Demonstrating demand response from water distribution system through pump scheduling

Ruben Menke a,⇑, Edo Abraham a, Panos Parpas b, Ivan Stoianov a

a Department of Civil and Environmental Engineering, Imperial College London, SW7 2BU London, UK
b Department of Computing, Imperial College London, SW7 2BU London, UK

HIGHLIGHTS

• Water distribution systems can profitably provide demand response energy.
• STOR and FFR are financially viable under a wide range of operating conditions.
• Viability depends on the pump utilisation and peak price of the electricity tariff.
• Total GHG emissions caused by the provision of reserve energy are <300 gCO2/kW h.
• These are lower than those from the major reserve energy provision technologies.

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form 23 February 2016
Accepted 25 February 2016

Keywords:
Demand response
Water distribution systems
Pump scheduling
GHG emission mitigation
Optimisation

ABSTRACT

Significant changes in the power generation mix are posing new challenges for the balancing systems of the grid. Many of these challenges are in the secondary electricity grid regulation services and could be met through demand response (DR) services. We explore the opportunities for a water distribution system (WDS) to provide balancing services with demand response through pump scheduling and evaluate the associated benefits. Using a benchmark network and demand response mechanisms available in the UK, these benefits are assessed in terms of reduced greenhouse gas (GHG) emissions from the grid due to the displacement of more polluting power sources and additional revenues for water utilities. The optimal pump scheduling problem is formulated as a mixed-integer optimisation problem and solved using a branch and bound algorithm. This new formulation finds the optimal level of power capacity to commit to the provision of demand response for a range of reserve energy provision and frequency response schemes offered in the UK. For the first time we show that DR from WDS can offer financial benefits to WDS operators while providing response energy to the grid with less greenhouse gas emissions than competing reserve energy technologies. Using a Monte Carlo simulation based on data from 2014, we demonstrate that the cost of providing the storage energy is less than the financial compensation available for the equivalent energy supply. The GHG emissions from the demand response provision from a WDS are also shown to be smaller than those of contemporary competing technologies such as open cycle gas turbines. The demand response services considered vary in their response time and duration as well as commitment requirements. The financial viability of a demand response service committed continuously is shown to be strongly dependent on the utilisation of the pumps and the electricity tariffs used by water utilities. Through the analysis of range of water demand scenarios and financial incentives using real market data, we demonstrate how a WDS can participate in a demand response scheme and generate financial gains and environmental benefits.

1. Introduction

Electricity storage schemes and grid management methods are becoming ever more important as the landscape of the electricity grid changes to more decentralised renewable production. The intermittent nature of these sources and the unavailability of contemporary technology for storing large quantities of electrical energy efficiently and cost effectively has led to a demand for new energy storage systems and more intelligent electricity demand management [1]. Edmunds et al. [2] highlight significant reductions in greenhouse gas (GHG) emissions from future UK
power grids when storage technologies are implemented, while Lau et al. [3] show considerable GHG emission savings can be achieved through a range of demand response programs. Demand response schemes have also been shown to enable a larger integration of wind power generation [4,5].

In demand response, an electricity consumer reduces or shifts its power consumption when requested to do so in exchange for compensation. For an electricity consumer with an electricity demand that is predictable into a future operational horizon, demand response (DR) is provided by reducing its electricity consumption compared to the predicted consumption [6]. Different DR mechanisms may impose requirements on how long the power reduction must last, how large it has to be, at what rate it must be reduced and within what time frame it must be achieved. A detailed summary of possible mechanisms and their properties is provided by Ma et al. [7]. Despite the potential of active demand management to increase renewables penetration [8], a large share of demand response services are currently provided through backup generators instead of demand shifts by consumers [9].

Moreover, the increased utilisation of renewable energies to power a water distribution system (WDS) has been shown to reduce the GHG emissions considerably [10]. This paper shows for the first time how a WDS can provide reserve energy through demand response by optimising the pump schedules.

We quantify the environmental benefits of demand response compared to alternative reserve energy systems as well as the financial profits that can be generated for the water network operator. In the literature, electricity usage management of a WDS has been considered employing time of use tariffs and maximum demand charges [11,12]. Demand response is a more dynamic and flexible service than maximum demand charges and time of use tariffs are part of the reality for a water utility in the UK, therefore Demand Response services must be provided additionally and not alternatively to these methods. The key challenges we address include the formulation of an optimal scheduling problem with demand response and its solution through model relaxations and state-of-the-art global optimisation tools. Furthermore, we validate our findings through simulation of the optimal operation of the WDS using real data from National Grid in a Monte Carlo simulation. Using this data, we show the potential application as an additional revenue stream that is new to water distribution companies, which could simultaneously provide the grid with more demand side response potential at low GHG emissions and competitive cost.

Frequency response and reserve energy mechanisms available in the UK are used as case studies to evaluate the financial and environmental implications of a WDS participating in DR. The financial and environmental benefits of participation are assessed by comparing the operating cost and GHG emissions when participating in DR to those of operations that minimise only the operating cost in a time of use tariff. To ensure this comparison is valid, we solve both schedules to a sufficiently small certifiable optimality gap; an optimality gap that is smaller than the model uncertainty is chosen [13]. When assessing the ability of a WDS to curtail its electricity usage at request to participate in the demand response market, we separate the hurdles to implementation into system and operational hurdles. The system constraints considered are the available financial rewards, the given electricity price structure and the water network’s pump utilisation rate. These dictate whether a demand response program can be considered financially viable. Examples of operational constraints are ramp rates, pump switching constraints or minimum network pressure constraints. This investigation focusses on the system hurdles using quasi steady state modelling and simplified operating constraints; we assume the operational hurdles can be met with available control and monitoring technologies and design expertise.

