservation price in this scenario, depending on the exact bids by U4 and U5 under Co-optimisation and Sequential. A lower/higher AS reservation price under Co-optimisation would reduce/increase WS consumer costs from Co-optimisation. We do not examine this in the illustrative scenario above.

Sequential markets can lead to no impact on WS scheduling and decreased WS consumer costs, but this is not a likely outcome unless AS markets are tight

Here we illustrate how **imperfectly estimated opportunity costs under Sequential markets** can (in limited circumstances) lead to **the same schedule** and to **lower WS energy prices** (hence a WS consumer cost from Co-optimisation).

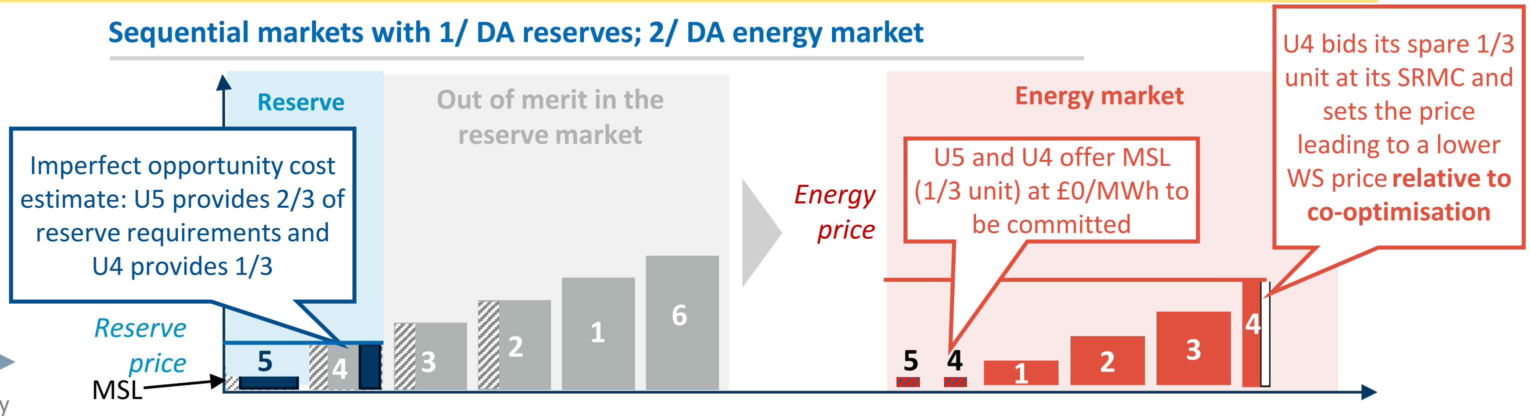
Co-optimised Energy and Reserve





As discussed in Section 4, this scenario can be observed in a **system with tight AS markets** (and thus high AS prices relative to WS prices)

Sequential markets with 1/ DA reserves; 2/ DA energy market



DA AS market

- All units expect the WS energy price to be the same as under Co-optimisation (i.e. set by unit 5).
- The same reservation as under Co-optimisation occurs and the price is set by the AS bid of U4.

DA energy market

- U5 and U4 bid in their MSL at £0/MWh to be able to provide reserves. U1, U2 and U3 bid in their marginal costs and are cleared in the WS market. The offer of U4 at its SRMC to satisfy demand.
- The **WS energy price is set by the SRMC of U4 which is lower than under Co-optimisation, while the reservation schedule and the final dispatch is the same.**
- As a result, **U5** in the example above is committed to provide AS, and subsequently turns out to be **worse off compared to if it chose not to run at all in either market**. In practice U5 is likely to learn from this experience, making such outcome not highly likely to sustain.[1]
- This paradoxical result occurs as under Co-optimisation WS market prices can reflect AS market prices, while **such feedback link is not present under Sequential markets.** Specifically, the difference between the WS price under Co-optimisation and Sequential markets is **driven by a unit that is reserved under Co-optimisation being at the same the marginal WS unit while the AS price is non-zero.**

The resulting WS energy price under Sequential Markets is lower than under Co-optimisation while all units estimate the same opportunity costs that resulted under Co-optimisation. The final dispatch schedule is the same under Co-optimisation, implying **no efficiency savings but increased producer surplus under Co-optimisation at the expense of increased WS consumer costs.**

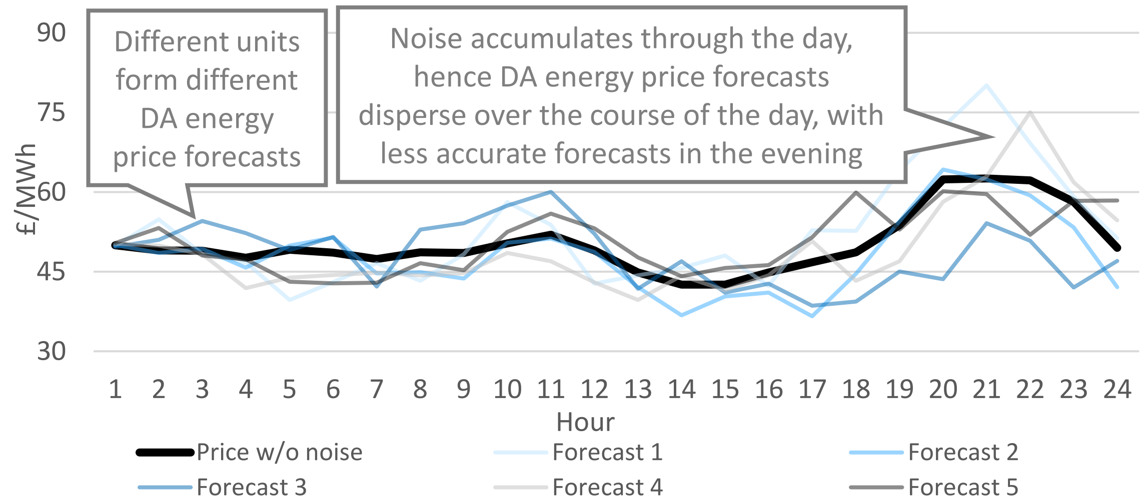
Note: (1) To reflect this, we introduce minor noise in the WS market price estimation and add a risk premium to the bidding of thermal units - see Appendices 4-5.

Appendix 4 – Noise estimation in the Sequential model

We add noise to each unit's DA energy price estimation, which changes every day for every generator

- We add noise to each unit's DA energy price forecast, to quantify the impact of imperfect WS price forecasting on the allocation of resources between WS and AS markets under a sequential design, and hence the benefits of co-optimisation when the need for making such imperfect forecasts is avoided.
- This is illustrated in the chart below, which shows:
  - **Price w/o noise** – i.e. the DA energy price units would expect if they formed accurate expectations of AS reservations, demand, weather patterns etc.; and
  - **Forecasts 1-5**, which represent the forecasts of five separate market participants, introducing a degree of imperfect forecasting.
- Noise across DA energy price forecasts leads to imperfect estimates of opportunity cost of participating in DA AS markets, hence leading to imperfect bidding in AS markets and imperfect AS reservations, hence distorting units' behaviour in the DA energy market.

