Ancillary services in Great Britain during the COVID-19 lockdown: a glimpse of the carbon-free future

Luis Badesa\textsuperscript{a}, Goran Strbac\textsuperscript{a}, Matt Magill\textsuperscript{b}, Biljana Stojkovska\textsuperscript{b}

\textsuperscript{a}Imperial College London, Department of Electrical and Electronic Engineering, London SW7 2AZ, UK
\textsuperscript{b}National Grid Energy System Operator, Faraday House, Warwick Technology Park, Warwick CV34 6DA, UK

Abstract

The COVID-19 pandemic led to partial or total lockdowns in several countries during the first half of 2020, which in turn caused a depressed electricity demand. In Great Britain (GB), this low demand combined with large renewable output at times, created conditions that were not expected until renewable capacity increases to meet emissions targets in coming years. The GB system experienced periods of very high instantaneous penetration of non-synchronous renewables, compromising system stability due to the lack of inertia in the grid. In this paper, a detailed analysis of the consequences of the lockdown on the GB electricity system is provided, focusing on the ancillary services procured to guarantee stability. Ancillary-services costs increased by £200m in the months of May to July 2020 compared to the same period in 2019 (a threefold increase), highlighting the importance of ancillary services in low-carbon systems. Furthermore, a frequency-secured scheduling model is used in the present paper to showcase the future trends that GB is expected to experience, as penetration of renewables increases on the road to net-zero emissions by 2050. Several sensitivities are considered, demonstrating that the share of total operating costs represented by ancillary services could reach 35%.

Keywords: Ancillary Services, Flexibility, Power System Stability, Renewable Energy

Nomenclature

\begin{align*}
\eta^c_s & \quad \text{Charge efficiency for storage unit } s. \\
\eta^d_s & \quad \text{Discharge efficiency for storage unit } s. \\
\pi(n) & \quad \text{Probability of reaching node } n \text{ in the stochastic scheduling.} \\
\Delta f_{\text{max}} & \quad \text{Maximum admissible frequency deviation at the nadir (Hz).} \\
\text{BEES} & \quad \text{Battery Storage Energy Systems.} \\
\text{BECSS} & \quad \text{Bio-Energy with Carbon Capture and Storage.} \\
\text{CCS} & \quad \text{Carbon Capture and Storage.} \\
\text{CfD} & \quad \text{Contract for Difference.} \\
C_g & \quad \text{Operating cost of generator } g \text{ (£).} \\
c_{\text{st}}^g & \quad \text{Startup cost of generator } g \text{ (£).} \\
c_{\text{nd}}^g & \quad \text{No-load cost of generator } g \text{ (£/h).} \\
c_{\text{m}}^g & \quad \text{Marginal cost of generator } g \text{ (£/MWh).} \\
D & \quad \text{Total system demand (MWh).} \\
D_{\text{shed}} & \quad \text{Demand shed (MWh).} \\
E_s & \quad \text{State of charge of storage unit } s \text{ (MWh).} \\
E_{s,\text{max}} & \quad \text{Maximum charge level for storage unit } s \text{ (MWh).} \\
E_{s,\text{min}} & \quad \text{Minimum charge level for storage unit } s \text{ (MWh).} \\
\text{EFR} & \quad \text{Enhanced Frequency Response (MW).} \\
\text{EFR}_s & \quad \text{EFR delivered by storage unit } s \text{ (MW).} \\
\text{EFR}_{\text{max}}^s & \quad \text{EFR capacity of storage unit } s \text{ (MW).} \\
f_0 & \quad \text{Nominal frequency of the power grid (Hz).} \\
g & \quad \text{Index for thermal generators.} \\
\text{GB} & \quad \text{Great Britain.} \\
H & \quad \text{Total system inertia (MVA} \cdot \text{s).} \\
H_g & \quad \text{Inertia constant of generator } g \text{ (s).} \\
H_{\text{Loss}} & \quad \text{Inertia constant of outaged unit (s).} \\
n & \quad \text{Index for nodes in the scenario tree of the stochastic scheduling.} \\
\text{OCGT} & \quad \text{Open-Cycle Gas Turbine.} \\
\text{OFDM} & \quad \text{Optional Downward Flexibility Management.} \\
\text{PFR} & \quad \text{Primary Frequency Response (MW).} \\
\text{PFR}_g & \quad \text{PFR delivered by generator } g \text{ (MW).} \\
\text{PFR}_{\text{max}}^g & \quad \text{PFR capacity of generator } g \text{ (MW).} \\
\text{PHES} & \quad \text{Pumped Hydroelectric Energy Storage.} \\
P_{s} & \quad \text{Charge rate of storage unit } s \text{ (MW).}
\end{align*}
1. Introduction

The outbreak of the disease COVID-19 in the final months of 2019 had a significant impact on electricity systems worldwide. Partial or total lockdowns were enforced by governments of several states, starting in China, followed by Italy and several European countries, as well as the United States. The measures taken varied from country to country, with several governments forbidding citizens from leaving their homes except for essential activities, and most of the industrial activity being put on halt. In Great Britain (GB), these measures caused the electricity demand to drop by up to 28% compared to levels in the same period in previous years.

Several studies have discussed the effects of lockdowns due to COVID-19 on electricity systems. Most of the focus has been put on the reduction in demand [2], with the subsequent drop in energy prices and carbon emissions [3], since demand in many countries was mostly covered by emissions-free renewables, which also benefit from low short-run marginal costs. Reference [4] has documented the correlation between reduction in electricity consumption in the United States and the rise in number of COVID-19 cases, degree of social distancing, and level of commercial activity. The work in [5] analyses the sharp drop of almost 40% in prices in the European carbon market in March 2020, driven by the surplus of emissions allowances as thermal generators were largely pushed out of electricity markets due to low demand. The authors in [6] analysed the demand reduction in Great Britain during the first week of the lockdown, highlighting some of the challenges that the system operator was expected to experience. The present paper focuses on another important aspect, that will only increase in importance as energy systems shift towards net-zero emissions: ancillary services.

