Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany

Michael Joos\textsuperscript{a}, Iain Staffell\textsuperscript{b,⁎}

\textsuperscript{a} Element Energy, 78 Margaret Street, London W1W 8SZ, UK
\textsuperscript{b} Centre for Environmental Policy, Imperial College London, London SW7 1NE, UK

ARTICLE INFO

Keywords:
Electricity
Renewable energy
Wind power
Reserve
Balancing
Curtailment

ABSTRACT

Britain and Germany saw unprecedented growth of variable renewable energy (VRE) in the last decade. Many studies suggest this will significantly raise short-term power system operation costs for balancing and congestion management. We review the actual development of these costs, their allocation and policy implications in both countries.

Since 2010, system operation costs have increased by 62% in Britain (with a five-fold increase in VRE capacity) and remained comparable in Germany (with capacity doubling). Within this, balancing costs stayed level in Britain (~4%) and decreased substantially in Germany (~72%), whilst congestion management costs have grown 74% in Britain and 14-fold in Germany. Curtailment costs vary widely from year to year, and should fall strongly when ongoing and planned grid upgrades are completed. Curtailment rates for wind farms have risen to 4–5% in Germany and 5–6% in Britain (0–1% for offshore and 15–16% for onshore Scottish farms).

Policy debates regarding the balancing system are similar in both countries, focusing on strengthening imbalance price signals and the extent that VRE generators bear the integration costs they cause. Both countries can learn from each other’s balancing market and imbalance settlement designs. Britain should reform its balancing markets to be more transparent, competitive and open to new providers (especially VRE generators). Shorter trading intervals and gate closure would both require and enable market participants (including VRE) to take more responsibility for balancing. Germany should consider a reserve energy market and move to marginal imbalance pricing.

1. Introduction

Both the UK and Germany have the declared goal to cut carbon emissions by 80% by 2050 compared to 1990 levels, which requires near-total decarbonisation of the electricity sector. While nuclear and carbon capture and storage (CCS) are plausible low carbon alternatives in the UK, public and political opinion limit Germany’s options to renewable energy sources (RES).

Renewable generation has grown dramatically in both countries. Since 2015 29% of Germany’s electricity was renewable, as was 24.5% of Britain’s \cite{[1,2]}. The German government aims to source at least 50% of electricity from RES by 2030 \cite{[3]}, and similarly a 45–55% share is needed to meet the UK’s Carbon Budget for 2030 \cite{[4]}. The EU targets at least 27% RES share in total final energy consumption by 2030 \cite{[5]}, which studies suggest would correspond to an electricity share of around 45–55% depending on modelling assumptions \cite{[6–8]}. The majority of new RES will come from wind and solar \cite{[9]}, since alternatives like hydro and bioenergy are constrained by limited resource and higher cost. Wind and solar are variable renewable energy (VRE) sources: their output depends on weather conditions and so can be forecasted but not fully controlled as can thermal plant. To offset deviations between forecast and actual wind and solar output reserves must be held and operated \cite{[10]}. VRE capacity has grown significantly in recent years. Combined wind (onshore and offshore) and solar capacity in the UK \cite{[11]} grew from 5.46 GW in 2010 to 27.25 GW at the end of 2016 \cite{[1]}, while capacity in...
Germany increased from 44.84 GW to 90.81 GW [11]. Fig. 1 shows the rise in capacity and penetration.

The cost of wind and solar has declined substantially in the last decade, such that the levelised cost of electricity (LCOE) for onshore wind and large solar is now lower than gas and nuclear [12–14]. However, LCOE does not tell the whole story: To accommodate VRE output while enforcing high standards for security of supply, costs are incurred in other parts of the system, mainly for holding and operating reserve and back up plants to manage variability and uncertainty of VRE output. These are so-called system integration costs.

Several studies conclude that electricity systems with high shares of renewable energy (up to 100%) are feasible in industrialised economies without a reduction of security of supply standards [15–17]. Grids have maintained or even increased their reliability during significant increase of VRE penetration (the ‘System Average Interruption Duration Index’ in Germany went from 18.67 min in 2006 to 10.45 min in 2015) [18,19]. However, decarbonising the electricity sector must be conducted cost-efficiently and would lose public acceptance if it led to a substantial increase of power prices for industry, businesses and residents.

Numerous studies exist on integration costs, mostly based on modelling (e.g. [20–23]). These concur that increasing VRE penetration will lead to increased system operation costs, which could hinder the political feasibility of highly or completely renewable electricity systems. However, the already significant penetration of VRE in Britain and Germany offer an opportunity to study the impacts that VRE integration is actually having on two major industrialised economies based on real market data.

We choose to focus on Britain and Germany for five reasons:

- they are two advanced industrial countries of similar economy and population, climatic conditions, and ambitious climate policies;
- they have comparable conventional power mixes and natural resources, notably limited opportunities for hydro storage and domestic biomass, and a recent increase in renewable generation capacity;
- they currently have a large installed wind capacity, both in terms of % penetration, and especially in terms of GW capacity;
- they are widely anticipated to have the largest installed wind capacities in Europe: for example, the ENTSO-E 2030 Vision see between 30 and 58 GW installed in Britain and 61–101 GW in Germany (together 40–50% of Europe’s total wind capacity) [24];
- the transparency of their systems operators and regulators means that substantial quantitative data is available, necessary to perform such a study.

While higher system operation cost due to VRE integration might not yet be observable in retail electricity prices (cp. Appendix D), detailed data on system operation costs are made publicly available by system operators and regulators which we use as evidence on cost impacts of system integration of VRE sources in the UK and Germany. For this reason, we focus on short-term integration costs of VRE sources. In contrast, quantifying long-term integration costs (such as for lower average utilisation of generation capacity) requires scarce data, involves controversy about methodologies and must be based on tentative assumptions about future electricity systems.

Short-term integration costs of VRE broadly consist of two components: grid congestion management and balancing. This paper explores the impacts that VRE is having in the British and German power systems, how related costs are allocated in both systems, how these costs have developed in the last 7 years and the ongoing policy debates in both countries regarding their reduction and cost reflective allocation. We focus mainly on wind over solar, as it is the major VRE technology in both Britain and Germany in terms of electricity generation and has higher impacts on grid congestion and system balancing.

The cost of grid congestion, especially compensating curtailed wind output, has gained media attention and fuelled the debate about curtailment management, how risk of curtailment should be shared and how it relates to a changing role of distribution grid network operators [25,26]. However, these costs are a temporary phenomenon that can be reduced by sufficient reinforcement of grid capacities, which is ongoing in both countries. The cost allocation of grid infrastructure investments, while being related is another question of great complexity, which is handled very differently in the UK and Germany, and is widely debated in policy, research and industry [27–29]. As this involves long-term considerations and assumptions, it is not discussed here.

We investigate reforms of the balancing system in both countries. This policy debate centres on three topics:

1. setting the right incentives for market participants to contribute to the balance of the system;
2. giving them sufficient possibilities and opportunities to act on these incentives; and
3. reducing balancing costs through improved system operation, including opening balancing markets to new providers, including VRE generators.

Another question is whether integration costs constitute system externalities [22]: Do VRE generators bear the costs that they cause elsewhere in the system to a sufficient extent in current market arrangements? If not, how could market arrangements be changed to internalise these costs?

The next section explores grid constraint costs, their cost allocation and recent development in Germany and Britain. Section 3 compares how system balance is managed in Britain and Germany and investigates cost developments in both countries. Policy options for balancing systems reform are discussed in Section 4. Section 5 then summarises and concludes. The data presented in this paper is available.
open-access via Zenodo to help facilitate future research in this area: http://doi.org/10.5281/zenodo.848028.

2. Constraint costs

This section explores the costs of grid congestion management and their allocation in Britain and in Germany. Redispatch measures are used to manage congestion by ramping down plant before the congestion and ramping up plant behind it, maintaining the energy balance of the overall system.

In both Britain and Germany wind generators are located mainly in the North due to better land and wind resource availability, while demand is centred in the South. This has put pressure on transmission grids, and several projects to reinforce or extend North-South network capacities are ongoing or planned in Germany [30] and Britain [31].