Application of noise on an indicative price curve



Methodology

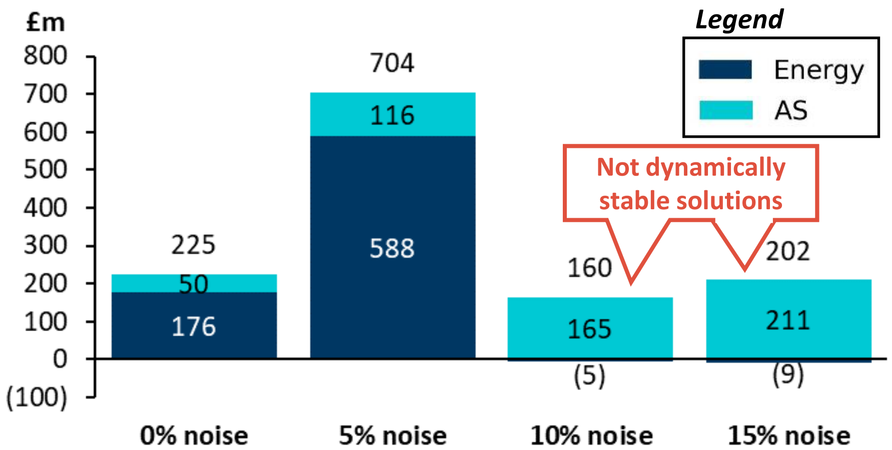
- We introduce "noise" to the sequential model to **account for the imperfect foresight** inherent for a unit bidding in a sequential system – i.e. individual units' inaccuracy in forecasting DA energy price *at the time of* bidding to provide AS.
- The modelled noise exhibits the following key characteristics:
  1. Each unit's noise is **correlated across hours within the day** (i.e. underestimating the price in one hour makes that unit more likely to also underestimate the price in the next hour). This causes noise to accumulate throughout the day, thus resulting in less accurate DA energy price forecasts in the evening, as it is more difficult to forecast over longer time horizons.
  2. Noise is **not correlated across days** (i.e. underestimating prices one day does not make that unit any more or less likely to underestimate or overestimate prices the next day).
  3. Except for storage units,[1] each unit's noise is **independent** of other noise profiles.
  4. The **average DA energy price forecast** across units is set to equal our modelling forecast.[2]
- We run a **sensitivity analysis for 2030** with differing levels of noise, as explained on the next slide.

Notes: (1) Storage units are grouped into portfolios to estimate DA energy prices (in order to reduce computational complexity), and hence storage units within the same portfolio share the same noise profile; (2) To simplify the modelling process, we use the co-optimised DA energy prices as the basis for the sequential forecasts.

Co-optimisation yields significant consumer cost savings under all noise sensitivities we modelled for 2030

- We tested the sequential model for **2030 with four different noise levels** – 0%, 5%, 10% and 15% Mean Absolute Percentage Error ("MAPE").
- **Under all noise sensitivities, co-optimisation yielded significant consumer cost savings (£160m - £704m in 2030)**, as shown in the chart on the right.
- The results in this report use the **low noise (5%) scenario**. This was chosen to provide the best balance between having "enough" noise to generate sequential AS prices which broadly align with historical prices and not having "too much" noise in which case thermal units would make economically incoherent decisions, making the modelling outcome less credible, as explained in the boxes below.
- Nonetheless, if we had used any of the other noise scenarios in our modelling then we would still expect **total consumer cost savings from co-optimisation of over £1bn (PV, 2025-35).**[1]

Consumer cost savings (£m) 2030 – Sequential (noise sensitivities) vs Co-optimised



Why not have <5% noise?

- **Increasing noise increases average AS prices**, as units make less accurate WS opportunity cost estimates. This can be seen in the increasing size of the light blue bars of the chart above as noise increases.
- **Even with 5% noise our sequential model underestimates historical AS prices**[2], so having even lower levels of noise (e.g. 1%, 2%, etc) would likely create an even wider gap between observed historical AS prices and those modelled in the sequential design.
- More generally, having **0% noise** (i.e. all units make identical and perfect opportunity cost estimates all the time) is an **unreasonable assumption**.[3]

Why not have >5% noise?

- As noise levels increase, it becomes more likely that **thermal units cannibalise their own profits**.
- With high noise, out-of-merit thermal units are more likely to mis-estimate their opportunity cost and bid into AS markets expecting themselves to be in merit in the WS market. If enough units make this mistake then the **'free' energy** provided by these price-taking units generating at MSL can be large enough to **lower the WS price**.
- This leads to significant outturn losses for these thermal units, making their original **WS price forecasts inconsistent with outturn WS prices**. Hence the significant WS energy cost to load savings in the 0% and 5% noise runs are not replicated in the medium and high noise runs.
- In reality, however, these units would learn from their mistakes and change their AS bidding strategy. Therefore, we **do not consider the 10% or 15% noise runs to be credible outcomes** as they **do not present a dynamically stable solution**.
- As a secondary point, the cost to load savings in the 10% and 15% runs arise **solely from AS markets**, indicating that despite co-optimisation improving allocation of resources between AS and WS markets, there is no benefit in reducing observed wholesale prices. This does not accord with the economic intuition explained earlier in the report.[4]

Notes: (1) £1.6bn PV (2025-35) with 0% noise, £1.1bn with 10% noise and £1.4bn with 15% noise. Approximated using the £4.9bn cost saving under 5% noise, apportioned for the relative cost savings in 2030; (2) See table on slide 122; (3) 0% noise leads to higher costs than co-optimisation because market participants cannot perfectly reflect their opportunity costs in AS bids, which may not be optimal on a market-wide level (e.g., storage units face power opportunity costs, see slide 109); (4) See, for example, slides 44, 46 and 100.

Appendix 5 – Opportunity cost methodologies in the Sequential model

The calculation of the opportunity cost of a thermal unit depends on its own technical parameters and expectations of the day-ahead dispatch and prices

Thermal generator bidding summary

Reserve direction of bid	Generation commitment expectation[1]	Hourly opportunity cost of reservation (£)				Hourly bid volume (*MW*)
Downward	Unit does not expect to be committed	*Start-up costs*[(2)] + (*MSL* + *downwards bid volume*) × (*SRMC* − *day-ahead price estimation*)				*Min. of {installed capacity, ramping speed × activation time}*
Downward	Unit expects to be committed	Zero				
Upward	Unit does not expect to be committed	*If*	*activation time* < *start-up time*	*then*	*Start-up costs*[(1)] + *MSL* × (*SRMC* − *day-ahead price estimation*)	
Upward	Unit does not expect to be committed	*If*	*activation time* ≥ *start-up time*	*then*	Zero	
Upward	Unit expects to be committed	(*day-ahead price estimation* − *SRMC*) × *upward bid volume* + *Start-up cost risk premium*				

With variable amounts of noise

Rarely happens

Key:
- *Likely accepted*
- *Possibly accepted*
- *Unlikely accepted*

Inframarginal plants are most likely to be accepted in upward reserves, as **their opportunity costs are estimated as the sum of the expected profit margin and a risk premium**

The 1 second activation time for response means thermal units are only eligible to provide reserve

Methodology

- A thermal generator's opportunity cost estimation mainly depends upon:
  - its **merit order expectations** (committed or not and profit margin); and
  - its **technical parameters** (start-up costs, MSL, installed capacity, ramp rate).
- Thermal units will then estimate their opportunity costs, based on expected **foregone revenues and/or technical costs of being available for activation**.
- Additionally, a **risk premium**[3] is added to reflect the start-up uncertainty of thermal generators' (to account for the fact that a plant can expect to be committed and be reserved up for roughly its expected profit margin, but if it doesn't end up in the day-ahead merit order it will incur start-up costs that haven't been internalized in any way).

Notes: (1) The input for this expectation is based on the Co-optimised model; (2) Start-up costs are allocated in all consecutive hours after a shutdown and before a start-up, which can be considered a risk-averse bidding strategy. Also, if the cost of anticipating an expected start-up and running the plant is cheaper than having two separate start-ups, the start-up cost will equal the former; (3) The risk-premium is proportional to the SRMC and is added onto the bid when the forecasted price is similar to the generators' SRMC.