A main challenge caused by low demand is related to grid stability: maintaining the network magnitudes, such as voltage and frequency, within safe levels is far from being trivial given these exceptional conditions. The authors in [7] discuss the preventive reactive-power management undertaken by the Indian system operator to avoid potentially dangerous voltage rises in the network. Such voltage rises could have been caused by an excess of reactive power, as there was a decrease in load while distributed renewable generation output did not decrease. In Great Britain, an island with no synchronous interconnection to other countries, low levels of inertia driven by high penetrations of non-synchronous renewable generation are a concern, which require large volumes of ancillary services to guarantee stability in the event of a contingency. A recent report [8] focused on the British system, highlighting that a significant proportion of demand in GB has been covered at times from non-synchronous renewables such as wind and solar PV, and providing a summary of balancing costs. However, no publication to date has provided a comprehensive analysis of the challenges experienced in GB, or the specific actions taken by the system operator and the reasoning behind them.

Furthermore, the lessons learned in GB are of uttermost relevance for the future decarbonised electricity grids. Some works such as [9] have already quantified the future cost of ancillary services, demonstrating that these will represent a very significant share of total operating costs for the electricity grid. However, an updated analysis is necessary given two major recent changes: 1) the ambition to sharply reduce emissions has greatly increased recently: as an example, the British system operator National Grid envisions that net emissions from the power sector will be negative by 2033 in 3 out of 4 future scenarios (achieved by including certain capacity of bio-energy with carbon capture) [10]. And 2) previous modelling tools did not have the capabilities to account for fast frequency response, which has been demonstrated as key to tackle the low-inertia problem (with 1GW of fast frequency expected to be procured by mid-2021 in GB [11]).

Several works have proposed optimisation models to schedule operating reserves [12], frequency response from thermal units [13] and synchronous inertia [14]. While frequency response is already co-optimised along with energy in some countries like the US, only the quasi-steady-state value is considered (i.e. the same volume of frequency response is scheduled as MWs would be lost after the largest generator outage). The quasi-steady-state value...
of frequency after a contingency was the only concern until recent years due to the high level of inertia in the grid, provided by synchronous generators as a by-product of energy. Nevertheless, decreasing inertia levels imply that a co-optimisation of inertia and response is necessary to compute the required volume of these services, as demonstrated by [13]. Furthermore, only recently has grid-scale battery storage become a cost-effective option, and this technology is the main provider of ‘fast frequency response’ in grids such as GB. The recent development of modelling tools that allow to also co-optimise fast frequency response [15] make it possible to analyse ancillary services requirements under the latest expectations for future generation mix.

In this context, the present work analyses the challenge of procuring ancillary services in GB during the COVID-19 lockdown, and demonstrates how these challenges will also be present in the future system as it becomes decarbonised. The contributions of this paper are two-fold:

1. To provide a comprehensive analysis of the stability challenge faced by the British electricity system during the COVID lockdown, as well as discussing the additional ancillary services procured by the system operator in this period and their associated costs.
2. To quantify the future costs for stability actions in GB and analyse their main drivers. To do so, a frequency-secured scheduling model is used, which explicitly considers the stability boundary within the optimisation. Several sensitivities are considered, demonstrating the importance of fast frequency response and size of the largest possible contingency.

The remainder of this paper is organised as follows: Section 2 reviews the need for and types of ancillary services procured by the GB system operator to guarantee stability during the lockdown period. Section 3 illustrates why the lessons learned during this period can inform the challenges that will be faced by the future high-renewable grids, using several results from a frequency-secured scheduling model that co-optimises energy production and ancillary services. Finally, Section 4 provides the conclusion and discusses future lines of work needed to enhance our understanding of best practices in ancillary-services procurement.

2. Ancillary services in GB during COVID-19

In this section a discussion on the operational challenges experienced in GB during the lockdown period is provided, as well as the actions taken by the system operator to overcome them. Finally, the consequences of this unique period are explained, both in terms of economic and environmental impact.

2.1. Challenges experienced

The challenges faced by the British electricity system during the lockdown period were due to three factors:

1. Depressed electricity demand.
2. Most renewables are not affected by market prices, due to their special market arrangements.
3. Nuclear stations are also not significantly affected by market prices, due to their inflexibility.

The first and second factors above caused periods of low net-demand (i.e. demand minus renewable generation, which must be covered by dispatchable generation), which in turn caused low system inertia: inertia is provided by synchronous generators (mainly gas- and coal-fired power stations), which were displaced by non-synchronous renewables such as wind and solar PV. The challenge of operating a low-inertia grid, which is further explained in Section 3.1, means that a higher volume of alternative ancillary services is needed to guarantee system stability.

The drop in demand caused a drop in energy prices, but most renewables in GB are incentivised to produce as much energy as possible regardless of the energy price. The reason for this is not only the low short-run marginal costs of these technologies (given that they do not incur fuel costs), but also their special market arrangements in GB: many renewable generators benefit from special financial instruments such as feed-in tariffs, Renewable Obligation Certificates and Contracts for Difference (CfDs). Although these three instruments vary in their characteristics, they all share a common feature: they shield renewables from low market prices, allowing these generators to still profit from generating even during periods of negative energy prices. A CfD, which is the financial support currently being awarded by the UK government to renewable generators successful in CfD auctions, guarantees a fixed strike price per MWh of energy paid to the generator, regardless of the price set in the wholesale market. This strike price effectively incentivises the renewable producer to output as much energy as possible, making it profitable even in the event of negative energy prices (since generators under CfDs are currently allowed to bid negatively in the wholesale market).

Simultaneously, the third factor contributing to the operational challenge during lockdown was that large nuclear units were predicted to maintain a high load factor, which would have increased the need for ancillary services. The reason for this increased need for ancillary services is that these units, in particular station Sizewell B, typically drive the largest power infraelastic loss in GB: since the system operator enforces the ‘N-1’ reliability requirement (meaning that the grid must withstand any single failure of a device), there must be enough ancillary services to cover for the outage of the largest plant online at any time.