2.1. Congestion management and its cost allocation in Germany

German TSOs contractually assure themselves the right to intervene with generation profiles in case of network congestion. In addition, they contract committed reserve capacities for congestion management, called Netzreserve (grid reserve) [32]. TSOs assess the demand for such reserve capacities which must be approved by the Bundesnetzagentur to be procured [33]. Grid reserve is seen only as a temporary measure which should be made obsolete by sufficient grid reinforcements and extensions. The government plans to extend the grid reserve until 2023 (it was planned to be implemented until 2017) [34].

Costs for congestion management in Germany are broadly divided into two categories: redispatch and curtailment of renewable feed-in, which are regulated by the Energiewirtschaftsgesetz (EnWG, Energy Act) §13. Redispatch measures constitute utilisation of contracted services from conventional generators or committed reserve plants although a court has ruled in 2015 that generators might have to be reimbursed to a higher extent than currently [32,35]. Curtailment of renewable feed-in is not contracted but must be compensated and due to priority dispatch of RES should only be used as a last resort to manage grid constraints [35,36]. Unlike redispatch measures, RES curtailment is not paired with measures in the opposite direction and thus affects system balance and must be compensated for, e.g. by utilising balancing power [37].

RES generators are compensated with 95% of their foregone revenues (Renewable Energy Act, EEG §15) by the grid operator they are connected to. Irrespective of compensation payments made to RES generators for curtailment, grid operators are legally obliged to invest in grid reinforcements sufficient to transport their output [32].² [38] The government allows TSOs to tolerate a maximum of 3% curtailed wind output in their network development plans to avoid disproportionately high grid investments [34].

It is suggested that due to the foreseen compensation, TSOs tend to curtail RES preferably since this avoids conflicts with generators that do not get compensated [39]. Costs for the grid reserve and RES curtailment are socialised via grid fees [37], which in Germany are paid for only by consumers [27].

The government argues that compensation of curtailed renewable output is necessary, otherwise the risk of curtailment would be priced into all new renewable plants – even those which will experience only minimal curtailment – and thus increase the cost of renewable energy in general [34].

2.2. Recent congestion management costs in Germany

After a modest increase in 2010–2013, costs of congestion management² increased significantly in 2014 and 2015 then decreased slightly in 2016 (Fig. 2). RES curtailment costs were €478 m in 2015, with 4722 GWh or 2.8% of RES output lost [40]. Costs for redispatch were €435 m [40], those for reserve plants were €228 m [41]. Reasons for this increase include [32]: 2 new conventional power plants going online in the North, a nuclear plant in the South being shut early, the large addition of wind capacity, stronger winds in 2015 (feed-in increased by 38% compared to 2014), delays to grid reinforcements, temporary outage of grid elements due to upgrades, and high exports to Austria.

In 2016 costs for compensation of RES curtailment and redispatch decreased. Preliminary estimates are €230.4 m for redispatch (11,475 GWh of feed-in reductions and increases), €372.7 m for curtailment of 3,743 GWh of RES output, and €256 m for reserve plants [41]. The sum of feed-in increases for redispatch was 5219 GWh, corresponding to 0.88% of gross inland consumption. A main reason for the fall in curtailment and redispatch costs is the cost-optimised dispatch of plants by the TSOs [41], 96% of curtailed feed-in was from distribution connected plant, however 89% of all curtailment was caused by transmission grid congestion. 86% of curtailed feed-in was from onshore wind, 11.3% solar PV, 1.2% biomass [41]. Wind curtailment has increased 27-fold between 2010 and 2016 (Table 1).³

The Bundesnetzagentur expects costs of congestion management to remain high in the coming years since grid expansion trails behind the expansion of renewable energy. It has warned the costs for redispatch alone could exceed €1bn by 2020. Relief is expected in 2024 when major North-South grid infrastructure projects are scheduled to be finished [43]. The government also plans to slow down the expansion of renewable capacity significantly due to the increased costs of congestion management and delayed grid expansion [44].

To better assess the meaning of absolute numbers it is important to put them in relation to the size of the power market. Total turnover of the power sector without taxes was €484bn in 2014 in Germany [45].

2.3. Congestion management and its cost allocation in Britain

Britain’s short-term balancing and grid congestion management is mainly conducted via the Balancing Mechanism. This is in principle a short-term energy market which takes place in the hour after gate

---

¹ Forced curtailment of VRE feed-in can be caused by grid constraints or negative residual load. While the latter is not yet an issue in Germany or the UK, it has already been observed in Denmark [38].

² We use the terms constraint costs and congestion management costs interchangeably.

³ Curtailment of RES and CHP output is compensated; redispatch costs include costs of redispatch and countertrading.
Table 1

| Development of congestion management and wind curtailment in Germany; data: [11,32,41,42,46]. |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| curtailed wind [TWh]          | 0.13        | 0.41        | 0.36        | 0.48        | 1.22        | 4.12        | 3.53        |
| share of total compensated    | 98.7%       | 97.4%       | 93.2%       | 86.6%       | 77.3%       | 87.3%       | 94.4%       |
| curtailed energy              | 37.79       | 48.88       | 50.68       | 51.71       | 57.38       | 79.20       | 77.42       |
| wind feed-in (net generation) | 0.33%       | 0.83%       | 0.70%       | 0.92%       | 2.08%       | 4.95%       | 4.36%       |
| curtailment rate              | 10.2        | 33.5        | 33.1        | 43.7        | 183.6       | 478.0       | 372.7       |
| total curtailment compensation payments [€m] | 48.4 | 129.4 | 164.9 | 114.9 | 186.7 | 435.4 | 230.4 |
| Net costs of redi} | 48.4 | 129.4 | 164.9 | 114.9 | 186.7 | 435.4 | 230.4 |
| Costs of reserve plants [€m]  | -           | 17.0        | 26.0        | 56.0        | 66.0        | 227.8       | 256.3       |
| total congestion management [€m] | 58.6 | 179.9 | 224.0 | 214.6 | 435.7 | 1141.2 | 859.4 |

closure and before real time with the system operator as the single buyer. Generators and suppliers can submit bids and offers that state what they expect to be paid to increase or decrease production or consumption by an incremental amount. National Grid also contracts various reserve capacities to deal with energy imbalances and grid congestion.

Thermal generators usually offer to pay National Grid to decrease their output as they save on fuel costs. Wind and solar farms have no fuel cost but lose Renewables Obligation Certificates (ROCs) when they curtail their output, since they receive certificates (each worth ~£40) per MWh delivered to the grid. They expect to be compensated for lost revenue and thus are paid to reduce their output.

The costs for congestion management are socialised via the Balancing Services Use of System (BSUoS) charge, which is paid for by all suppliers and generators on a pro rata basis [47]. There is no locational element to the charging methodology [48]: Constraint management costs are shared proportionally to the energy consumed or produced during the corresponding settlement period.

Only generators which can participate in the balancing mechanism, mainly large transmission connected plants, get compensated for curtailment. However most of VRE capacity is connected to the distribution grid. These generators receive no compensation for curtailment, and must accept the risk of curtailment while benefitting from a lower cost non-firm connection [49].

2.4. Recent costs of congestion management in Britain

Britain’s constraint costs have increased in recent years from £86 m p.a. during 2005–08 to £303 m during 2014–17 (Fig. 3). Constraint costs have risen on average by £22.7 m per year over this period, or by £5.8 m for each TWh generated by wind (although, of course, other factors influence this). The biggest source of bottlenecks is insufficient network capacity between Scotland and England. Ongoing grid reinforcements are supposed to relieve these constraints [31,48], for example the Western link, a 2.2 GW subsea HVDC connection between Scotland and Wales, expected to be finished in 2017.

The government introduced a new policy called Connect and Manage regarding grid connections of new generation, which offers generators grid access before the completion of wider grid reinforcements necessary to accommodate additional electricity flows [51]. The policy was introduced to incentivise investment into new plant, in particular renewables, which would have been deterred by long grid connection delays. The long-term benefits of new generation capacity available to the power system are considered to outweigh the short-term costs of constraint management [53]. Constraint costs due to Connect and Manage made up only 3% of overall constraint costs in 2013, the main part caused by unavailability of transmission assets during the major programme of grid upgrades in Scotland and elsewhere [51]. However, constraints due to this regime increased as more and larger generators joined, from £69.4 m in 2014 to £121.7 m in 2015. Notably, National Grid expects these costs to drop to nearly zero in 2017/2018 [54].