The opportunity costs of reserving storage units result from: (A) missing a key hour for an arbitrage; and (B) having the required SoC to provide the service

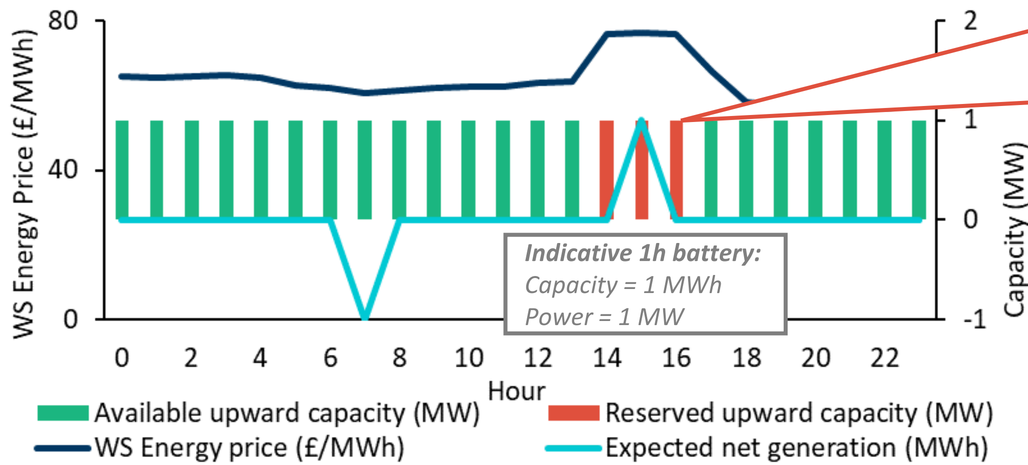
Methodology

- Reserving a storage unit in one hour of the day **may impact its functioning for the rest of the day.** This is because for a storage unit to be reserved it must:
  - Have **available capacity to either accept or release energy**, depending on the direction it is reserved ("power" implications); and
  - Have a **level of energy already stored or readily available**, in case activation is required ("energy" implications).

In the next two slides we examine in more detail the difference between:

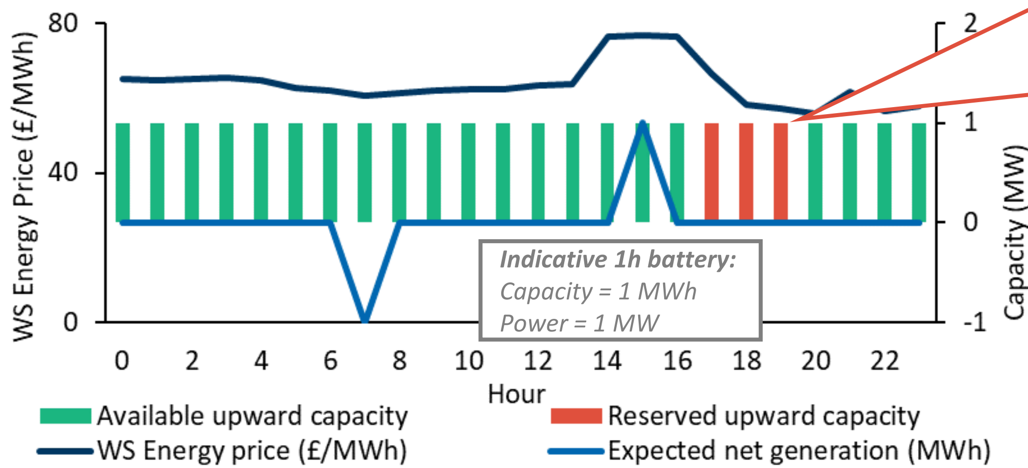
- **Power-related opportunity cost,** which is positive during the **key hours for an arbitrage;** and
- **Energy-related opportunity cost,** which is positive during hours when **energy requirements** (derived from being reserved) are **not aligned with the expected DA dispatch** of storage units.

A **Reserve Up – Power opp. Cost (£/MW) – Battery reserved during peak**



If the battery is reserved upwards **during** the peak hour, it **won't be available to discharge** and will **miss the full arbitrage revenues**

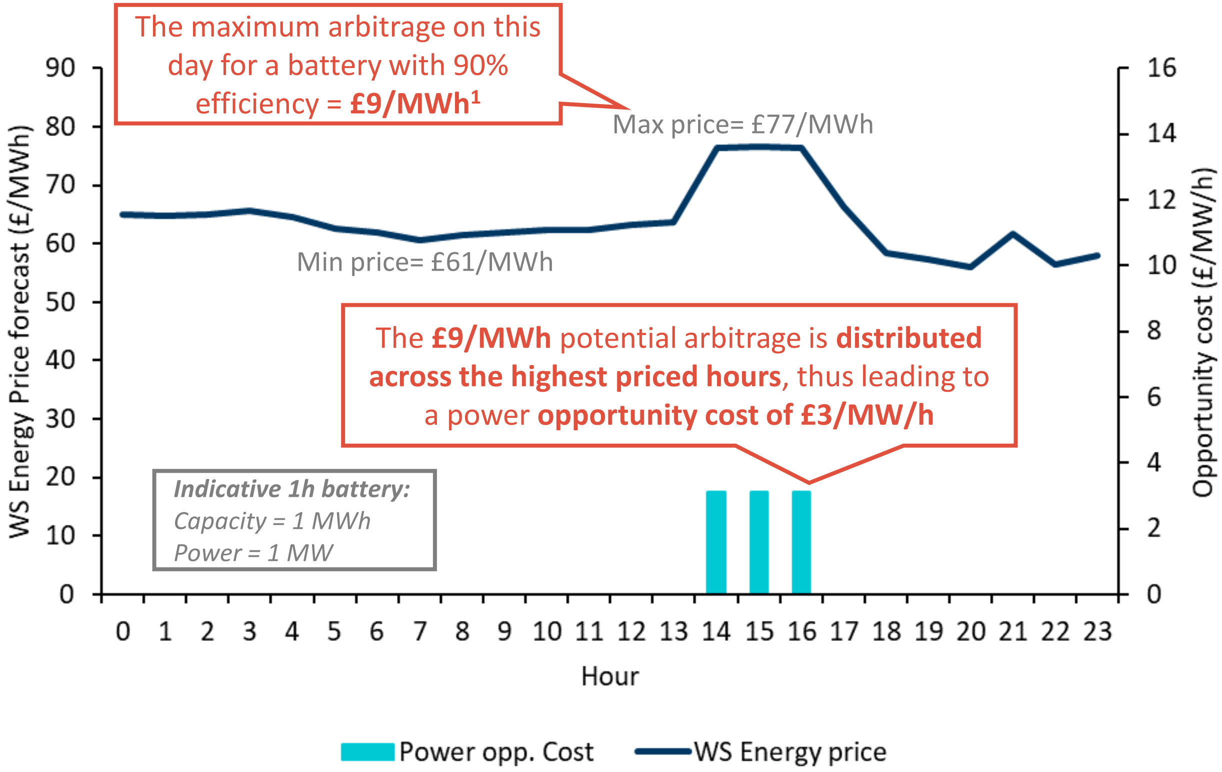
B **Reserve Up - Energy opp. Cost (£/MW) – Battery reserved after peak**



If the battery is reserved upwards **after** the peak hour, it will have to ensure it has **sufficient energy** to discharge if activated and therefore may **miss arbitrage revenues during the price peak**

A storage unit's opportunity cost of upward reservation during a price peak equals the foregone wholesale market arbitrage profits, spread over peak hours

A **Reserve – Power opportunity cost estimation of a battery (£/MWh)**



- Power opportunity costs are highest for **upward services in peak power price hours** and for **downward services in minimum power price hours**.
- In a **central dispatch system** (i.e. co-optimised), units do not have to create bids under uncertainty, so there is **no power opportunity cost**.