Nuclear plants were likely to maintain high load factors because they are somewhat indifferent to low energy prices: given the inflexibility of these units in GB (long commitment times and high startup costs), it is typically preferable for these plants to stay online through periods of low prices rather than turn off.

In summary, the periods of very low net-demand com-
bined with high nuclear output increased the need for ancillary services in GB during the COVID-19 lockdown. As an illustrative example, the lowest ever observed national demand occurred during the lockdown: 13.4GW, overnight on Sunday June 28th 2020 [10]. To compare with 2019, the lowest demand was of roughly 18GW that year. In the next section, a discussion on the actions taken by the system operator in GB to overcome this operational challenge is presented.

2.2. Measures taken

The British government enforced a lockdown some days after other European countries had: lockdown in the United Kingdom started on 23rd March 2020, while in Spain it had started on 13th March and in the Lombardy region of Italy restrictions had been in place since around 21st February. This delay gave some valuable days to the GB system operator, National Grid, to conduct studies on how low demand could drop during coming months, since there was some concern were the lockdown to extend into the summer (which is the period of lowest demand in GB). By early April 2020, roughly a week after the lockdown was enforced, National Grid had developed a projection for the expected levels of demand in coming months: demand was expected to drop by up to 20% compared to the same periods in previous years [17].

These studies conducted by National Grid foresaw periods when it would be impossible to guarantee system stability. The system operator estimates that the capacity of synchronous units needed online to guarantee stability is in the order of 8 to 9GW, while the lowest possible demand foreseen during the summer was just above 12GW. Given that the installed renewable generation and about 5GW of nuclear power stations on the system were expected to generate as usual, as they are somewhat indifferent to demand drops (explained in Section 2.1), National Grid realised that there was a key necessity: the system was going to require more downward flexibility [18].

To obtain this downward flexibility, National Grid created a new product, called Optional Downward Flexibility Management (OFDM). The goal was to either increase the demand of the transmission system, or decrease the amount of non-flexible providers outputting onto the system. This OFDM product provided a commercial route for providers with whom National Grid did not have a commercial flexibility arrangement with: over 4.5GW of assets signed up for this service, predominantly embedded wind and solar, but also a fair share of demand turn-up assets.

The OFDM service was used 4 times, with a total cost of the actions of approximately £7mn [18]. These instructions were used both during daytime (such as from 6am to 4.30pm due to high solar output) and nighttime periods (as demand is lowest during night, which was combined with high wind in these periods). It is notably illustrative that three of these four OFDM instructions happened on the second bank-holiday weekend in the month of May (22nd to 25th of May 2020), since demand is traditionally lowest in GB during bank holidays and weekends.

Another measure taken by the system operator was to sign a bilateral contract with the nuclear power station Sizewell B. Through this contract, National Grid can instruct Sizewell B to halve its output: given that this is the largest generator in GB (with 1.3GW rated power), it defines the needed volume of ancillary services through the ‘N-1’ reliability requirement; therefore, part-loading this unit effectively reduces the need for ancillary services. The cost of this bilateral contract has been reported to be of up to £73mn [19].

Furthermore, National Grid also takes actions to guarantee system stability by committing out-of-merit synchronous generators in the balancing market, called the Balancing Mechanism in GB. As an example, 2.3GW of gas plants were brought online through the Balancing Mechanism on the Easter Monday bank holiday (13th April 2020) with the purpose of increasing inertia and procuring additional frequency capability [20].

Finally, the last measure taken was to create a fast-track application process to the Accelerated Loss of Mains Change Programme [21], targeted to embedded resources. Through this fast-track programme, generators receive payments from National Grid to change their protection settings, so that they stay connected to the grid under high Rate-of-Change-of-Frequency conditions. The programme will be completed in coming years, but the fast track created during the lockdown had the goal of decreasing the need for one key ancillary service that was scarce during this period of low net-demand: inertia.

In summary, the GB system operator took four different measures to avoid blackouts during the challenging lockdown period:

1. To create the OFDM product in order to procure downward flexibility.
2. To sign a bilateral contract with the large power station Sizewell B, allowing to halve its power output.
3. To take stability-related actions through the Balancing Mechanism, as usual practice before the lockdown.
4. To create a fast-track route for the Accelerated Loss of Mains Change Programme, which will reduce the need for inertia.

2.3. Economic and environmental consequences

The actions to procure additional ancillary services explained in Section 2.2 implied additional costs. Ancillary services costs were 3 times higher in the period May-July 2020 compared to the same months in 2019: £302mn in 2020 compared to £101mn in 2019 [22].

On the positive side, the high penetration of renewable generation during this period has produced records of minimal carbon intensity: May 2020 was the lowest carbon-intensity month for electricity so far in Great Britain, with May 24th showing the lowest ever carbon intensity of the
system (46gCO₂/kWh produced on the transmission network) [18]. One of the reasons for this drop in carbon emissions during lockdown was the halt in coal usage: the UK did not use any coal for power generation for just under 68 consecutive days, from April to June 2020. This set a record for the longest coal-free period since the industrial revolution, exceeding by far the previous record of 18 days achieved in 2019 [23].

3. Future trends

Even though electric demand is expected to increase in GB in coming decades due to the electrification of heat and transport, net-demand is expected to further decrease due to an even higher rate of increase of renewable penetration to meet the emission targets [10]. In addition, an increase in the largest power infed loss with the commissioning of large nuclear plants such as Hinkley Point C in coming years [21], will increase the need for ancillary services to maintain system stability.

These two effects, the low inertia caused by low net-demand and the high value of the largest power infed, will exacerbate in coming decades the operational challenge already experienced during the lockdown of 2020. To illustrate this point, a frequency-secured scheduling tool is used in this section considering National Grid’s latest view on the generation mix of the future, demonstrating that the cost of ancillary-services provision will increase as GB decarbonises its grid to meet the target of net-zero carbon emissions by 2050.