2.4.1. Curtailment rates of wind farms

Due to the lack of public information on British wind farm curtailment, we collate original data on the quantity and cost of curtailment to explore the scale of this issue. Curtailment data on the level of individual onshore and offshore wind farms is available from the settlement company Elexon [55]. This covers farms which participate in the Balancing Mechanism, most of which are connected to the transmission grid – i.e. offshore farms, and larger onshore farms in Scotland only. Most of Britain’s wind farms are connected at distribution-level: 258 farms above 10 MW, compared with 38 connected to the transmission system [56].

We calculated the level of curtailment for 44 Scottish onshore farms and 18 offshore farms in England and Wales for the period 2012–16 using the method in [57], detailed in Appendix E. The onshore farms we observe represent around 30% of total onshore feed-in, and the offshore farms represent around 80% of offshore feed-in [1,56]. Of the 44 onshore Scottish farms we observe, 10 operate under Connect & Manage [58], 20 are embedded in the distribution grid, 24 are connected to the transmission grid.

Offshore wind farms show very little curtailment, with annual rates of 0.00–0.63%. These are unaffected by congestion on lines from Scotland, and would require higher compensation than onshore farms to be curtailed (due to forgone subsidies, see below), and so would be used as a last resort. Figs. 4 and 5.

In contrast, Scottish onshore farms have shown high curtailment rates over the last five years. Individual farms ranged from 0% to 32% of their annual output curtailed, depending on their location and operator preferences. 5 The volume-weighted average during 2012–16 is 10.61%, more than three times Germany’s average (2.98%). Maximum,

5 Farms which bid lower prices to be curtailed see higher rates (cp. Appendix E.).
volume-weighted average and plant average have all increased from 2012 to 2015, but decreased in 2016.

Combining the offshore and onshore data gives the overall curtailment rates of the observed wind farms: The volume-weighted average is shown in Fig. 6, and later in Table 2. If all this curtailed wind output had instead been fed into the grid, this would have raised Britain’s wind capacity factor from by 1–2% in 2015. With 5.68% curtailment observed in 51.2% of feed-in, the lowest value for gross capacity factor is 34.7% = (33.7% \cdot 51.2% + (1 - 51.2%)) / (1 - 5.68%). If all of Britain’s wind farms were curtailed at this rate, the gross capacity factor would be 35.7% = 33.7% / 1 - 5.68%.
2.4.2. Absolute curtailment, total costs and prices

The quantity and cost of wind curtailment has grown substantially in Britain, from 45 to 1123 GWh between 2012 and 2016, and from £5.9 m to £81.9 m (Fig. 7). While the share of wind constraint in overall constraint costs has increased, the bulk of constraint costs come from payments to gas plants (Fig. 8). It is also important to note that constraint costs were high (comparable to levels of 2014 and 2015) in 2008 and 2011, when VRE penetration was still at around 5% (Fig. 3).

Wind farms participating in the balancing mechanism will offer to reduce their output at the cost of their foregone revenues from selling Renewables Obligation Certificates (ROCs) and electricity. While onshore windfarms receive only 0.9 ROCs per MWh, offshore windfarms get 2 ROCs per MWh [61,62] and so request higher compensation and subsequently get curtailed less. Fig. 9 shows the monthly volume-weighted average price that National Grid paid wind farms to reduce their output, which has fallen 68% since 2011.

These curtailment prices are well above the ROC buyout and market prices, averaging £75/MWh in 2015, and £85/MWh in 2016. This is 1.9 and 2.2 times the ROC buyout price of £44–45/MWh in these years [63,64]. National Grid suggested that wind farms might place prohibitively high bids to avoid being curtailed due to bilateral contracts with suppliers who have an obligation to collect a certain number of ROCs [65]. These prices may also reflect a cautionary approach to potential wear and tear impacts, additional operational cost, obligations due to PPA provisions, or simply rent-seeking behaviour from operators [66].

2.5. Comparing curtailment in Britain and Germany

Congestion management costs have increased sharply in both Britain and Germany in recent years, due to expansion of renewable capacity and other developments such as new thermal plant coming online or ongoing grid upgrades. Constraint costs in both countries are expected to decrease once major transmission grid infrastructure projects are finished. The implications of constraint costs have to be compared to grid investment costs for it is more economic to accept some curtailment than to expand the grid indefinitely [10].

Wind curtailment rates have increased in both countries along with wind penetration. British wind farms are curtailed at a significantly higher rate than those in Germany, but our plant sample might not be representative for the complete wind capacity in Britain, as distributed wind farms outside of the Balancing Mechanism could not be analysed.
Nevertheless, we find that wind curtailment is a strongly localised problem in Britain, caused by congestion of the connection between Scotland and England.

Table 2 compares recent developments of wind curtailment and its compensation in Britain and Germany. Average prices per MWh of curtailed wind output have decreased in Britain, whereas in Germany they seem more volatile. Curtailment costs per MWh of feed-in are lower in Britain, since a smaller share of the wind farms is compensated (only BM-participating plant).

Fig. 10 shows curtailment rates versus penetration level of wind in Britain and Germany. British wind farms appear to suffer around twice the curtailment of German ones for a given penetration and curtailment seems to increase with penetration at a steeper slope in Britain than in Germany. These findings come with two major caveats: Onshore wind farms in Northern Ireland, and most distribution grid connected farms in England and Wales do not participate in the Balancing Mechanism and so could not be considered in the analysis.

3. Balancing systems of Britain and Germany

The balancing system is the set of technical and economic institutions that maintain active power balance of the electricity system over the period of seconds to hours [68]. Other ancillary services such as reactive power for voltage support and black start capabilities are reviewed elsewhere [69,70].

Supply and demand must always match to a high precision to maintain a stable system frequency. An imbalance of more than a few percent for a few seconds will lead to frequency deviating beyond tolerance levels that could damage generating equipment and lead to system collapse. In principle, this balance is ensured by the wholesale market: Suppliers contract electricity in the wholesale markets according to their demand and generators produce only as much electricity they deliver to or take off the grid. After delivery, every BRP’s schedule is compared to their actual delivery by trading with each other in the settlement period at 11:00 the previous day [76,77]. In Britain BRPs are called Balance and Settlement Code (BSC) parties, and deliver day-ahead schedules for each ½ hour period at 11:00 the previous day [78], which can be changed until one hour before real-time [73]. Intraday markets close 30 min (Germany) and 1 h (Britain) ahead of delivery [79].

After delivery, every BRP’s schedule is compared to their actual feed-in and consumption. In Germany, BRPs can adjust their schedules after delivery by trading with each other in the day after market until 16:00 the following day [34,71]. Such a market does not exist in Britain, but has been proposed by EdF [80]. Each BRP’s imbalance in a settlement period is multiplied with the imbalance price of that period to give the imbalance cost to pay. The calculation of imbalance prices is covered later in Section 3.3.

BSPs can be generators or consumers which reserve part of their capacity for pre-agreed time periods at the TSO’s disposal and are usually remunerated for capacity availability (€/MW per hour) and energy utilisation (€/MWh). They must complete a prequalification process to prove their capability to provide services to the required standards. Regulators determine the design of the imbalance settlement and the balancing power markets (e.g. procurement system, prequalification).

3.2. Balancing market design

3.2.1. The balancing market in Germany

Germany follows the ENTSO-E terminology [81]; balancing power is called control power and differentiated into three categories: primary control (PC), secondary control (SC) and tertiary control (TC) power.