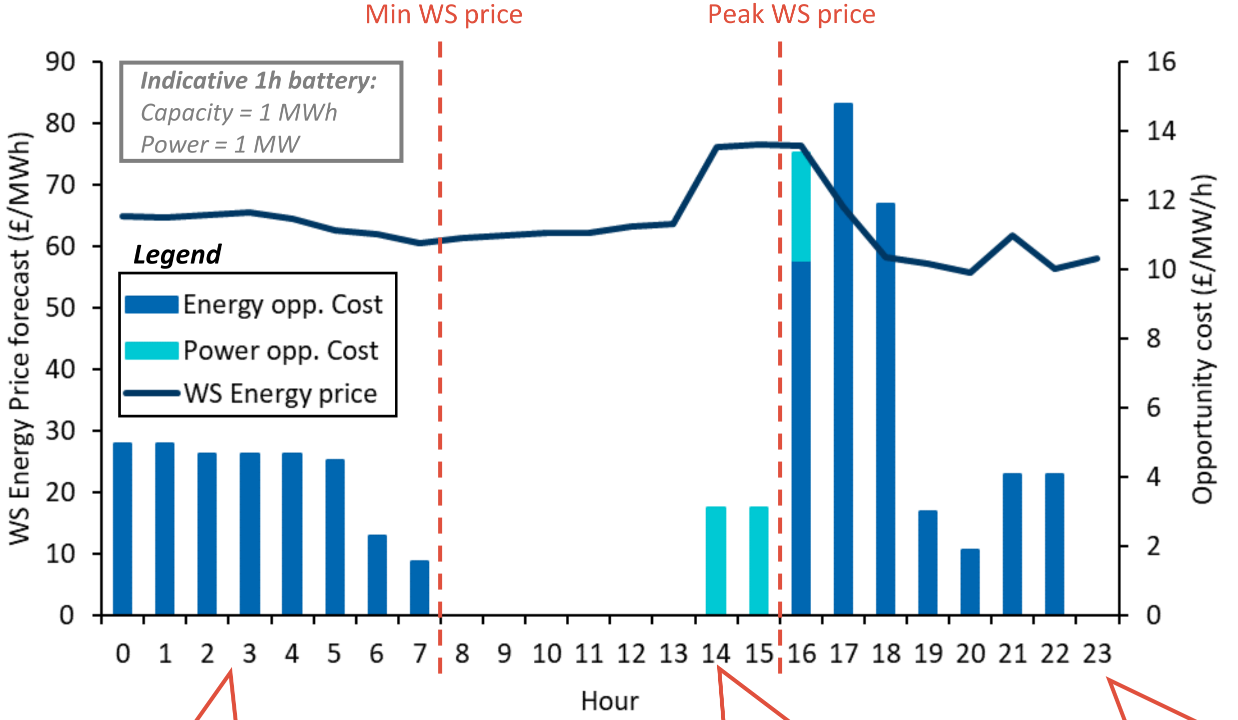
Methodology

- **Power opportunity costs** reflect the **uncertainty storage assets face when estimating their bids**. The opportunity cost ensures that a storage unit will be compensated for missing its best cycle (i.e. the charging and discharging profile that will maximise wholesale market arbitrage revenues) if it participates in reserve markets in that hour.
- However, there are often multiple hours of peak power prices within a day, so we assume that the **maximum arbitrage is distributed across all hours** that are within 10% of the expected peak price.
    - In the example on the chart to the left, the maximum arbitrage revenue is earned by charging at 7am (when WS Energy price reaches a minimum of £61/MWh) and discharging in any hour from 2pm to 4pm (when WS Energy price reaches maximum of £77/MWh).
    - The £9/MWh maximum arbitrage is spread across the three peak price hours, to give an opportunity cost of £3/MW/h in each of these hours.

Note: (1) Max arbitrage = max price – min price / efficiency = 77 – 61/0.9 = 9.

Storage bids so that if reserved, it can meet activation requirements, and deviate from its optimal energy dispatch (from a DA power price perspective)

B **Reserve Up**[1] **– Energy opportunity cost estimation of a battery**



Pre-minimum price (Midnight – 7am):
Opportunity cost reflects the cost of having to charge earlier than expected to provide upward reserve

After minimum price to penultimate peak hour (8am – 3pm):
Opportunity cost equals zero, as storage units can discharge and charge again one hour later , so they can provide reserve up without deviating from optimal energy dispatch

Last peak hour onwards (4pm – 11pm):
Opportunity cost initially reflects the cost of discharging after the peak for less. But then, as the energy price drops, a cheaper way to provide reserve arise - cycling a second time after the peak and internalizing the cost of a second cycle

Methodology

- In addition to the power opportunity cost (outlined on the previous slide), a storage unit's **energy opportunity cost** reflects the cost of deviating from the optimal energy dispatch, in order to have a sufficient state of charge at the start of the hour to provide AS in that hour (and hence be activated if called upon). As such, it is:
  - **Lower for Response than Reserve**, since we model Response with a 15-minute duration, whereas Reserve has a 1-hour duration; and
  - **Lower for units with longer storage duration**, as these units are more likely to have enough energy stored (e.g. lower for PS than for a 1h battery).
- Storage units face energy opportunity costs in **both the co-optimised and sequential models**.
- The opportunity cost therefore varies through the day, as shown in the example to the left.

Note: (1) Upward and downward positive opportunity costs occur at different times; they occur before off-peak and after peak hours for upward reserves, and they occur in between off-peak and peak hours for downward reserves.

Appendix 6 – Mathematical formulation of the co-clearing of AS markets under the sequential model

We co-clear all AS markets in the sequential model, using the calculated bids to find the minimum cost solution, whilst respecting physical constraints

*The co-clearing is performed on an **hourly[1] basis**, with the solutions of different hours being independent from each other*

Inputs

Set of values for which the co-clearing will be performed (i.e. units' bids, technical parameters of generators and reserve product)

From previous steps of the sequential model

We have calculated $bid\ price(g,m,d)$ and $bid\ volume(g,m,d)$

From our set of assumptions

We have $demand(m,d), activation\ time(m,d), MSL(g), ramp\ rate\ (g)$

Objective function – $f(cleared\ bid\ volumes)$

This represents the aggregate opportunity cost of providing reserves with a given clearing. We want to clear all the markets so that the minimum cost solution to this function is found. Its formulation is:

$$f(cleared\ bid\ volumes) = \sum cleared\ bid\ volume(g,m,d) \times bid\ price(g,m,d)$$

Where *cleared bid volume* is a variable for every generator (g) , market (m) and direction (d) that meets four constraints

C1 – Demand for every market must be met

The aggregate reserved capacity of all generators (g) in a market (m) and direction (d) must equal its demand, therefore, for every market and direction:

$$\sum cleared\ bid\ volume(g,m,d) = demand(m,d)$$

C2 – Reserved capacity must be physical feasible

The maximum capacity a generator can provide occurs when producing at MSL (requires start-up), therefore, for every generator:

$$\sum cleared\ bid\ volume(g,m,d) \leq av.capacity(g) - MSL(g)$$

C3 – Reserved capacity across markets must be compatible

A generator cannot be reserved in two markets if activating for one prevents it from activating in the other, therefore, for every generator and direction:

$$\sum cleared\ bid\ volume(g,m,d) \leq ramp\ rate(g) \times \max\{activation\ time(m,d)\}^{1}$$

C4 – Reserved capacity must be lower than bid volume

The reserved capacity of a generator (g) does not exceed its original bid, therefore, for every generator in every market and direction:

$$0 \leq cleared\ bid\ volume(g,m,d) \leq bid\ volume(g,m,d)$$

Optimal solution

Minimizing the $f(cleared\ bid\ volumes)$ we obtain the optimal solution, which corresponds to the co-clearing of ancillary services under a sequential system

Inputs | Variable | Constraints | Optimal solution

Notes: (1) Every input, variable, and solution is hourly and can vary from one hour to another; (2) $\max\{activation\ time\ (m,d)\}$ corresponds to the activation time of the slowest reserve, using it as a constraint, ensures that the reserved capacity of generator complies with all activation times

Appendix 7 – State of charge management for storage units participating in AS markets

In the sequential model, we enforce state of charge constraints on reserved storage units to ensure they have sufficient energy stored to provide AS

- In the co-optimised model, only storage units with sufficient state of charge are procured to provide AS, but in the **sequential model** the AS markets are cleared **before** the wholesale power market, hence we have to enforce SoC restrictions in the wholesale market to reflect AS reservations.
  - When a storage unit is reserved **upwards**, it must have **sufficient energy to discharge** if activated, hence we enforce a **minimum SoC** constraint.
  - Conversely, when storage is reserved **downwards**, it must have **sufficient available energy to load** if activated, hence we enforce a **max. SoC** constraint.
- When batteries are accepted in both response and reserves the min SOC and max SOC must be **combined**. For example, for 1h batteries the min SOC becomes 50% and the max SOC becomes 50%.[2] The constraints for **1h batteries** have been chosen to ensure there will always be a feasible range for SoC, even if providing upward and downward services simultaneously.
- **Response** has a **shorter duration** than reserve, leading to **weaker SoC constraints**. For example, if a 4h battery is procured for Reserve Up then it is assumed to need to have a minimum SoC of 25% (1 hr generation / 4hr storage), but if it is procured for Response Up it is assumed to only need a minimum SoC of 6.25% (15 mins generation / 4h storage).
- Each **pumped storage** unit has a different storage duration, hence the min SoC constraints vary by unit.[3]