Note that the present analysis focuses on ancillary services to guarantee frequency stability, which is the main concern in low-inertia power systems. With increasing loading level in the grid, voltage stability could also be compromised. Other concerns could also arise, such as a higher difficulty in respecting rotor-angle stability, and even new types of potential instability driven by the fast dynamics of power-electronics-based generation [25]. These types of stability are not considered in this paper.

3.1. Conditions for frequency stability

National Grid, as the system operator in Great Britain, must take actions to guarantee that the frequency of the grid is kept within safe boundaries after the loss of any generator (an operational practice commonly known as the ‘N-1 reliability requirement’). These safe boundaries are defined as:

1. The Rate-of-Change-of-Frequency (RoCoF) must not exceed 0.125Hz/s, in order to avoid triggering RoCoF-sensitive relays that could further exacerbate the generation loss by disconnecting distributed generation. This RoCoF limit is currently being relaxed to 0.5Hz/s and 1Hz/s in certain locations through the Accelerated Loss of Mains Change Programme, as the cost of RoCoF-related ancillary services is expected to significantly increase in coming years if this limit is not relaxed [20].

2. The lowest frequency value, the frequency nadir, must not drop below 49.2Hz. This boundary avoids the activation of Low Frequency Demand Disconnection, which would entail significant economic penalties for the parties responsible, as occurred after the large power outage of August 9th 2019 [27].

In order to guarantee that the two above frequency boundaries are respected, the energy market must be constrained so that the generation dispatch contains sufficient ancillary services. These ancillary services for frequency support in GB are: 1) inertia; 2) Enhanced Frequency Response (EFR); and 3) Primary Frequency Response (PFR) [28]. These ancillary services help contain the frequency drop that would be caused by a generation outage, as described in Figure 1.

Mathematically, the conditions to guarantee frequency stability take the following shape. The RoCoF is guaranteed to be within safe limit if the following inequality holds [29]:

\[
\frac{H}{f_0} \geq \frac{P_{\text{Loss}}}{2 \cdot \text{RoCoF}_{\text{max}}}
\]

where \(H\) is the system inertia provided by all synchronous generators online (with units of GVA·s), \(f_0\) is the nominal frequency of the grid (50Hz in European grids), \(P_{\text{Loss}}\) is the largest power infed loss (units of GW) and \(\text{RoCoF}_{\text{max}}\) is the maximum admissible RoCoF (units of Hz/s).

The frequency nadir is guaranteed to be above the predefined limit if the following inequality is respected [15]:

\[
\left(\frac{H}{f_0} - \frac{\text{EFR} \cdot 1s}{4 \cdot \Delta f_{\text{max}}}\right) \cdot \text{PFR}_{\text{10s}} \geq \left(\frac{P_{\text{Loss}} - \text{EFR}}{4 \cdot \Delta f_{\text{max}}^2}\right)^2
\]

where \(\Delta f_{\text{max}}\) is the maximum admissible frequency deviation (in Hz), EFR is the aggregate from all devices providing this service (units of MW, EFR must be fully delivered by 1s after the contingency), and PFR is the aggregate from all devices providing this service (units of MW, PFR must be fully delivered by 10s after the contingency).

Note that the EFR service is currently being migrated into a service with similar characteristics named Dynamic Containment [30].
3.2. Frequency-secured scheduling

In order to understand the need for ancillary services in the future carbon-free GB grid, several relevant case studies are run here using a frequency-secured stochastic Unit Commitment model with two important characteristics: 1) Considers uncertainty from renewable generation, by using a scenario tree of possible RES realisations; and 2) Guarantees frequency stability, by enforcing frequency constraints.

The objective function of this frequency-secured scheduling minimises the expected value of system operating costs (fuel costs, start-up costs) to meet demand:

$$\min \sum_{\forall n} \pi(n) \left( \sum_{\forall g} C_g(n) + VoLL \cdot D_{\text{shed}}(n) \right)$$

Where the operating cost of thermal units is given by:

$$C_g(n) = c^\text{st}_g \cdot y_g^\text{st}(n) + c^\text{nl}_g \cdot y_g(n) + c^\text{m}_g \cdot P_g(n)$$

The objective function is subject to engineering constraints in the generation mix, such as start-up times and ramp rates for thermal plants. The full description of these constraints can be found in Appendix A. The binary variables in the UC are relaxed to continuous in order to decrease computation time. The objective function is also subject to the frequency-stability constraints (1) and (2), to guarantee that the scheduling solution would allow to keep the RoCoF and frequency nadir within required limits in the event of any credible contingency.

The frequency-stability constraints in Section 3.1 map the sub-second dynamics of post-contingency frequency evolution (shown in Figure 1), into any timescale desired for the optimisation, such as the hourly timescale used in this scheduling model. These constraints make sure that sufficient ancillary services will be available to contain any credible frequency drop: inertia, EFR and PFR are decision variables in the scheduling optimisation.

Inertia refers to the energy stored in the rotating masses on synchronous generators (e.g. nuclear, gas, biomass). These masses act as an energy buffer that stores kinetic energy as they rotate: this energy is spontaneously released in the event of a generation outage, decreasing the rotating speed of these masses and therefore decreasing the electrical frequency of the grid. A higher number of synchronous generators committed increases system inertia.

Frequency Response is a power injection from different assets such as generators and batteries. Generators have slower dynamics in frequency-response provision, making them candidates for the PFR service (which has a requirement of being fully delivered by 10s after a contingency). Batteries and other devices controlled through power electronics converters have much faster dynamics, allowing them to provide EFR (which must be fully delivered by 1s after a contingency).

All these ancillary services can be thought of as a form of ‘insurance’ to prevent blackouts: for example, part-loaded generators providing frequency response operate at a lower-than-optimal power output, therefore increasing their per MWh operating cost. The headroom in part-loaded generators provides this insurance in the form of frequency response, that will act if a generation outage occurs. Another example of this insurance consists in bringing out-of-merit thermal units to increase the system inertia: these thermal units brought online displace non-synchronous renewables, therefore increasing the fuel costs in the system, as demonstrated in detail in [31].