Table 3

<table>
<thead>
<tr>
<th>Balance Market Design in Germany, based on [68,71,83].</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Control</strong></td>
</tr>
<tr>
<td><strong>Auction timing</strong></td>
</tr>
<tr>
<td><strong>Tenderable period</strong></td>
</tr>
<tr>
<td><strong>Availability windows</strong></td>
</tr>
<tr>
<td><strong>Minimum bid size</strong></td>
</tr>
<tr>
<td><strong>Pooling of capacity</strong></td>
</tr>
<tr>
<td><strong>Price components</strong></td>
</tr>
<tr>
<td><strong>Bid selection</strong></td>
</tr>
<tr>
<td><strong>Commitment</strong></td>
</tr>
<tr>
<td><strong>Remuneration</strong></td>
</tr>
</tbody>
</table>

Fig. 10: Curtailment rates against penetration levels for wind power in Britain and Germany; data: [1,11,42,56,59]. For context, China experienced 15% curtailment for a 3.2% penetration in 2015 [67].
Operating Reserves (STOR) are activated on a timescale of around 10 min to replace fast reserves and lost generation \cite{84,85}. In ENTSO-E terminology these approximate to PR, SR and TR respectively, although the boundaries are not completely aligned.

The reserve sizing methodology for Frequency Response and STOR is specified by the Security and Quality of Supply Standard (SQSS). Frequency Response is sized to cover the pre-defined largest loss event – currently 1800 MW, corresponding to the loss of half of the anticipated Hinkley Point C capacity. STOR is sized to meet the largest requirement and additional capacity beyond that is contracted based on forecast of available production and balancing market conditions and driven mainly by economic aspects. Typically, 2300 MW are contracted for \cite{47,86,87}.

Most balancing services are procured via competitive tender, as summarised in Table 4. Frequency response can either be provided as a mandatory service (for power stations \(\geq 100\) MW) or as a commercial service. Both are remunerated both in terms of availability and utilisation \cite{88,89}.

A primary difference between system management in Britain and Germany is the Balancing Mechanism. This can be seen as a reserve energy market operated after gate closure. Generators and suppliers can offer to increase or decrease production or consumption for a specified compensation. In the hour between gate closure and delivery, the TSO will accept some of these bids and offers according to expected imbalances in the settlement period, and BRPs’ contracted positions are adjusted accordingly.

The introduction of such a reserve energy market is currently discussed in Germany (see Section 4.4.4), where currently only BSPs who have secured bids in the capacity auctions can provide balancing energy. It is also envisaged in the draft of the European Network Code on Electricity Balancing \cite{93,94}.

Procurement cycles, contract and product lengths are longer which favours BSPs with large generation fleets \cite{95}. It is also suggested that there is over-procurement of balancing power and balancing activity is taken out of the market \cite{96}. However, long-term contracts give the TSO and BSPs an opportunity to hedge risk, might make BSPs bid at lower prices for long-term security, make investment in flexible capacity more attractive \cite{90,97} and are complemented by a balancing energy market (the BM) which efficiently provides short-term flexibility to the system \cite{79}.

National Grid has recognised the need to restructure and simplify balancing services to become more accessible, transparent and address changed system needs \cite{98,99}. Both countries have similar shares of availability (2/3) and utilisation payments (1/3) in balancing costs \cite{79}.

3.3. Imbalance settlement and cost allocation

3.3.1. Cost allocation and imbalance price in Britain

All costs incurred in managing the power system (including reserve availability and utilisation payments, reactive power and black start capabilities, accepted bids and offers in the balancing mechanism and the TSO’s internal costs) are socialised via the Balancing Services Use of System (BSUoS) charge. This is paid by every supplier and generator portfolio. It is calculated for each settlement period by dividing the costs incurred by the amount of electricity consumed and fed in during the period \cite{51}. Fixed costs like availability payments are spread across the year. This gives a price per MWh for system services, which is multiplied by the amount of electricity fed in or consumed by each party to calculate their charge for that period.

Since the Electricity Balancing Significant Code Review (EBSCR) of 2014/2015 \cite{100}, the imbalance price aims to reflect the marginal costs caused by energy imbalances (excluding other factors such as constraint management) and the scarcity of generating capacity \cite{73}.

3.2.2. The balancing market in Great Britain

National Grid, Britain’s TSO, procures several services to manage the transmission system. We focus on those which maintain and restore the active power balance. Frequency Response reserves are first to react, mostly within 10–30 s. They are activated automatically and help to keep the frequency from further decreasing. National Grid then instruct Fast Reserves, which must be activated within 2 min. Finally, Short Term

These must be fully activated within 30 s, 5 min and 15 min respectively. PC and SC reserves are automatically activated and in proportion to frequency deviation, TC reserves are scheduled in 15 min intervals at fixed values. A more detailed description is given in Appendix F.

PC reserves are shared throughout the Continental Europe synchronous system. 3000 MW are procured to compensate for the simultaneous loss of the two largest generators (1500 MW nuclear reactors) \cite{71}. Each balancing area procures a share of this reserve according to their total electricity generation, so in 2015 Germany procured 583 MW.

Sizing methods for SC and TC reserves differ significantly between European TSOs. Germany’s approach is based on the convolution of probability density functions to estimate the combined risk due to forecast errors in load and VRE output, plus plant and line outages. The reserve sizes are updated every 3 months, and for Q3 2017 were +1.92 and \(-1.84\) GW of SC, and +1.37 and \(-1.82\) GW of TC \cite{82}.

PC, SC and TC capacities are allocated in separate auctions on a common platform for all four TSOs \cite{82}. Table 3 summarises the current design of balancing power auctions, and Appendix F gives further details.

The current design was implemented by the regulator Bundesnetzagentur in 2011. It assumed responsibility in 2006 after concerns over inefficiency, lack of transparency, market barriers to new and smaller entrants and market power of incumbents \cite{83}. Minimum bid sizes have been reduced and the frequency of auctions have increased from biannually \cite{83}. The number of prequalified suppliers has increased substantially, and balancing power prices have either remained constant or decreased (Fig. 11). While prices are volatile and influenced by a variety of factors, more providers should increase competition and cost efficiency of balancing markets \cite{68}.
The settlement of imbalances is not maintained to recover costs but solely to incentivise BRPs to be in balance (see Appendix I for details of the British settlement). Historically, short BRPs (generating less or consuming more energy than contracted) paid the System Buy Price and long BRPs received the System Sell Price and these differed [101]. In November 2015, a single price system was established, i.e. short balancing groups pay the same price that long balancing groups receive [100]. This aims to reduce costs for small parties which are unable to influence the system imbalance and whose imbalances over longer periods usually average zero [47,51,101].

3.3.2. Cost allocation and imbalance price in Germany

While the costs for availability (capacity payments) are allocated among consumers via grid fees, utilisation payments are recovered through the imbalance costs that every BRP pays (or receives) for being out of balance [35]. The imbalance price is calculated by dividing the TSO’s net costs to utilise balancing power by the net balancing energy in each ¼ hour [34]. Some caveats to this calculation are given in Appendix I. Germany follows a single price system [83].

3.4. Balancing Costs of VRE generation

3.4.1. Impact of VRE plant on reserve requirements

VRE plants cause balancing costs due to output forecasting errors which must be offset by operating dedicated reserve or peaking plants. Many studies find that balancing costs should rise with wind penetration because of increased reserve requirements, with a wide range of estimates given [22,23]. Reasons include differences in sizing methodologies and other power system characteristics, and primarily the different timescales of uncertainty assumed. Reserves must be substantially larger if used to offset day-ahead forecast errors rather than errors of forecasts 1 h before real time (gate closure in many electricity markets).

Several meta-studies report an increase of secondary and tertiary reserves by 2–11% of installed wind capacity up to penetration levels of 20%, while primary reserves remain unaffected since they are mainly determined by the largest contingency [23,68,102]. One industry approximation is the “3+5 rule”: reserves equalling 3% of demand plus 5% of wind generation [103].

Increased reserve requirements do not necessarily imply investment in new reserve capacity, as plant which formerly provided energy could be used for reserve services [23]. Downward reserves in particular are reported to be used more frequently with higher VRE penetration, but such reserves could be provided by curtailing excess wind power while still keeping curtailment at 1% or less of total wind energy [102,103].

3.4.2. Factors determining balancing costs

The magnitude of VRE forecast errors is a primary driver for the size of balancing costs. Key factors in the size of these errors are:

- the installed VRE capacity;
- the accuracy of forecasting tools;
- the time horizon of forecasting [104];
- the mix of VRE technologies (e.g. if PV has lower forecast errors than wind); and
- the geographic spread of VRE plants (as a larger area yields less correlated outputs and errors, and thus smaller aggregated errors) [20].