State of Charge ("SoC") constraints in sequential model when storage unit is accepted in AS markets (%)

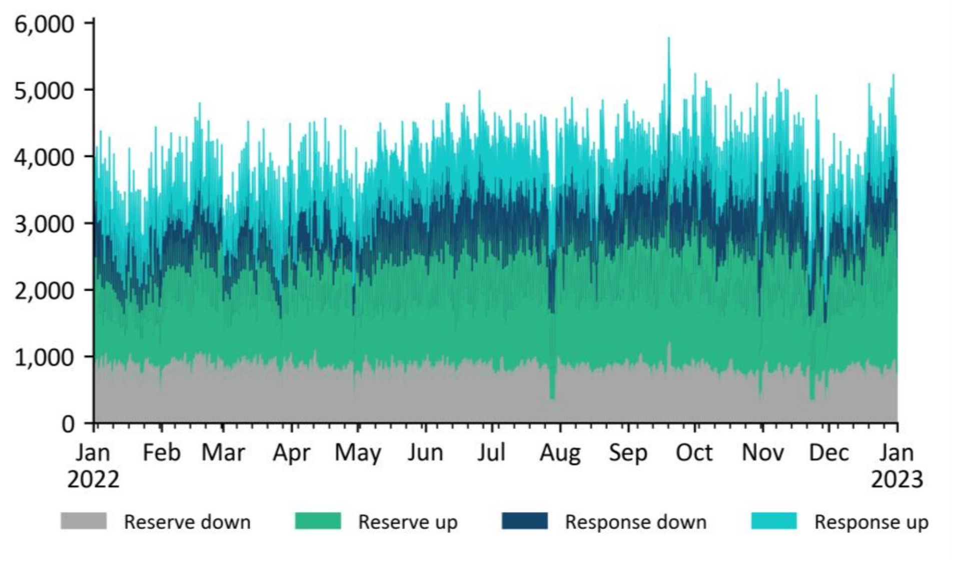
Product	Resource	Product duration	Max participation (% of capacity)	Min SoC when upward bid fully accepted	Max SoC when downward bid fully accepted
Response	1h batteries	15 mins	75%	25%	75%
	4h batteries		90%	6.25%	93.75%
Reserve	1h batteries	1 hour	25%	25%	75%
	4h batteries		50%	12.5%	87.5%
	Pumped storage[1]		50%	Varies by unit	Varies by unit

Notes: (1) To allow long-duration pump storage to correctly estimate their opportunity cost under the sequential model we do not enforce any constraint on its power capacity when being reserved but instead: (a) the reserved quantity is added as energy demand, and (b) the same quantity as the additional energy demand is added to the reservoir in the first hour of the modelled day (in addition to the default initial state of charge). This is a modelling tool to enable the model to correctly model pumped hydro storage and does not bias the results. In 2022, pumped storage max participation = 20% of capacity; (2) Except in the first hour of the day in years after 2022 when the initial state of charge is 25%, the min SOC can maximum be 25% for both 1h and 4h batteries. We set the initial state of charge equal to 50% in 2022; (3) Min SoC = Reserve Up provision (MW) / total storage capacity (MWh).

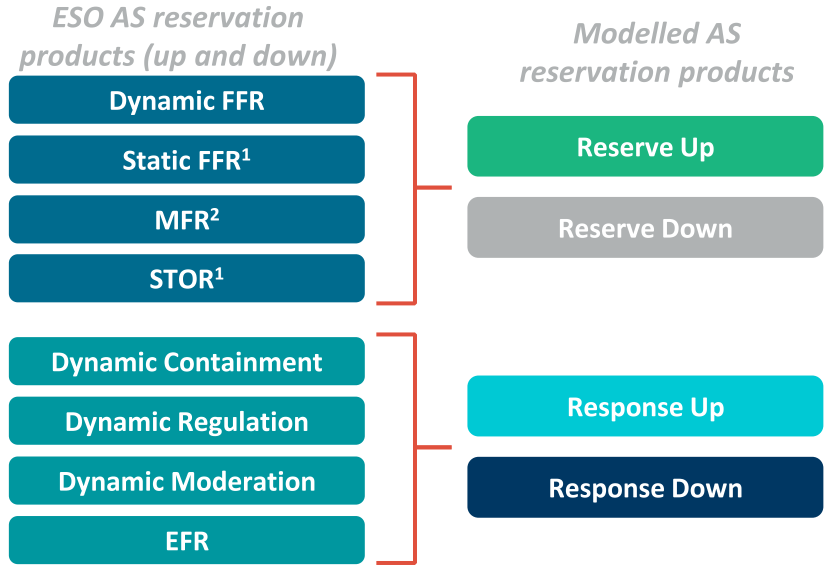
Appendix 8 – Historical 2022 AS volumes

We simplify historical AS to provide a clear comparison between market designs and we match historical AS volumes

Historical AS volumes, 2022 (MW/h, daily average)



Aggregation of ESO AS products for historical assessment



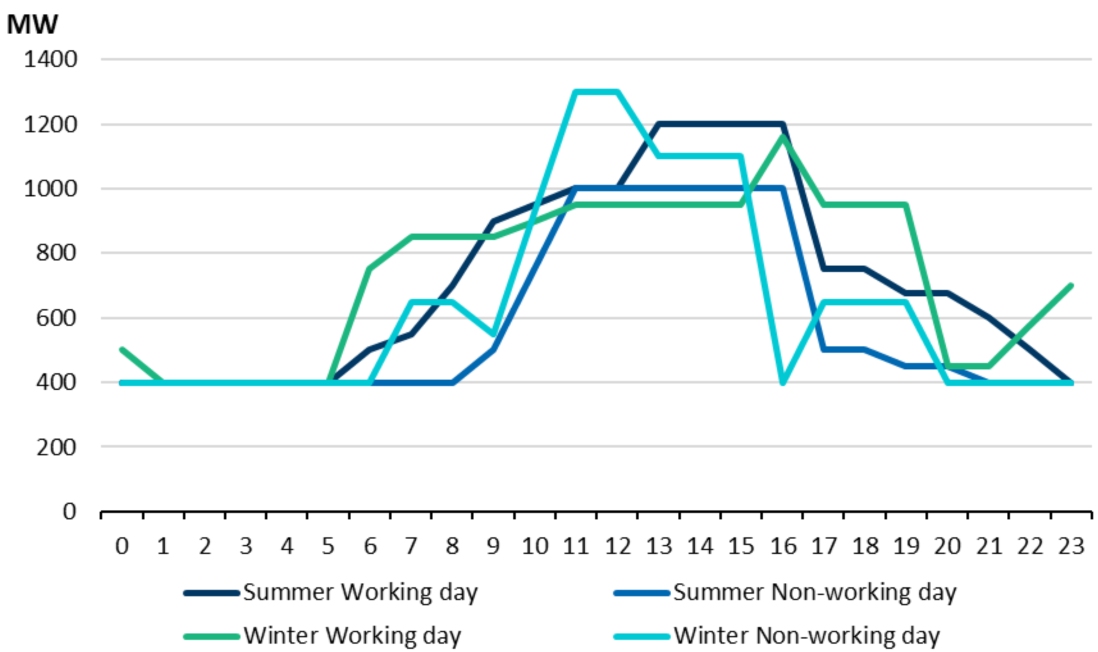
- We **aggregate historical AS** into Reserve and Response Up/Down, as explained in slide 95 and set out above, in order to make the modelling tractable. **We therefore match historical cleared AS volumes.**
- We further simplify the product activation times (1 second for Response and 10 mins for Reserve) and durations (15 mins for Response and 1 hour for Reserve) and clear all AS markets hourly, when in reality these properties all vary by service.
- **2022 AS volumes were very volatile** (as set out in the chart above) reflecting the differing AS requirements across time; the development of existing AS markets (e.g. Dynamic Containment volumes increased through the year after being introduced in late 2021) and the introduction of new AS markets (e.g. Dynamic Regulation and Dynamic Moderation).