3.3. Ancillary services costs by 2030

The frequency-secured scheduling introduced in Section 3.2 is used here to understand the future trends on ancillary services provision. The generation mix considered is described in Table 1 and corresponds to the ‘Leading the way’ scenario within National Grid’s 2020 Future Energy Scenarios [10], described as the ‘fastest credible path to decarbonisation’. This scenario achieves a carbon-neutral electricity system by 2030, and is the result of a capacity-expansion optimisation which considers a set of different technologies as generation candidates, subject to regulatory constraints such as emissions limits.

### Table 1: Characteristic of thermal plants

| Number of Units | Hinkley Point C | Nuclear | Gas CCS | OCGT | Biomass | BECSS |
|-----------------|----------------|---------|---------|------|---------|-------|
| Rated Power (MW)| 1800           | 1350    | 500     | 100  | 500     | 500   |
| Min Stable Generation (MW) | 1350 | 1000 | 250 | 50 | 450 | 450 |
| Ramp Rate (MW/h) | 100            | 100     | 250     | 50   | 50      | 50    |
| No-Load Cost (£/h) | 400           | 400     | 4500    | 3000 | 5000    | 3500  |
| Marginal Cost (£/MWh) | 8             | 8       | 46      | 200  | 25      | 15    |
| Startup Cost (£’000) | N/A           | N/A     | 10      | 0    | 20      | 20    |
| Startup Time (h) | N/A            | N/A     | 4       | 0    | 4       | 4     |
| Min Up Time (h) | N/A             | N/A     | 4       | 0    | 4       | 4     |
| Min Down Time (h) | N/A            | N/A     | 1       | 0    | 1       | 1     |
| Inertia Constant (s) | 5             | 5       | 5       | 5    | 5       | 5     |
| Max PFR deliverable (MW) | 0             | 0       | 50      | 20   | 0       | 0     |
Electric demand has an annual peak of 60GW and a minimum value of 20GW, accounting for daily and seasonal trends. The RES fleet is composed of 30GW of solar PV and 68GW of wind. In addition, the system contains 14GW of storage, including the 2.8GW existing Pumped Hydroelectric Energy Storage (PHES) plus a further 2GW assumed to be built by 2030, and 9.2GW of Battery Energy Storage Systems (BESS). PHES has a round-trip efficiency of 75% and 5h duration, while BESS has a round-trip efficiency of 90% and 2h duration. In the base case, 1GW of EFR is procured, as per current plans by National Grid from mid-2021 onwards.

The results for a scheduling simulation spanning one full year of operation are presented in Figure 2. The cost of ancillary services is computed by running two simulations, one with the frequency-stability constraints (1) and (2) enforced, and another one without these constraints (therefore corresponding to an energy-only scheduling). The difference between the operating cost in the first simulation and the operating cost in the second simulation gives the cost of ancillary-services provision.

Figure 2 illustrates that the share of total system operating costs corresponding to ancillary services would rise to 15% by 2030 under the ‘Leading the way’ scenario. For comparison, Figure 2 also displays the estimated energy and ancillary services costs in 2015, when only 2% of costs corresponded to ancillary services.

It is insightful to consider the distribution of system inertia throughout the year, as presented in Figure 3. In 2016, 220GVA·s was the most common value, therefore inertia-related actions taken by the system operator were rare. On the other hand, the simulation results for 2030 show a distribution of inertia skewed towards significantly lower values. This distribution contains three peaks corresponding to the most common values of system inertia throughout the year:

- Peak 1, at 74GVA·s, corresponds to hours for which only nuclear and gas plants are online providing a baseline of inertia, in order to respect the nadir limit given by eq. (2). Although biomass plants have lower marginal costs than gas plants in Table 1, the minimum stable generation of gas plants make them preferable in these hours as a source of inertia: this lower minimum stable generation allows to accommodate more energy from RES, which are assumed to have zero marginal costs, therefore reducing overall system costs.
- Peak 2, at 92GVA·s, corresponds to periods of high RES output when it is effective to commit a combination of biomass plants to provide inertia (since the inertia provided by biomass plants is cheaper than that from gas plants, given the lower marginal costs of biomass units), and some gas plants to provide the necessary PFR.
- Peak 3, at 165GVA·s, corresponds to periods of high net-demand, when dispatchable, synchronous plants must be turned on for energy purposes, and therefore inertia effectively becomes a by-product of energy.

To further understand the annual distribution of inertia for 2030 presented in Figure 3, an example of a week of operation of the system is included in Figure 4. The first two
days of this week show periods of high net-demand, when the gas plants increase their output to cover the system demand. On the other hand, the last three days show periods of zero net-demand, when RES curtailment is present for a sustained period. Even though in these days all the demand could be covered by non-synchronous RES generation, a minimum amount of gas plants is kept online to provide inertia: in the current GB system, this would be achieved by bringing online out-of-merit thermal units through the Balancing Mechanism, in order to increase inertia and therefore guarantee system stability.

On a final note, it is important to clarify that RES do not provide ancillary services in the studies presented in this paper. That is currently the case in most countries in the world, as significant market barriers exist for these generators. Notably, ancillary services markets in GB are currently cleared well ahead of real time, which makes it impossible for RES to accurately forecast their potential contribution to these services. In addition, the current RES fleets, which are based on grid-following inverters, are widely agreed to be less effective than synchronous machines to contain frequency drops. Nevertheless, system operators are working to remove market barriers, and the adoption of grid-forming inverters could remove the technical limitations. Once RES are fully able to provide ancillary services, the stability challenge in decarbonised grids could be reduced, but the trend of increasing overall costs for these services projected in this paper is still likely to occur.

3.4. Fast frequency response to compensate for low inertia

One of the keys to tackle the low-inertia challenge is the procurement of fast frequency-response services (such as EFR in GB), as demonstrated by works like [32]. Here we run a sensitivity analysis for the volume of EFR procured by 2030, for which the results are presented in Figure 5. These results demonstrate that the cost of ancillary services could more than double if the volume of EFR procured is halved from the currently planned 1GW, reaching a share of total system costs of 35% in this case (since projected energy costs for 2030 are of £6.5bn/year, as presented in Figure 2). On the other hand, procuring an additional 0.5GW of EFR (making the total volume of 1.5GW) would reduce ancillary-services costs by roughly £700m/year, highlighting the system value of this service. Note that during the 2020 lockdown, 200MW of EFR were being procured [33].