Several other factors influence balancing costs [22]:

- The correlation of VRE forecast errors with load forecast errors and other imbalances;
- The flexibility of the residual plant mix (e.g. flexible hydro can provide lower cost balancing than large coal units) [105,106];
- The design and liquidity of intraday and balancing markets; and
- What costs are rolled into the study: some include only the costs of reserves while others also include costs of intraday trading and

---

Table 4

Balancing market design in Britain; based on: [90–92].

<table>
<thead>
<tr>
<th>Frequency Response</th>
<th>Tender timing</th>
<th>Tenderable period</th>
<th>Availability windows</th>
<th>Minimum bid size</th>
<th>Pooling of capacity</th>
<th>Price components</th>
<th>bid selection</th>
<th>Commitment</th>
<th>Remuneration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>monthly</td>
<td>1 ~ 23 months</td>
<td>at least 2 h for at least 1 month (recommended)</td>
<td>± 10 MW (± 1 MW from Apr – 2017)</td>
<td>allowed</td>
<td>capacity &amp; energy</td>
<td>cost benefit analysis</td>
<td>energy price merit order</td>
<td>pay at bid (capacity and energy)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5

Balancing costs for wind and solar reported in the literature.

<table>
<thead>
<tr>
<th>Study</th>
<th>Context</th>
<th>Energy Penetration</th>
<th>Balancing Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heptonstall [21]</td>
<td>Various countries, 2016, model</td>
<td>VRE up to 30%</td>
<td>£0-5/MWh</td>
</tr>
<tr>
<td>Sijm [102]</td>
<td>Various countries, 2014, model</td>
<td>VRE up to 30%</td>
<td>£1-6/MWh</td>
</tr>
<tr>
<td>Holttinen [23]</td>
<td>In thermal systems, 2013, model</td>
<td>VRE up to 50% in UK/Ireland</td>
<td>£15-45/MWh</td>
</tr>
<tr>
<td>Holttinen [23]</td>
<td>In hydro systems, 2013, model</td>
<td>VRE up to 50% in Germany</td>
<td>£2-4/MWh</td>
</tr>
<tr>
<td>Pudjianto [107]</td>
<td>Various EU countries, 2013, model</td>
<td>VRE up to 50% in Denmark</td>
<td>£0-6/MWh</td>
</tr>
<tr>
<td>Hirth [22]</td>
<td>Various countries, 2015, model</td>
<td>VRE up to 50% in Spain</td>
<td>£0-13/MWh</td>
</tr>
<tr>
<td>Hirth [22]</td>
<td>Historic (market data), 2015</td>
<td>VRE up to 50% in Germany</td>
<td>£1-7.5/MWh</td>
</tr>
<tr>
<td>Holttinen [23]</td>
<td>Historic (market data), 2013</td>
<td>VRE up to 50% in Denmark</td>
<td>£1-4-2.6/MWh</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VRE up to 50% in Spain</td>
<td>£1-3-1.5/MWh</td>
</tr>
</tbody>
</table>

---
3.4.3. Balancing cost estimates

Table 5 lists several estimates of the balancing costs imposed by VRE generation, with more detail given in Appendix C.

Balancing costs are relatively low even at higher penetration levels up to 40% of electricity demand. Estimates for grid-related integration costs are also modest, typically single-digit €/MWh [22], which suggests that the flexibility already built into power systems to cope with demand and generation uncertainty will accommodate the additional uncertainty from a significant VRE share [108,109]. However, integrating higher levels of VRE (e.g. > 50% of demand) will require transformation of the power system with large scale deployment of alternative flexibility options such as demand-side management (DSM) and storage to maintain cost efficiency [108,110,111] as well as transformation of power markets to provide and remunerate short-term flexibility sufficiently [84,112,113].

It is also important to note that while most VRE integration literature focusses on balancing costs [114], these make up a smaller part of energy bills than costs of grid infrastructure. In Britain combined balancing and constraint costs account for 1–1.5% of bills [112,115], whereas grid infrastructure accounts for about 15% [48]. Similarly, system operation costs made up 1% of Germany’s residential retail electricity price in 2015 (0.31 of 29.80ct/kWh) [19]. For context, electricity production costs made up 40% of domestic electricity bills in Britain in 2015 and 46% in Germany in 2015 (including renewable support) [40,116].

3.5. Recent developments of balancing costs

3.5.1. Germany

The socialised costs of ancillary services in Germany have reduced significantly between 2007 and 2014, mainly due to balancing costs falling despite VRE capacity tripling [19] (Fig. 12). However in 2015 they reached their highest value in the last decade due to increasing grid congestion. While decreasing in 2016, total costs were still 26% higher than in 2010.

Fig. 13 shows the absolute size of German balancing markets, which has fallen 72% between 2010 and 2016. Capacity payments form about 2/3 of the market, the bulk of which are for SC reserves. Total balancing costs and contracted reserve capacities have fallen due to recent developments in German balancing markets. The four TSOs are increasing cooperation, merging the four balancing areas into one and working with European neighbours to apply imbalance netting; reformed regulation improved competition and efficiency in the balancing market, and intraday liquidity has increased [35,68].

3.5.2. Britain

In Britain, the BSUoS charges (which recover all system operation costs) have constantly increased in recent years, from £746 m in 2010 to £1207 m in 2016 (+62%) whilst VRE capacity has quadrupled (Fig. 14).8

The right panel of shows the market size of the balancing market in Britain, which has been reducing gradually since 2013. Frequency response constitutes the bulk of costs, unlike in Germany, where the share of analogous PC reserves is much smaller. This is partly because Britain is an island system that cannot share reserves over a wider synchronous area [118], and also because the characteristics of British frequency response overlap with SC reserves to some extent.

The average BSUoS price has quadrupled over the last 15 years to £2.47/MWh during summer 2016 (Fig. 15). This energy-weighted average is increasing by £0.16/MWh a year; a trend which precedes the

---

7 2016 numbers for reactive power, black start and losses were not available at the time of writing, and so were estimated to remain the same as 2015 values.
8 Fast Reserve: availability & utilisation excluding bids and offers acceptances in the Balancing Mechanism; STOR: availability & utilisation; Frequency Response: availability only.
development of VRE capacity. Wind output is correlated with higher BSUoS costs and solar with lower costs; when weighted by wind output the average cost is 17% above the energy-weighted average, and the solar-weighted average is 23% lower. These are diverging as VRE capacity rises: the spread between wind- and solar-weighted BSUoS charge is increasing by 6.5% points per year. June 2017 recorded the highest balancing costs on record: £3.68/MWh time-weighted; £4.44/MWh wind-weighted and £3.17/MWh solar-weighted.

PC and frequency response reserves are sized according to the largest generator that could be lost, which is usually a nuclear plant [71,118]. It is argued that these costs should be allocated to those who cause them [68], which would act as a strong disincentive to build ever-larger units (such as Hinkley Point C). National Grid expects that by 2030 primary frequency response requirements will have increased to 3–4 times the level of 2015, mainly due to new-build nuclear increasing the largest infeed loss [119]. Strbac estimates that reducing the largest generator from 1800 to 500 MW lowers VRE system integration costs by 44% due to lower frequency response requirements and thermal must run [110].

A look at the monthly balancing costs in Britain reveals a mixed picture. The total costs of STOR, Fast Reserve and Frequency Response have decreased by 6% (Fig. 16). The trend over the whole period saw the cost of frequency response and fast reserve rise by 5% and 26% respectively, while the cost of STOR fell by 40%; however, it should be noted that these costs have all reverted to their pre-2012 levels in recent months [50].

Fig. 17 shows that the contracted capacity of STOR has varied over a wide range while gradually increasing. The volume of fast reserve has increased, while frequency response has decreased [50].

Fig. 18 shows that holding volumes of mandatory frequency response provided by large thermal stations decreased by 50%, while volumes of commercial frequency response increased by 26%. The cost of procuring mandatory frequency response decreased by 62% while costs of commercial response have increased by 72%. Declining mandatory response volumes might reflect the retirement and lower
utilisation of coal plants: the share of electricity generated by coal has fallen by three-quarters between 2012 and 2016 whilst half of the capacity (8 plants, 14 GW) has closed down [1, 57].