Notes: (1) Static FFR and STOR are upwards only; (2) MFR Primary and MFR Secondary are only included in Reserve Up, MFR High is only included in Reserve Down.

Appendix 9 – Calibration of the balancing up reserve as a function of wind output

The Balancing Reserve Up requirement depends on time of day, season and wind output

Balancing Reserve Up requirement – time varying component (MW)



Balancing Reserve Up requirement – wind varying component (% of wind forecast)

Wind capacity factor (%)	Additional Reserve Up (% of wind forecast)
0%	0%
3%	50%
6%	35%
13%	25%
25%	16%
38%	13%
50%	10%
63%	8%
75%	7%
88%	6%
100%	5%

Note that as wind **output factor** increases, the **proportion** of output that must be procured as reserve **decreases**, but increasing deployment of **wind capacity** causes total Balancing Reserve Up demand to **increase** across the modelling horizon

- In the forward-looking assessment we aggregate three ESO reserve services (Balancing reserve, Slow reserve and Quick reserve) into the modelled 'Reserve' requirements.[1] The requirement for these services are all fixed, except for **Balancing Reserve Up**, which includes **time-varying and wind-varying components**.
- The time-varying component of the Balancing Reserve Up requirement is highest in hours of peak WS energy demand (left chart). The requirement **varies by (i) season, (ii) day of the week, and (iii) hour**.
  - For the **national model**,[2] to estimate the **hourly wind-varying component** of Reserve Up, we:
    1. Aggregate **hourly, post-constraint, wind output** across GB to a **GB national figure**…
    2. … and calculate the **average capacity factor** across wind units.
    3. In the table above, we calculate the **wind-varying component of Reserve Up** that needs to be procured for a given capacity factor.[3] e.g. if there is 40 GW of total wind capacity and post-constraint wind output in a given hour is 50% of total wind capacity (i.e. 20 GW), then the additional reserve requirement is 10% of wind output (i.e. 2 GW/h).

Notes: (1) See slides 38-39 for an explanation of reserve requirements in the national model; (2) See slides 69-71 for an explanation of reserve requirements in the zonal model; (3) We linearly interpolate between data points.

Appendix 10 – Details on maximum AS participation assumptions per technology

We have assumed restrictions on AS participation to reflect different technical characteristics and incentives facing different technologies

Maximum participation in AS markets by technology (% of capacity)

Resource	Response[1]	Reserve[1]	Comments / observations
Short-duration (1h) batteries	75%	25%	We restrict battery AS provision to ensure batteries have sufficient state of charge to provide AS. Batteries are assumed to be able to offer a smaller share of their capacity in reserve markets (25-50%) compared to response markets (75-90%) due to the longer activation time (1hr vs 15 mins). We do not permit BTM[2] and V2G[3] batteries to provide AS. CAES[4] and LAES[5] can only provide reserve (with the long-duration 50% restriction).
Long-duration (4h) batteries	90%	50%	
Pumped storage	No participation	50%[6]	We restrict pumped storage reserve provision to provide units with flexibility to still participate in wholesale markets and to align with historical observations.
Wind, solar and run-of-river hydro	No participation	No participation / 100%	Only run-of-river hydro can provide upward reserve - wind and solar assets are restricted, as it is considered uneconomic. We allow wind (including CfD assets, excluding OBZ assets) and hydro to participate in downwards reserve.
Demand-side response	No participation	18%	The proportion of DSR active in AS markets is based on 2022 observations. This prevents DSR from "flooding" the reserve market since DSR is very cheap to reserve at DA (but expensive to activate in RT).[7]
Thermal	No participation	90%[8]	We restrict nuclear generators from reserve markets due to their inflexibility. We restrict thermal units to 90% reserve provision to so that units would be able to participate in wholesale markets if prices increase intra-day.[9,10]

Notes: (1) We exclude any unit with capacity under 1 MW from participation in AS markets; (2) Behind The Meter; (3) Vehicle-to-Grid; (4) Compressed Air Energy Storage; (5) Liquid Air Energy Storage; (6) In 2022, we set pumped storage max participation to 20% of capacity to better align the participation of pump storage in the modelling with observed participation in AS markets; (7) However, there is significant uncertainty regarding future DSR participation in reserve markets. A higher share of participation in reserve could have significant impacts on the reservation prices; (8) 90% of installed capacity after deducting min stable level; (9) For biomass, hydrogen and waste generators FES 2022 does not provide an SRMC, so we reverse-engineered SRMC from the model (knowing the FES 2022 capacity and generation forecasts); (10) We do not model intra-day markets, but we restrict the thermal participation to allow for the real-world possibility of such intraday participation.

Appendix 11 – Historical 2022 WS Energy and AS prices

The prices in our Sequential model are lower than historical 2022 prices, thus providing a conservative estimate of the benefits of co-optimisation

Average WS Energy prices[1], Historical vs Sequential (£/MWh, time-weighted)

	Historical	Sequential
2022	205.66	164.63

- The modelled **Sequential WS price is lower than the historical 2022 average WS price** (£165/MWh vs £206/MWh time-weighted average).
- We match historical demand, IC flows and outages at large generators, but we assume market participants bid into the WS Energy market at marginal cost.[2,3] Therefore, we **cannot capture factors other than market fundamentals** which likely contributed to the high WS energy prices observed in 2022.
- We have **sense checked our modelled prices with ESO** and, given that the aim of this report is to compare market outcomes under sequential and co-optimised designs, **we do not aim to perfectly replicate historical market outcomes**.
- Nonetheless, by under-estimating the WS energy price we under-estimate cost to load in 2022 and hence we provide a **conservative estimate of the benefits of co-optimisation** in that year.
- The energy crisis in 2022 caused WS prices to increase significantly relative to adjacent years (£206/MWh in 2022 vs £118/MWh in 2021 and £94/MWh in 2023). The modelled Sequential price is in between the historical 2022 price and the historical price in adjacent years; the (absolute) benefits of co-optimisation in those years are likely to be lower than those observed in 2022.

Average AS prices, Historical vs Sequential – 2022 (£/MW/h, load-weighted)

	Historical	Sequential
Reserve Up	18.45	13.50
Reserve Down	12.39	0.00
Response Up	17.31	5.36
Response Down	4.58	4.90

- Similarly, the modelled Sequential AS prices are **under-estimates** relative to the observed historical average prices (right table).
- Again, we cannot capture factors other than market dynamics which may have contributed to these high prices, particularly for Response where new services were released in 2022 (i.e. DM and DR) and existing services were in the early stages of development (i.e. DC began in late 2021).
- The **historical Reserve Down price is positive**, whereas the modelled price is zero. The Dynamic FFR and MFR High services which we have aggregated into Reserve down are response services in reality, which are predominantly provided by batteries with positive opportunity costs. However, we model them as Reserve services (which dispatched wind and thermal units can provide at zero opportunity cost) as they do not meet the one second activation time for the modelled Response services, as agreed with ESO.

Notes: (1) EPEX Spot Annual Market Review 2022 (link); (2) In practice, all non-convex costs (start-up costs) are currently internalised in the bids of market participants in the WS market in GB; (3) Except for when they are reserved for AS, when they bid in for free.