It is important to clarify that, since EFR is provided by BESS which do not incur fuel costs, the only system operating cost associated to EFR is a lost opportunity to use the BESS fully for energy arbitrage. In practice, BESS would be remunerated for providing EFR with a certain price set in an auction for this service. This revenue, along with other revenue streams such as the price differentials from doing arbitrage in the energy market, would allow BESS owners to recover the investment cost on these assets. Note however that the scheduling model used in this paper only considers the operational timescale of the power system, assuming a fixed capacity mix, and therefore investment costs are out of the scope of this analysis.

3.5. Impact of largest loss and reliability standards

In previous sections, the largest loss infed driven by the large nuclear plant Hinkley Point C has been considered to be fixed, i.e. \( P_{\text{Loss}} = 1.8GW \). Here, we consider two alternative cases:

- The system is secured against an ‘N-2’ loss, which considers the simultaneous loss of Hinkley Point C power station and half of the 2GW rated capacity of the double-circuit HVDC interconnector IFA. The value of \( P_{\text{Loss}} \) is then set to 2.8GW.
- A variation of the above case, where the size of the largest possible loss is optimised in the scheduling model, by optimally part-loading the large nuclear station if beneficial for the system as a whole. To compute this optimal part-loading strategy, \( P_{\text{Loss}} \) is considered a decision variable in constraints \([1]\) and \([2]\), while respecting the minimum stable generation for Hinkley Point C in Table \([1]\).

The results for these cases are presented in Figure 6. A volume of 1.5GW of EFR is considered to be procured
in all three cases, since it was verified that 1GW of EFR would not be enough to secure the ‘N-2’ case, given the generation mix considered in Table [1].

These studies are motivated by the need to understand the cost of a tighter reliability standard, since the most recent large power outage in GB (on the 9th of August 2019) was caused by an ‘N-2’ outage [27]. The results in Figure 7 demonstrate that securing the system against any two simultaneous losses increases the cost of ancillary services by more than 6 times, even in the presence of 1.5GW of EFR. On the other hand, optimally sizing the largest possible loss is shown to achieve very significant savings. However, in this third case Hinkley Point C station is chosen by the optimal scheduling to operate on average at 1.45GW throughout the year; therefore, these savings from optimal part-loading can only be considered as an ideal scenario, since energy and ancillary services are fully co-optimised in the scheduling simulation in this case. In practice, nuclear part-loading would be done ex-post (e.g. through the Balancing Mechanism in GB, which can adjust the day-ahead energy market dispatch), since Hinkley Point C will benefit from a CfD and therefore is incentivised to produce as much energy as possible. This matter is further discussed in the next section.

3.6. Future market arrangements for renewables and nuclear plants

In addition to the increased need for ancillary services in a low-inertia system, the special market arrangements currently in place for most renewables in GB complicated the challenge during the 2020 lockdown. In coming years, renewables might still benefit from instruments such as Contracts for Difference (CfDs), as the business case for merchant renewables (i.e. renewables that simply obtain revenue from the wholesale market) is still under debate, due to issues such as price cannibalisation (i.e. renewables would not be able to capture high prices, given the correlation between wind or solar output at different locations, respectively). Under a CfD, renewable generators would still be incentivised to produce as much energy as possible, regardless of the level of demand. Nevertheless, the latest round of CfD auctions held in GB fords any recipient of a CfD to bid negatively [34], which could help alleviate the renewable output during periods of very low demand (when energy prices could potentially be negative).

Furthermore, the large nuclear plant Hinkley Point C, which will drive the largest loss in GB in coming years, will operate under a CfD with a strike price of £92.5/MWh (in 2012 prices, to which inflation must be added) [24], placing it at around 45% above average wholesale day-ahead electricity prices in the last decade [35]. This implies that it will be incentivised to run at output roughly at all times, which would highly increase the need for ancillary services due to the large size of this plant driving the ‘N-1’ reliability requirement. Some studies have proved that part-loading large nuclear units when net-demand is low can significantly decrease system operating costs and reduce overall emissions [36], therefore mechanisms for making this strategy compatible with a CfD should be designed.

3.7. Applicability of the results to other countries

The results presented in this paper have been obtained by using the projected future British electricity system as a test case. As an island, Great Britain is already suffering the low-inertia problem induced by increasing renewable penetration, which was particularly acute during the COVID-19 lockdown. Other island nations such as Ireland and Australia are experiencing the same issue, as well as Texas, which has no synchronous interconnection to other grids in the continental United States.

However, the results presented here could also be of relevance to system operators and stakeholders in the energy sector in other countries. While continental grids like the US and Europe currently benefit from sharing inertial support through AC transmission corridors, once renewable penetration is sufficiently high in all the interconnected countries, the low-inertia issue will also appear in continental grids. Therefore, the lessons drawn from the simulations conducted in this paper will be of interest for these systems.

Note that the present paper makes no assumption on the particular market arrangements for procuring ancillary services in GB: the cost of ancillary services presented here is simply the opportunity cost computed with the frequency-secured Unit Commitment, as explained in Sections 3.2 and 3.3. The only British particularity in terms of ancillary services that is considered in this paper is the existence of the ‘fast frequency response’ service (i.e. EFR), which also exists with different terminology in Australia, Ireland and Texas. While this product does not yet exist in most countries, it is not unreasonable to expect that a similar product will eventually be created by other system operators to counteract the low-inertia problem. The advantage of defining a ‘fast frequency response’ product is to recognise the faster dynamics of certain grid-connected assets such as battery storage, and the results presented here demonstrate that such service can bring significant cost savings.