Fig. 19 shows the average price for each service. While prices for Fast Reserve and Commercial Frequency Response have increased, prices of Mandatory Frequency Response and STOR have decreased, the later due to oversupply of STOR capacity because of low spark spreads increasing the attractiveness of balancing markets to generators [96]. Commercial Frequency Response prices vary between £20–45/MW/h while Mandatory Frequency Response prices are mostly between £3–5/MW/h.

Balancing costs per MWh of demand were about 30% higher in Britain than in Germany in 2015, but both are lower than average for Europe [120].

4. Policy options to address balancing costs

Several options exist to reduce balancing costs, which can be classed as [10]:

(i) improved forecasting,
(ii) improved spot markets and
(iii) improved balancing management and markets (including

European integration).

We first discuss the potential of (i) and the main mechanism to incentivise BRPs to contribute to it: The imbalance settlement (Section 4.1). (ii) can be seen as the tools BRPs must be given to act on the imbalance settlement's incentives: liquid and refined spot markets (Section 4.2). Finally, we look at (iii), and ways to improve operation (Section 4.3) and design of balancing markets (Section 4.4). The emphasis of this section is on increased balancing activity and responsibility of VRE generators, so other ways to reduce balancing costs such as TSO cooperation or support of storage and DSR are covered in less detail.

4.1. A stricter imbalance settlement

Fig. 20 shows an example of imbalance prices in Germany [71].
IEA reports a mean absolute error (MAE) of 3% of installed wind capacity in Germany for day-ahead forecasts and 1% 1 h ahead of real time [126]. National Grid aims for 3.25-4.75% accuracy in its day-ahead wind energy forecast [127], and German TSO 50 Hz states that actual wind output deviates by 2% (RMSE) from day-ahead forecast [128,129].

Data mining has gained prominence for improving forecast accuracy [130]. Including hub height wind speed and wind farm operational data (in addition to satellite data) into forecasting improves accuracy of output and ramping rates significantly [131]. Such data is often not made available by wind farms due to privacy laws or simply lack of information technology [132]. An obligation for wind farms to share some performance and measurement data with forecast providers, central meteorological institutions or TSOs may be a promising route forwards. However, such forecast improvements need to be incentivised via an efficient imbalance settlement.

4.1.2. The imbalance price: incentive and price signal

A central part of Germany's electricity market reform is to strengthen the obligation of BRPs to uphold commitments, ensuring their contracted consumption and generation matches actual consumption and generation. This is seen as the essential market mechanism (along with free price formation) to address long-term security of supply and short-term balance of the system. High peak prices in the spot markets and imbalance prices should incentivise a more active management of balancing groups. This would decentralise system management by deferring system balancing responsibilities from the TSO to market participants and thereby fuel competition of different flexibility options. In combination with decreasing technology costs of storage [133], marginal imbalance prices might make combining a wind farm with storage economic to avoid high imbalance charges [134]. BRPs will want to hedge against peak prices and imbalance risk by increasing intraday trading and entering long-term supply and option contracts [34]. Such contracts offer possibilities to remunerate capacity and should thus contribute to long-term security of supply by ensuring that flexible capacity can be financed. Similarly it is argued in Britain that the imbalance price reflects the marginal value of electricity in the power system, signalling to investors what remuneration they can expect for offering flexible capacity [47,51,101,135]. The imbalance price thus serves two purposes: It incentivises balancing activity by BRPs but is also a price signal for investment in new capacity.

Chaves-Ávila compares balancing systems and imbalance settlements in the Netherlands, Belgium, Denmark and Germany, and concludes that average pricing and the single price system in Germany reduce incentives to BRPs to be balanced [136]. Hirth argues the imbalance price in Germany is too much regarded as a vehicle to allocate balancing costs and should be seen as an economic incentive to BRPs to avoid imbalances [68]. Imbalances can be reduced in many ways: by improving forecasting tools, updating forecasts more frequently, trading more actively on intraday markets, shifting from hourly to ¼ hourly scheduling and dispatching assets more accurately [68]. BRPs will deploy and improve such methods only as long as they are provided with a return on investment, which is determined by the imbalance spread, the difference between the imbalance price and the spot market price.

For efficient resource allocation, the imbalance price should reflect the marginal costs of deploying balancing power, not the average costs. Operators should therefore introduce either uniform pricing in balancing markets (instead of the current pay-as-bid mechanism) or imbalance pricing based on marginal costs (the energy price of the most expensive reserve deployed). Note that the second approach allows the TSO to receive more revenue than its expenditure.

While Britain has made imbalance prices better reflect marginal balancing costs, several options to change the current German imbalance price formula (based on average costs) are being discussed in response to a government white paper [34]. Notable among these are the inclusion of capacity payments [137,138] (opposed by ACER [139]), and the use of uniform pricing [93,140]. See Appendix I for more detail.

4.1.3. Passive balancing

BRPs can react to imbalance price signals by deviating deliberately from their schedule to reduce the system imbalance: so-called passive balancing. Hirth argues that passive balancing by BRPs could be a good alternative to a reserve energy market suggested by ACER [68,94,139].

To enable passive balancing, balancing prices or system imbalances have to be published close to real time to allow BRPs to react to prices signals [136]. Imbalance prices are published every 1 and 2 min in the Netherlands and Belgium, 30 min after real time in Britain [55], but not until 1 month afterwards in Germany [136].

Chaves-Ávila and van Hulle both promote passive balancing [39,141]. However, the Bundesnetzagentur's stance is ambivalent: One reason for establishing a one price system is that imbalances which benefit the system should not be punished. However, BRPs retain legal obligations to be in balance with their contracted positions [142]. Each balancing area’s imbalance as well as combined imbalance across Germany in each ¼ hour is published no later than 15 min afterwards [82]. The Bundesnetzagentur has rejected proposals to publish system imbalances in real time, as it fears this would incentivise passive balancing [93]. Britain’s Ofgem has stated that one reason to move to a single imbalance price was to avoid disproportionate efforts by BRPs to balance their positions since this might be uneconomic from a system perspective [47,101].

4.2. Enabling balancing by market participants

Increased liquidity on spot (mainly intraday) markets, gate closure closer to real time and shortening trading intervals can transfer more balancing responsibility from the system operator to market participants and make the balancing process more efficient with more accurate forecasts being available closer to real time [10,126,143].

The main incentive for BRPs to accurately forecast their output comes from the imbalance price. A prerequisite for BRPs to act on these incentives and self-balance is liquid day-ahead and intraday markets [141]. Providing this responsibility will also increase the liquidity of these markets [120]. Shorter trading intervals help to reduce intra-interval imbalances enabling suppliers and VRE generators to better approximate their actual demand/output profile by their contracted profile [68] and thus reduce the need for balancing energy (Fig. 21) [144].

The 2015 solar eclipse across Europe demonstrated the balancing capability of liquid intraday markets. Less than a fifth of the balancing capacity that TSOs had anticipated was utilised, since the timing of the eclipse was precisely known, allowing market participants to trade on the intraday market [34].

In Germany, power can be traded in 1 h as well as in 15 min contracts (the later since 2011) [35,145,146]. In addition to continuous trading, auctions were introduced in 2014, which have proven attractive to small VRE generators and increased intraday volumes [79].

The intraday market volume has grown continuously in the last years (Fig. 22). Intraday schedule changes have increased significantly: from 0.1 million in 2008 (a volume of 20.5 TWh) to 2.8 million (134.9 TWh) in 2015 [40]. One reason for this is the fluctuating feed-in of VRE that is balanced in the intraday market [42].

Intraday markets are a rather new trading platform and remain an active field of research, regarding their design, market participants’ behaviours, the interplay with day-ahead and balancing markets, and their increasing importance for balancing fluctuating VRE feed-in [147–150]. Their design needs to be coordinated with the design of day-ahead markets [149] and the imbalance settlement [151] to avoid arbitrage opportunities, as well as with balancing markets to avoid liquidity moving from intraday to balancing markets [93]. Balancing and
intraday approaches of neighbouring markets should be aligned to enable further integration of European markets and cost benefits [152]. A further question is whether continuous trading or discrete auctions are better suited for balancing purposes [148,153]. Genoese shows that the direct selling option in Germany is highly attractive for VRE generators and incentivises more accurate forecasting and clearing of forecast errors by intraday trading [154].