Appendix 12 – Uplift methodology and results

We provide generators with uplifts to compensate for being scheduled in a way which makes them economically worse off compared to not running at all

- In the co-optimised model, we calculate uplifts to account for the fact that, in identifying the optimal WS and AS schedules, the model **only accounts for the marginal cost** and hence does **not account for generators' start-up costs.**
- Therefore, in cases where **thermal units** incur losses over the course of a day, they are compensated using uplifts. RES units and batteries do not have start-up costs, hence they never make a loss and we do not calculate uplifts for them. [1]
- We classify uplift payments as 'Reserve' uplifts if the generator is primarily online to provide reserves or 'Energy' uplifts if the generator is primarily online to provide energy (see equations to the right).
  - In both the Sequential and Co-optimised models, we include an '**Energy uplift**' to compensate units which came online primarily to provide WS Energy but made a loss that day. We calculate the energy uplift based on the Co-optimised model and assume it is constant across both models.
  - However, we only include the 'Reserve uplift' in the Co-optimised model. We do not include a '**Reserve uplift**' in the Sequential model because generators make AS bids fully informed of their own technical characteristics and with a unit-specific WS price forecast, and hence are able to bid to recover all expected costs.[2]
- The calculated uplift payments directly increase cost-to-load, as set out below:

DA cost-to-load definition

DA cost-to-load = (**DA AS reservation costs** + **'Reserve uplift'**) + (**DA WS energy costs** + **'Energy uplift'**)

Uplift calculation methodology

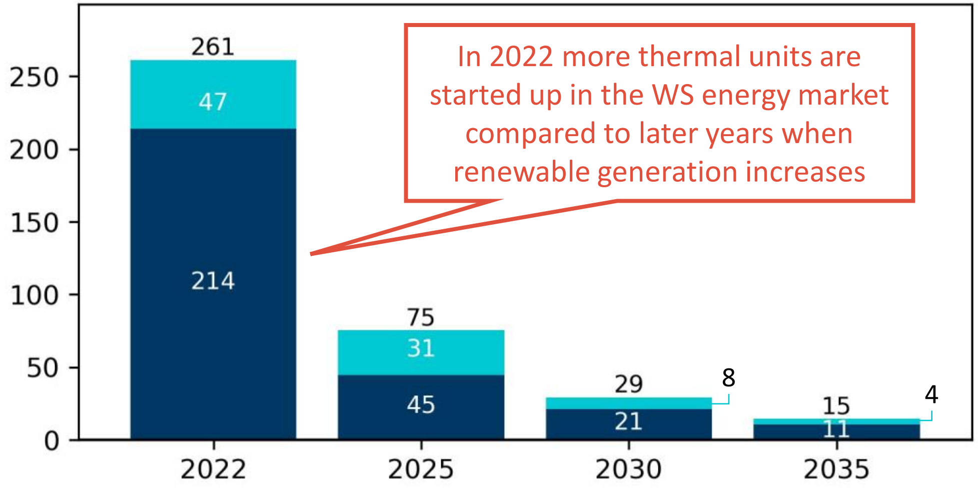
1. **For each thermal unit, we take the sum of its net profit[3] over the course of a day and check if it is negative or not**
   - if **negative** → **uplift > 0**
   - if **positive** → **uplift = 0**
2. **Given a negative profit, we classify it as an 'energy uplift' or a 'Reserve uplift' based on how many hours the generator had been reserved and how many hours the generator was generating**
   - if hours reserved **>** (hours generating – hours reserved) → **uplift classed as 'Reserve uplift'**
   - if hours reserved **<** (hours generating – hours reserved) → **uplift classed as 'Energy uplift'**
3. **Once we have calculated reserve and energy uplifts for each thermal unit for each day, we aggregate them into total reserve and energy uplifts for the year.** [1]

Annual reserve and energy uplift figures are set out on the next slide

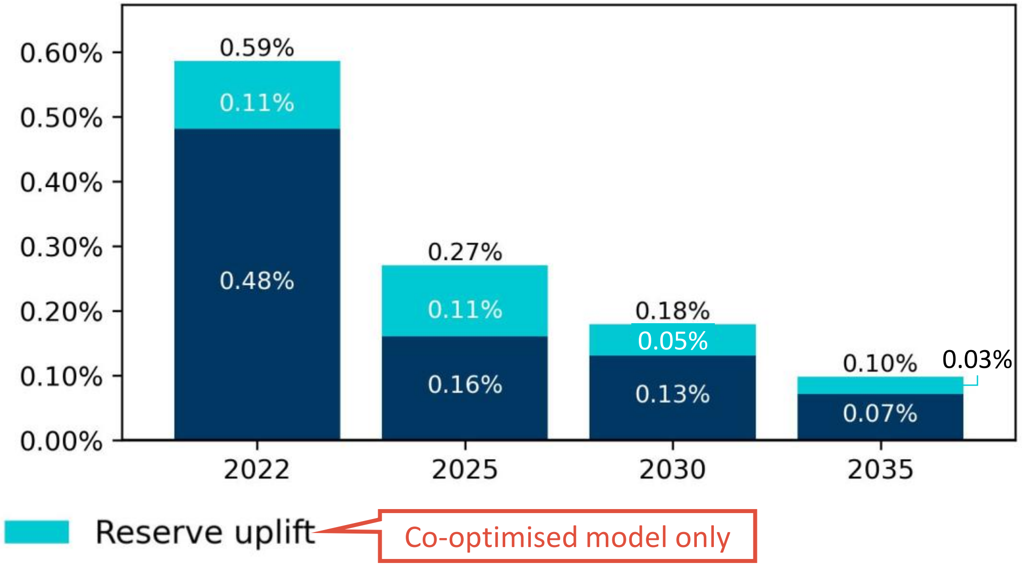
Notes: (1) We also do not calculate uplifts for CHPs given a negative profit in WS energy and AS markets does not necessarily imply a negative total profit due to the money earned from generating heat; (2) Units do not always recover all costs in the sequential model because individual units may mis-estimate the WS energy price. In the 2025 results there are instances when units committing to providing Reserve Up (and hence committing to generate at MSL) cause WS prices to fall and therefore the units make a loss, cannibalising themselves; (3) Net profit = WS Energy revenue + AS revenue – variable costs (i.e. fuel & emissions) – start costs.

Uplifts payments decrease over the modelling period as thermal units play a reduced role in WS energy and AS markets

Uplifts (£m) – All years



Uplifts as a proportion of total cost-to-load (%) – All years



- As explained on the previous slide, **uplifts are payments to make generators whole** if the received daily revenue from the marginally-priced wholesale and/or AS markets is insufficient to cover all operational costs (i.e. SRMC + start-up costs). 'Reserve uplifts' refer to uplifts for units that, for more than half of the hours that they are online in a day, also participate in reserve markets. In other words, these units are likely started up specifically to provide reserve.
- **Uplifts are highest in 2022** in both absolute terms (£261m, left chart) and relative terms (0.59% of cost to load, right chart) and then **decrease over the modelling period** (reaching £10m and 0.1% of cost to load in 2035).
  - At the start of the modelling period, thermal units play a key role in providing both WS energy and Reserve Up. Thermal units are more likely to be in-merit in either of these markets in the earlier years, hence leading to more instances where a thermal unit sets the price but makes an overall loss because the price only covers its variable production costs (i.e. fuel and emissions costs), without covering start-up costs.[1] This therefore requires increased uplifts.
  - As the modelling period progresses, the thermal installed capacity decreases (from 41 GW in 2022 to 18 GW in 2035) and the capacity factor of the remaining thermal units also decreases (from 37% in 2022 to 20% in 2035). Thermal units are therefore increasingly displaced from both WS energy and Reserve Up markets by RES and battery units, and hence uplifts fall.

Note: (1) We assume units bid into the WS Energy market at their SRMC (i.e. excluding start-up costs), as explained on the previous slide.

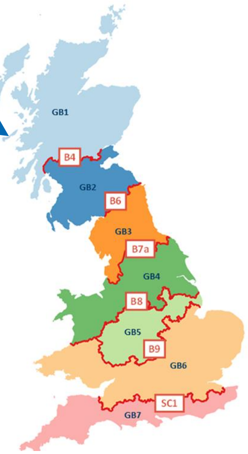
Appendix 13 – Zonal Reserve Up detailed calculations

Using the current approach for zonal, reserve demand in each region is often driven by the level of wind output in other regions

- We have selected two hours where GB1 has the same wind output in 2025.
- The amount of reserve procured in GB1 varies significantly, driven by wind output in other zones…
- …a difference of ~3x in the example hours shown.
- This also leads to hours where a large amount of reserve demand is concentrated in individual regions, causing demand and supply to struggle to balance.