For quantifying the value of and need for the different ancillary services in a given country, it might nonetheless be relevant to consider its particular market arrangements. A working group within the ‘Conseil international des grands réseaux électriques’ (CIGRE), gathering experts from system operators and academia across the world, is in the process of collecting operational practices for ancillary-services procurement in different countries. The interested reader can refer to the technical brochure which summarises the findings of the working group, expected to be published in Q1 2021 under the title ‘Impact of high penetration of inverter-based generation on system inertia of networks’.
4. Conclusion and future work

The lockdown measures following the COVID-19 outbreak in GB has provided some valuable insights on the practical challenges of operating a system with low net-demand. While these challenges had already been predicted, they were not expected to be present until some years in the future, as the penetration of renewables keeps increasing to eventually reach the legal requirement of net-zero emissions by 2050 imposed by the British parliament. The exceptional circumstances of the pandemic, and the subsequent depression in electric demand, have shown that these challenges are very much real.

This paper has demonstrated and discussed why these challenges will become increasingly relevant as the share of renewables in the system grows. While the difficulty in managing the GB system will likely be higher than in continental grids, given that it is an island where the low-inertia problem is already present, other countries will certainly experience similar issues in the future as they progressively move towards a carbon-free energy system.

Regarding future work that could lead to improved practices in procurement of ancillary services, it is notably relevant to understand how to make compatible the energy-based, financial-support mechanisms for certain generators (e.g. Contracts for Difference) with the provision of inertia, frequency response and an optimised largest in-feed loss. This issue is particularly key in GB, where the future driver of the largest possible loss will be a nuclear plant benefiting from a CfD, which could exacerbate the stability challenge during periods of low net-demand.

Acknowledgements

The authors are grateful for the valuable support and funding received by the UK EPSRC project ‘Integrated Development of Low-Carbon Energy Systems’ (IDLES, Grant EP/R045518/1) and EU H2020 TradeRES project (ID: 864276).

The first author is also grateful to Dr. José Enrique “Villica” Villa Escusol for the conversation that sparked the writing of this paper.

References

[1] ESO operational transparency forum, COVID-19 preparedness Report, National Grid ESO (August 2020). URL https://data.nationalgrideso.com/plans-reports-analysis/covid-19-preparedness-materials

[2] J. López Prol, S. O, Impact of COVID-19 measures on short-term electricity consumption in the most affected EU countries and USA states, iScience (2020) 101639. doi:10.1016/j.isci.2020.101639

[3] Global energy review 2020: The impacts of the Covid-19 crisis on global energy demand and CO2 emissions, Report, International Energy Agency (June 2020).

[4] G. Ruan, D. Wu, X. Zheng, H. Zhong, C. Kang, M. A. Dahleh et al., A cross-domain approach to analyzing the short-run impact of COVID-19 on the U.S. electricity sector, Joule (September 2020). doi:10.1016/j.joule.2020.08.017

[5] M. Eikerhoff, L. Zeltzerberg, Can the EU ETS weather the impact of Covid-19?, Report, Centre for European Policy Studies (June 2020).

[6] K. Bell, G. Hawker, Electricity demand during week one of COVID-19 lockdown article, UK Energy Research Centre news (March 2020). URL https://ukerc.ac.uk/news/electricity-covid-lockdown/

[7] H. Zhong, Z. Tan, Y. He, L. Xie, C. Kang, Implications of COVID-19 for the electricity industry: A comprehensive review, CSEE Journal of Power and Energy Systems 6 (3) (2020) 489–495.

[8] I. Staffell, R. Green, R. Gross, T. Green, Electric insights Q-2 2020 quarterly report, Report, Imperial College London & Drax Group plc (September 2020).

[9] Roadmap for flexibility services to 2030: A report to the Committee on Climate Change, Report, Imperial College London and Pöyry Management Consulting (2017).

[10] Future energy scenarios, Report, National Grid ESO (2020).

[11] Dynamic containment Report, National Grid ESO (January 2020). URL https://www.nationalgrideso.com/document/162381/download

[12] A. Sturt, G. Strbac, Efficient stochastic scheduling for simulation of wind-integrated power systems, IEEE Transactions on Power Systems 27 (1) (2012) 323–334. doi:10.1109/TPWRS.2011.2165558

[13] H. Chávez, R. Baldick, S. Sharma, Governor rate-constrained OPF for primary frequency control adequacy, IEEE Transactions on Power Systems 29 (3) (2014) 1473–1480. doi:10.1109/TPWRS.2014.2298838

[14] F. Teng, V. Trovato, G. Strbac, Stochastic scheduling with inertia-dependent fast frequency response requirements, IEEE Transactions on Power Systems 31 (2) (2016) 1557–1566. doi:10.1109/TPWRS.2015.2434837

[15] L. Badesa, F. Teng, G. Strbac, Optimal portfolio of distinct frequency response services in low-inertia systems, IEEE Transactions on Power Systems 35 (6) (2020) 4459–4469. doi:10.1109/TPWRS.2020.2997194

[16] National Grid ESO coronavirus preparedness, Report, National Grid (1st July 2020).

[17] Electricity summer outlook report Report, National Grid ESO (April 2020). URL https://www.nationalgrideso.com/document/167541/download

[18] Power responsive summer insights: Covid-19 system operability challenges, Industry podcast series, National Grid (July 2020).

[19] National Grid pays Sizewell B owner for halving power output, News article, The Times (August 2020).

[20] National Grid ESO coronavirus preparedness, Report, National Grid (15th April 2020).

[21] Assessment of risks resulting from the adjustment of RoCoF based loss of mains protection settings, Report, Energy Networks Association (2015).

[22] Monthly Balancing Services Summary (MBSS) Report, National Grid ESO (August 2020). URL https://data.nationalgrideso.com/balancing/mbss

[23] 67 days coal free - a major milestone on the journey to a greener Britain, News article, Ecotricity (July 2020).