4.3. Improving balancing operation

Options to improve balancing operation mainly relate to reserve sizing and TSO cooperation. Reserves should be sized dynamically since demand for them depends on instantaneous conditions of the power system. Adjusting the size of contracted balancing power according to system needs on a more frequent basis avoids over-procurement of power [139,155–157].

Larger balancing areas allow forecast errors to cancel each other out. In 2009–10, following an investigation by the Bundesnetzagentur into balancing efficiency, the four German TSOs established a co-operation called "Netzregelverbund" or Grid Control Cooperation (GCC), which integrated their four balancing areas into one, with joint reserve sizing and activation [68,83]. TSO cooperation is viewed as essential to integrate larger amounts of VRE [103]. Four stages of increasing cooperation can be distinguished: (i) imbalance netting, (ii) common reserve dispatch, (iii) common reserve procurement and (iv) common reserve sizing. The EU intends to harmonise and integrate balancing markets of its member states, led by the Association of European Energy Regulators (ACER) and TSOs (ENTSO-E). The benefits from integrating European balancing markets are estimated at £1.3–2.7bn per year [158].

Bottom-up initiatives such as the International Grid Control Cooperation (IGCC) have already emerged and proven highly beneficial at the stage of imbalance netting [40,159,160] and advanced integration can provide further benefits [120,157,161–164].

4.4. Open balancing markets to VRE

The bulk of reserve capacity and energy is provided by large fossil and hydro plants [71]. Competitive and transparent balancing markets require barriers to be removed for new providers of balancing power. More frequent auctions (e.g. daily) and shorter lengths of balancing products (e.g. 1 h) would allow more and alternative providers such as demand side response (DSR), storage and VRE to participate in balancing markets [34,40,68,165–168]. This provides additional revenue streams for these providers and should also make balancing markets more competitive and efficient, thereby reducing costs.

We first focus on technical and economic aspects of VRE providing balancing power (Sections 4.4.1 and 4.4.2), then discuss system benefits of such provision (Section 4.4.3) and existing market barriers (Section 4.4.4).

4.4.1. Technical capabilities

Wind farms are well-equipped to provide balancing power due to their ability to change output very quickly over a wide range without increasing maintenance costs significantly. In contrast, ramping thermal plants is often restricted by high minimum stable generation, limited ramping rates and thermal fatigue [68,119,169].

Wind and solar farms already meet most requirements for frequency support, reactive power and some system restoration services as prescribed in grid codes. Where enhanced capabilities are required, technological solutions exist (fast, reliable communication and controllers) but are not economic in current market frameworks [170]. Costs for installing the necessary hardware to provide frequency support are negligible (below 1% of turbine capital costs), but opportunity costs can be high.

Pilot projects have shown the capability of wind farms to provide balancing services with high reliability: downward SC in Belgium [169], and TC in Germany [171]. In a two-year pilot with the four German TSOs, a pool of two wind farms (89 MW) qualified for provision of 60 MW negative TC capacity [172]. In Denmark wind farms are...
already actively participating in balancing markets providing negative balancing energy to increase their revenue [39,173].

A particular challenge for wind farms providing reserve capacity and energy is quantifying the capacity that can be offered and the utilised reserve energy. Probabilistic forecasting enables wind parks to offer balancing power with the same reliability as conventional providers [174]. The share of forecasted feed-in (in terms of capacity) that can be offered using this method can be increased with shorter forecast horizons, shorter product lengths and pooling of wind farms. The quantification of delivered reserve energy is more challenging, with solutions that either have insufficient accuracy to meet current standards or impose high opportunity costs [170,174]. See Appendix M for more detail.

Inertia is another critical area. Large thermal synchronised generators dampen the rate of change of frequency (RoCoF) of the system during contingencies such as losing a plant, passively through the mechanical inertia of their rotating turbine mass. Where wind is connected by power electronics to the system it does not contribute to system inertia at present. System inertia is falling with more wind farms going online and large coal and nuclear stations retiring, making frequency changes more rapid and difficult to manage [84,109]. This issue is more emphasised in the debate in Britain than Germany, since the effects are more significant in islanded systems than large interconnected systems [118]. However, with appropriate frequency controllers wind plant could provide fast frequency response similar to that of conventional plant [118].

4.4.2. Opportunity costs and coordination with subsidies

Providing positive reserve would require wind plants to continually limit their output, resulting in high opportunity costs. Therefore most studies argue that only provision of negative reserve would be economic [68,169,175]; however, this changes at times when prices are negative [173]. Since 2014, German RES subsidies cease when wholesale prices are negative for at least 6 consecutive hours [176], meaning provision of positive reserve capacity would have zero opportunity costs for a wind farm. Strbac argues there could be other system conditions when the marginal value of reserve would be higher than the marginal value of energy [110]. National Grid is considering new service provision arrangements to incentivise VRE generators to de-load constantly to offer frequency response [119].

Hirth argues that providing negative reserve comes at zero opportunity costs for wind farms, but does not for thermal generators when market prices cannot cover their operational expenses but they must still operate at their minimum stable generation plus the offered reserve capacity [68]. In addition to providing reserve availability, the opportunity costs of utilising that capacity must also be considered. If VRE generators receive subsidies per MWh fed into the grid, as in Britain and Germany, they lose subsidies when their negative reserve capacity is utilised. Furthermore, if output is curtailed that had not been sold beforehand, the farm loses the payments for this output via the imbalance settlement.

This structure of opportunity costs for VRE generators providing balancing reserves suggests that their most attractive option might be offering downward reserve capacities which are rarely utilised, i.e. SC and TC reserves. In 2014 Germany used only 7–8% of the SC and TC reserve volume tendered [35]. Enabling VRE generators to provide these reserves instead of fossil plant would also replace more fossil electricity per MW of reserve capacity than in the case of positive reserves. As mentioned before, capacity payments make up 2/3 of overall balancing costs in Germany [83].

In order to make provision of negative reserve economic, either wind farms must be compensated at least by the opportunity cost of foregone subsidy [39,65], or subsidies should be paid for output that was curtailed to provide negative reserve energy [177]. It is argued that a subsidy proportional to energy inhibits VRE generators from providing negative reserve and reduces VRE generators’ incentives to be in balance, and should therefore be avoided or adapted [169,178]. Britain is also discussing whether provision of certain balancing services should be made mandatory for VRE generators, as for large thermal generators [110]. Since only a fraction of the installed capacity would be needed to provide sufficient frequency support, this approach is considered ineffective [118], and would also inhibit innovation and competition [141,167].

4.4.3. Benefits

The provision of balancing services mainly by thermal generators leads to high thermal ‘must run’: plants avoid turning off as they must provide reserve capacity even if cheaper renewable electricity is available [179]. Therefore, high penetrations of VRE and corresponding emission reductions cannot be achieved without enabling VRE generators to provide balancing services [110,118].

Strbac estimates that wind providing inertia and primary frequency control reduces integration costs by 54% [110], as less part-loaded thermal plant is needed to provide these services and subsequently more renewable energy can be accommodated, allowing higher VRE utilisation. Teng shows that provision of synthetic inertia by wind plants in a Britain power system with 60 GW wind reduces system costs by a factor of 10 [118]. Jansen shows SC costs would be reduced by 24% in Germany if wind could provide it [180].

The system benefit of wind providing downward reserve is especially high in times of low demand and high wind output [175]. Provision of upward reserve by wind plants is argued to reduce system operation cost, even if it is remunerated at the lost opportunity cost [181]. Participation of wind plants in balancing markets, actively or passively, would also increase operational efficiency since it improves communication and coordination between system operators and VRE generators [39].

Opening balancing markets to smaller and alternative providers also offers revenue streams for important flexibility technologies like storage and demand side response. Furthermore, a larger number of providers increases competition in the balancing markets and should thus improve efficiency [40].

4.4.4. Market barriers

While VRE technologies have sufficient capabilities to provide balancing power, barriers to market participation remain due to inadequate market design [177]. These prevent “non-discriminatory, fair, objective, transparent, market based and economically efficient” procurement as demanded by ACER [182]. Three market reforms are mainly discussed to eliminate these barriers [68]:

(i) changing the procurement regime: shorter lead times and product lengths, smaller minimum bid sizes, allowing pooling;
(ii) an energy only market for reserve energy with gate closure close to real time; and
(iii) passive balancing.