7am 11th April 2025, High GB1 wind, low wind rest of GB

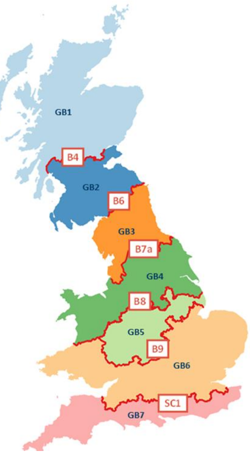
GB1
Wind output: 4,995 MW
Capacity factor: 52%
Reserve demand: 1,038 MW



Rest of GB
Wind output: 2,759 MW
Capacity factor: 9%
Reserve demand: 573 MW

6pm 8th March 2025, high GB1 wind, high wind rest of GB

GB1
Wind output: 4,995 MW
Capacity factor: 52%
Reserve demand: 356 MW



Rest of GB
Wind output: 25,606 MW
Capacity factor: 80%
Reserve demand: 1,824 MW

		7am 11th April 2025			6pm 8th March 2025		
		GB1	Rest of GB	GB total	GB1	Rest of GB	GB total
Wind output	a (GB1); b (GB total)	4,995 MW	2,759 MW	7,754 MW	4,995 MW	25,606 MW	30,601 MW
Wind capacity factor		52%	9%	19%	52%	80%	74%
Additional reserve demand (GB)	c	1,611 MW			2,180 MW		
Wind output as proportion of total GB wind output	a ÷ b	64%	36%	---	16%	84%	---
Zonal additional reserve demand	d = b × c	1,038 MW	573 MW	1,611 MW	356 MW	1,824 MW	2,180 MW
Additional reserve as proportion of wind output	d ÷ a	21%	21%	21%	7%	7%	7%

Quantity of flexible reserve procured in **GB1** is **3x lower**, despite **same local wind output**

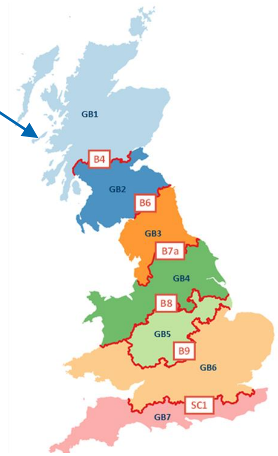
Alternatively, we can proportionally procure for each zone individually, but this will lead to a difference in reserve market size between national and zonal

- An alternative approach is to assess the amount of reserve required in each zone based on its individual wind output…
- …which leads to a consistent procurement of reserve in a zone for a given level of wind output in that zone, independent of conditions in other zones…
- …but leads to differences in the size of the total reserve procured compared to national.

7am, 11th April 2025, High GB1 wind, low wind rest of GB

GB1
Wind output: 4,995 MW
Capacity factor: 52%

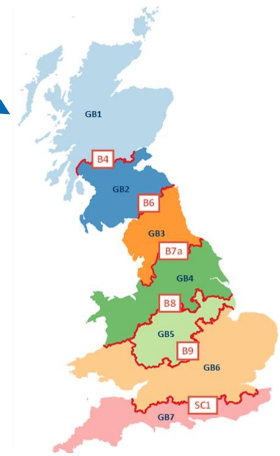
Rest of GB
Wind output: 2,759 MW
Capacity factor: 9%



6pm 8th March 2025, high GB1 wind, high wind rest of GB

GB1 (19%)
Wind output: 4,995 MW
Capacity factor: 52%

Rest of GB
Wind output: 25,606 MW
Capacity factor: 80%



		7am, 11th April: GB1	Rest of GB	GB total		6pm 8th March: GB1	Rest of GB	GB total
Wind output	a	4,995 MW	2,759 MW	7,754 MW	b	4,995 MW	25,606 MW	30,601 MW
Wind capacity factor		52%	9%	19%		52%	80%	74%
Additional reserve demand (GB)		N/A				N/A		
Wind output as proportion of total GB wind output		N/A				N/A		
Zonal additional reserve demand	c	484 MW	585 MW	1,069 MW		484 MW	1,610 MW	2,093 MW
Additional reserve as proportion of wind output	c ÷ a	10%	21%	14%		10%	6%	7%

Flexible reserve requirement fallen by c.40% relative to previous approach…

…although less impact (c.5%) when wind output spread more evenly across GB

In the zonal model, 'wind-varying' reserve is calculated within zone and then any 'shortfall' is added back to improve comparability with the national model

Zonal Reserve Up calculations - 7am, 11th April 2025

	Calculation	GB1	GB2	GB3	GB4	GB5	GB6	GB7	GB total	
Wind output	a	4,995 MW	2,285 MW	185 MW	251 MW	2 MW	28 MW	8 MW	7,754 MW	
Wind capacity	b	9,607 MW	9,203 MW	5,427 MW	8,189 MW	1,287 MW	6,577 MW	1,339 MW	41,628 MW	
Wind capacity factor	c = a ÷ b	52%	25%	3%	3%	0%	0%	1%	---	
Additional reserve (% of wind output)	d See Appendix 9	10%	16%	48%	50%	3%	7%	10%	---	
Zonal 'wind-varying' reserve	e = a × d	484 MW	369 MW	89 MW	125 MW	0 MW	2 MW	1 MW	1,069 MW	f
Load proportion	g	3%	6%	8%	20%	13%	39%	11%	---	h − f
'Shortfall' allocation	j = g × h	14 MW	33 MW	43 MW	110 MW	71 MW	213 MW	58 MW	542 MW	i =
Total 'wind-varying' reserve	k = e + j	498 MW	402 MW	132 MW	234 MW	71 MW	215 MW	59 MW	1,611 MW	h
'Fixed' and 'time-varying' reserve	m = g × l	68 MW	162 MW	212 MW	535 MW	347 MW	1,040 MW	286 MW	2,650 MW	l
Total Reserve Up	n = k + m	566 MW	563 MW	344 MW	770 MW	419 MW	1,255 MW	345 MW	4,261 MW	

- We initially calculate **'wind-varying'** reserve within each zone dependent upon the **hourly wind capacity factor within that zone** (see Appendix 9).
- However, this leads to a **'shortfall'** whereby the total GB 'wind-varying' reserve requirement is lower in zonal reserve markets (1,069 MW in the example above) than a national reserve market (1,611 MW).
- To ensure aggregate markets remain **consistent** and allow a **conservative comparison** of system costs across market design, we **allocate this 'shortfall' across zones in proportion to load.**
- This is then added to the **'fixed' and 'time-varying' portions of reserve** (which are also calculated in proportion to zonal load) to calculate the **total Reserve Up** requirement in each zone.

Glossary

Glossary

Actual / Co-optimised / Fully co-optimised	Automated optimisation of reserve, response and WS energy markets, considering all relevant requirements	**HND**	Holistic Network Design
AS	Ancillary Services	**Implicit co-optimisation**	Markets clear with implicit consideration of multiple requirements
BM	Balancing Mechanism	**LtW**	FES 2022 Leading the Way
CCS	Carbon capture and storage	**MAPE**	Mean Absolute Percentage Error
CfD	Contract for Difference	**MFR**	Mandatory Frequency Response
Client / ESO	National Grid Electricity System Operator	**MSL**	Minimum Stable Load
Contract	Agreement between FTI and ESO, dated 05 October 2023	**NOA**	Network Option Assessment
Counter-factual / Sequential / Separate	Reserve, response and WS Energy markets are cleared at different times	**Partly co-optimised**	Automated optimisation of multiple markets, considering only some requirements.
DA	Day-ahead	**Power opportunity cost**	Opportunity cost for storage units of being unable to charge / discharge in WS energy market at optimal hour
DC	Dynamic Containment	**PS**	Pumped storage
DESNZ	Department for Energy Security and Net Zero	**PV**	Present value
DM	Dynamic Moderation	**REMA**	Review of Electricity Market Arrangements
DR	Dynamic Regulation	**RES**	Renewable energy sources
DSR	Demand-side response	**ROC**	Renewable Obligation Certificate
EAC	Enduring Auction Capability	**RT**	Real-time
EFR	Enhanced Frequency Response	**SoC**	State of Charge
Energy opportunity cost	Opportunity cost for storage units of maintaining sufficient state of charge to provide AS	**SRMC**	Short-run marginal cost
ETYS	Electricity Ten Year Statement	**STOR**	Short-term operating reserve
FES	Future Energy Scenarios	**WM**	Wholesale market
FFR	Firm Frequency Response	**WS**	Wholesale energy
FTI	FTI Consulting LLP		

Experts with Impact™

FTI CONSULTING™