[24] N. Hatziargyriou, J. V. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich et al., Definition and classification of power system stability revisited & extended, IEEE Transactions on Power Systems (2020). doi:10.1109/TPWRS.2020.3041774

[25] Frequency changes during large disturbances and their impact on the total system, Report, Ofgem (2014).
Appendix A. Unit Commitment formulation

The optimisation problem described in Section 3.2 is subject to the constraints included in this appendix. Decision variables are italicised to distinguish them from constants. For the sake of simplicity, the formulation presented here corresponds to the deterministic UC, while the interested reader can refer to [12] for a detailed description of the generalisation of these constraints to be applied to the stochastic UC.

The only binary decision variables in this formulation are \( y^g \) and \( y^s \), while all other are continuous. The results presented in Section A.3 through Section A.5 have been obtained with these binary variables being relaxed, in order to decrease computational time.

The load balance constraint, assuming an hourly time-step in the scheduling optimisation, is determined by:

\[
\sum_{g} P_g + \sum_{s} \left( P^d \cdot P^s \right) + P_{\text{RES}} - P_{\text{RES}}^{\text{surf}} = D - D_{\text{shed}} \tag{A.1}
\]

Where RES curtailment is limited by these inequalities:

\[
0 \leq P_{\text{RES}}^{\text{surf}} \leq P_{\text{RES}} \tag{A.2}
\]

**Appendix A.1. Constraints for thermal units**

The thermal units described in Table 1 are subject to the following constraints. The power limits:

\[
y_g \cdot P_{\text{max}} \leq P_g \leq y_g \cdot P_{\text{max}} \tag{A.3}
\]

The ramp limits:

\[
-P^r \cdot y_g(t-1) \leq P_g(t) - P_g(t-1) \leq P^r \cdot y_g(t) \tag{A.4}
\]

The commitment state of a unit is ‘on’ if it was generating in the previous time-step or has started generating in the current time-step, unless it has been shut down in the current time-step:

\[
y_g(t) = y_g(t-1) + y^g(t) - y'^g(t) \tag{A.5}
\]

The thermal unit starts generating if it was started up ‘\( T^g \)’ hours before:

\[
y^g(t) = y^g(t - T^g) \tag{A.6}
\]

The generator is allowed to be started up if it was ‘off’ in the previous time-step and has been ‘off’ for at least ‘\( T^\text{mut} \)’ hours:

\[
y^g(t) \leq [1 - y_g(t-1)] - \sum_{i=t-T^\text{mut}} y^g(i) \tag{A.7}
\]

Regarding the contribution of a thermal unit to PFR, this is limited by its headroom and by the PFR capacity of the unit:

\[
P_{\text{FR}}^g \leq y_g \cdot P_{\text{max}}^g - P_g \tag{A.9}
\]

\[
P_{\text{FR}}^g \leq P_{\text{FR}}^{\text{max}} \tag{A.10}
\]

**Appendix A.2. Constraints for storage units**

PHES and BESS units are subject to the following constraints. The only binary variable is the charge/discharge switch, \( y_s \), which prevents the storage unit from charging and discharging simultaneously. All other variables are continuous.

The energy stored is defined by:

\[
P^{\text{min}} \leq E_s \leq P^{\text{max}} \tag{A.11}
\]

\[
E_s(t) = E_s(t-1) + \left( \eta_s \cdot P^c_s(t) - \frac{P^d_s(t)}{\eta_s} \right) \tag{A.12}
\]

The charge and discharge power are constrained by:

\[
0 \leq P^c_s \leq (1 - y_s) \cdot P_{\text{max}}^s \tag{A.13}
\]

\[
0 \leq P^d_s \leq y_s \cdot P_{\text{max}}^d \tag{A.14}
\]

The contribution of BESS to EFR is determined by:

\[
\text{EFR}_s \leq \text{EFR}^{\text{max}} \tag{A.15}
\]

\[
\text{EFR}_s \leq y_s \cdot P_{\text{max}}^d - P^d_s + P^c_s \tag{A.16}
\]

The last constraint accounts for the fact that BESS can swiftly switch from charging to discharging, therefore effectively providing a higher volume of EFR if it operates in charging mode.
Appendix A.3. Definitions of frequency-related magnitudes

This frequency-secured UC guarantees frequency stability by including constraints (1) and (2). The expressions that appear in these two constraints are defined in this subsection.

The total system PFR is given by:

\[ \text{PFR} = \sum_{g} \text{PFR}_g \]  
(A.17)

The total system EFR is defined by:

\[ \text{EFR} = \sum_{s} \text{EFR}_s \]  
(A.18)

System inertia is the aggregate of the inertia from each thermal plant online at a given time:

\[ H = \sum_{g} H_g \cdot \text{P}_{\text{max}}_g \cdot y_g - \text{P}_{\text{max}}_{\text{Loss}} \cdot H_{\text{Loss}} \]  
(A.19)

which accounts for the loss of inertia from the outaged generator.

The ‘fixed’ largest loss is defined as:

\[ P_{\text{Loss}} = \text{P}_{\text{max}}_{\text{Loss}} \]  
(A.20)

While the ‘optimised’ largest loss, achieved by optimally part-loading large units in the scheduling, is determined by:

\[ P_{\text{Loss}} = P_g \quad \text{for } g = \text{Largest Unit} \]  
(A.21)

The above constraint assumes that the largest unit is a must-run unit, i.e. it is committed at all times. This is the case for the GB system described in Table 1, where the largest unit is the nuclear plant Hinkley Point C.

Appendix A.4. Uncertainty from RES

Uncertainty in RES generation is included in the stochastic Unit Commitment through a scenario tree. This scenario tree discretises the probability density function of forecast errors for the RES generation. Full details on how to compute this discretisation and the probabilities of reaching each of the nodes, \( \pi(n) \), can be found in [12].

Figure A.7 displays a generic example of the use of scenario trees within the stochastic scheduling. The stochastic UC is run with a 24h lookahead: at each time-step, a scenario tree is built for the following 24h hours. A rolling-planning approach is also used, consisting in only keeping the optimisation solution for the current-time node in the scenario tree (the progenitor of all nodes) and discarding all other decisions. This approach has been shown to provide lower cost solutions for systems with a high renewable penetration.