Balancing reserves are often procured several days to weeks in advance (even years in Britain), for time periods of several hours to months (Table 3 and Table 4 above). For such time horizons, the uncertainty and variability of wind resources is usually too high: the size of forecast errors and variation of feed-in capacity mean only a small share of VRE output can be reliably offered as reserve energy. Calendar day procurement and hourly product lengths for all three kinds of reserve have been proposed in Germany [140]. Elia suggest daily procurement and 4 h blocks of SC reserves to enable wind participation; estimating that individual farms could go from offering just 1–65% of their in-feed as downward balancing when moving from monthly to daily procurement [169].

Positive and negative reserves should be procured separately to enable more providers to compete [79,173,183]. Minimum bid sizes should be reduced as far as possible to enable smaller plant to
participate [165,182,184]. Pooling wind farms to provide balancing services should be enabled, since it reduces uncertainty of the output due to geographical smoothing. Furthermore, harmonisation and integration of balancing markets should be increased to enable cross border participation and sharing of resources [183]. The lack of clear specifications and requirements by TSOs is an obstacle to wind farms enhancing their technical capabilities to provide ancillary services [170]. Pay as bid remuneration instead of uniform pricing of products can consolidate market power: It incentivises participants to bid as close as possible to the estimated marginal price, thus disadvantaging small players with less capabilities to forecast these prices [185].

The German government has initiated reforms of balancing markets addressing some of the issues above [34] (see Appendix K for further details) and National Grid has started a process to reform the procurement system of balancing services in Britain which currently still blocks alternative providers to participate [99]. Chaves-Ávila argues that procurement of reserves via organised markets is more transparent than bilateral contracting, and that Britain’s transparency could be increased if some of National Grid’s many balancing schemes were merged and opened to alternative providers [39].

On the other hand, National Grid reduced the minimum bid size for frequency response to 1 MW [186], and wind farms actively participate in the Balancing Mechanism – thus realising the balancing energy market envisioned by ACER, which is currently absent in Germany. Regulation in Britain seems to be more favourable towards passive balancing than in Germany; however, it is not actively encouraged in either market (Section 4.1.3).

4.5. Are balancing costs externalities?

In Germany and Britain most VRE generators have to sell their power on the wholesale market to participate in subsidy schemes [167–189] (see Appendix J), and have the same balancing responsibilities as other market participants. The obligation for new VRE plants to sell their output directly (as opposed to the TSO having to sell it) is seen as the first step towards greater exposure of VRE to market signals in Germany [154,176,190]. A goal of policy is to incentivise VRE generators to adjust their positions through trading in the spot markets, reducing the use of more expensive balancing reserves by the TSO after gate closure [191,192]. The European Commission advises that VRE should bear balancing responsibilities and recommends establishing liquid intraday markets to enable them to be in balance with their contracted positions. Furthermore, their costs from balancing responsibility should be considered when assessing the optimal level of support [193,194]. ENTSO-E states RES producers should be bound by the same duties and responsibilities as all other electricity generators. Providing incentives for RES producers to correctly forecast their feed-in and hedge their volatility improves system security and economic efficiency [177,195]. Britain’s recent imbalance price reform already seems to have increased some VRE generators’ balancing activity [196]. Due to the higher ratio of their imbalances to their output, VRE are more exposed to uncertainty and imbalance risk than conventional generators [197,198]. Rather than focusing on system operation cost, ongoing grid fee reforms in Britain and Germany aim at allocating transmission grid investments to some extent with distributed generation, due to increased power flows from distribution to transmission grids and perceived market distortions caused by exemption of distributed generators from transmission grid charges [199,200].

Increased responsibility for, and contribution to, system balancing has been required from German wind farms in the form of obligatory capabilities to handle a wider range of grid frequencies (47.5–50.2 Hz) [201] and the ability to monitor and control output remotely [202]. Since 2012, wind farms in the wholesale market are allowed to provide balancing power, and can earn prices on the spot market above their strike price. This increases wind generators’ exposure to the market value of their output, giving the incentive to produce a less fluctuating output profile using advanced wind turbine design [34,203].

Despite VRE having the same balancing responsibilities, it is argued that the imbalance settlement regime may not be cost reflective, as costs are not allocated in a way that who causes these costs also bears them [204]. Around 2/3 of balancing costs (the availability costs of reserves) are socialised in Germany and Britain [79]. While VRE generators pay imbalance charges like all BRPs, part of the balancing costs they cause is socialised in form of higher grid fees and imbalance prices for all participants [105]. However, cost allocation based on causation may be impossible, or associated with high transaction costs due to the complexity of interactions in the electricity system [143], so reforms to make the imbalance settlement more cost reflective will always have some shortcomings.

5. Conclusions

This paper examined recent developments of short-term integration costs of variable renewable energy (VRE) sources, congestion management and balancing costs, in Britain and Germany. Congestion management costs have increased recently in both countries to unprecedented levels. In Germany, these costs are dominated by compensation payments to wind farms for curtailment, whereas in Britain the costs for redispatch of gas plants make up the bulk of costs. Ongoing and planned major grid upgrades are expected to relieve congestion to a large extent in both countries. We show that 4.4% of German and 5.6% of (metered) British wind energy was curtailed in 2016, a total of 4.65 TWh. This energy cost €426 m to curtail, and could have saved approximately 2.1 MT of CO2 if it had been utilised.10

Balancing costs have stayed constant in Britain and have decreased in Germany, despite VRE capacity increasing five-fold and two-fold respectively. This goes against most modelling studies which predict an increase of reserve requirements and costs with increasing VRE penetration levels. This supports the hypothesis that significant flexibility already exists in these systems, and so accommodating more intermittent renewable output can be managed by improved system operation without necessarily raising costs.

Three policy goals should be pursued to enable and incentivise such improved system operation by system operators as well as market participants:

i) a stricter imbalance settlement,
ii) refined and more liquid intraday trading, and
iii) transparent, competitive and liquid balancing markets.

The first is discussed in Germany and already practised in Britain, and will incentivise more accurate forecasting and better portfolio management by VRE generators as well as suppliers. The second has been facilitated in Germany by a move to 15-min trading and moving gate closure to 30 min before real time, Britain still has half hourly trading intervals and gate closure 60 min before real-time.

The German balancing market was reformed in 2011 to improve transparency and competition. Since then prices and absolute costs have fallen. Further reforms are enabling more potential providers of balancing power, in particular VRE generators to participate. Wind turbines have sufficient capability to provide all three kinds of control power reserves, but if prevented from doing so, VRE output will have to be suppressed by must-run fossil stations providing reserves. Therefore rearranging balancing power procurement to enable VRE participation is necessary to reduce both costs and carbon emissions. The costs and benefits of making specific services mandatory or voluntary have to be weighed up.

Britain should move to a balancing system with shorter product

10 Assuming carbon intensities of 470 kg/MWh in Germany and 350 kg/MWh in Britain.
lengths and more frequent auctions instead of tenders. This could also help the system operator to merge products and improve transparency and accessibility of balancing markets. The short-term balancing energy market suggested by ACER is already in place in Britain while Germany lacks such a market. Harmonisation and integration of European balancing markets as well as bottom-up TSO cooperation help to reduce balancing costs and should be pursued further.

In summary, integrating large amounts of wind energy into power systems has not proven to be the major impediment that was once feared. System operators are managing this integration better than expected, but there is plenty of scope for countries to still learn from one another.

Acknowledgements

We thank Alex Coulton (Low Carbon Contracts Company), Eamonn Bell (RenewableUK), Phil Heptonstall and Rob Gross (Imperial College London), Dimitri Pescia (Agora Energiewende), Joachim Nick-Leptin (RenewableUK), Phil Heptonstall and Rob Gross (Imperial College London), Dimitri Pescia (Agora Energiewende), Joachim Nick-Leptin (RenewableUK) and not necessarily represent the views of the authors, and do not necessarily reflect their organisations.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.rser.2018.01.009.

References

### September 2016:


106. 50 Hz. Personal communication with Wilhelm Appler, energy policy manager; 2016.