2. Demand response

2.1. Service description

In the United Kingdom, National Grid operates the electricity grid, maintaining it as tightly as possible around the desirable frequency of 50 Hz. In case of a significant drop in frequency, as illustrated by Fig. 1, National Grid recognises two mechanisms relevant for this work, frequency response and reserve energy. Within two seconds of an incident that causes the frequency to drop, the frequency response services are brought on-line to stabilise the grid. Reserve energy providers are then brought on-line to enable the fast responding frequency services to be switched off so they could be used again at future events. The services considered here that can provide frequency response (FR) through demand response are the Firm Frequency Response (FFR) and Frequency Control by Demand Management (FCDM). The reserve energy provision service considered here is the Short Term Operational Reserve (STOR), which is brought on-line within 20–30 min [14]. The faster reacting fast reserve service is not considered since the 50 MW minimum power delivery is prohibitively large for a typical WDS.

The first demand response service considered for the WDS case studies is STOR since the technical requirements suggest that it can be implemented in a WDS more readily. A STOR provider offers a steady demand reduction and must deliver the reduction within 4 h after being called and may be required to reduce the demand for up to 2 h. However, the tender records show that the mean call duration in 2013 was 82 min and that National Grid prefers services that can respond within 10–20 min [9]. Since the minimum offered power requirement for STOR participation in the UK is 3 MW, only large WDSs would be able to participate in a STOR scheme directly. However, through an aggregator, a company that aggregates several consumers and bids their capacities to National Grid, a smaller WDSs could participate in these mechanisms by sharing the profits generated with the aggregator. To offer STOR National Grid recognises a range of pathways to suit the wide range of suppliers. The pathways modelled here are based on offering STOR services during both availability windows or just in one, this can be achieved through tendering either a committed or flexible service. The STOR windows and tariff structure is described in further detail in Section 2.2 and in Fig. 4.

The second method for demand response energy provision considered here is the provision of frequency response services through FFR or FCDM. National Grid requires that an FFR provider is able to deliver a minimum of 10 MW response power; smaller users can offer FFR through an aggregator. For the secondary response service, which is considered here, the response must occur within seconds and be maintained for a few minutes. The service may be tendered for any time period, with National Grid preferring tenders that can offer and deliver the service most times. Furthermore, there are requirements detailing the metering and communication systems in place as well as pre-qualification assessments that need to be performed [15,16]. FCDM is a bespoke service arranged through bilateral agreements with National Grid. In general an FCDM provider must provide the demand reduction within 2 s of instruction and deliver for a minimum of 30 min. The minimum demand reduction to be delivered is 3 MW, which may be achieved by aggregating a number of smaller loads at same location. FCDM calls occur only ten to thirty times per annum [17].

For our analysis, the frequency response services FFR and FCDM are approximated by removing the minimum power delivery constraint and requiring the WDS to be able to deliver demand response throughout the day. The event duration for which water must be supplied to customers with reduced pump power is set to 30 min. For the analysis of the financial viability of DR the
difficulties associated with sudden pump switches are neglected as we use quasi steady state models to represent the energy consumption; we consider the pump switching speed a technical issue to be resolved with the hydraulic (transient) modelling of the individual systems and their local control. A range of measures that can be employed to enable the pump switching speeds required for frequency response are discussed in Section 6.

Both FFR and STOR services are tendered at the beginning of the month or season respectively. Potential providers place bids with their cost and availability restrictions. Using historical data the range of energy provision and revenue generated from the reserve energy provision can be calculated. The financial rewards for a successful tender are summarised in Table 1. Firm Frequency Response services can also specify a nomination fee in their tender. However, an analysis of the tenders in 2015 shows many successful tenders do not include such a fee [19]. For modelling simplicity, we assume the WDS does not charge a nomination fee either. Frequency response services provide a small amount of energy – in a short (<30 min) high power (>10 MW) burst – at short notice, while reserve energy services provide the reserve power over a longer time period but do not react as quickly as the frequency response services. This is reflected in the payment structure, where the revenue contribution of the availability and nomination fee payment is much more significant for frequency response services than for reserve energy services [9]. FFR and STOR are tendered in each of the six seasons National Grid considers and pricing varies significantly between seasons. The services are more lucrative for the providers in winter due to tighter margins in the grid [18]. The services are more lucrative for the providers in winter due to tighter margins in the grid [18].

The environmental benefits of demand response from WDS is assessed by comparing the change in GHG emissions due to a change in operating schedule to the operational GHG emissions from competing energy storage mechanisms. The three largest power plants for STOR provision are Open Cycle Gas Turbines (OCGT), pumped hydro and diesel internal combustion engines [20,21].

The GHG emissions attributed to the WDS operation are due to the time dependent energy consumption and the associated GHG emissions of the electricity grid. To define the GHG emissions caused by demand response the operations when providing demand response the operations when the service is available and when the response energy is provided are considered separately. The GHG emissions from either operation are compared to the normal operation when not providing demand response services. The cost of providing demand response are evaluated using the same method. However, as the GHG emissions are not explicitly minimised in the objective function the change in GHG emissions can be both negative or positive while the cost are minimised and thus a deviation from the normal optimal operating schedule is expected to incur a cost for the operation when providing demand response.

The GHG emissions and cost due to the provision of response energy in the case of a demand response event can readily be calculated in terms of the energy provided during an event. To compute the GHG emissions and cost due to the availability to provide demand response, assumptions based about the usage of the service must be made as the GHG emissions and cost are incurred for the power provided and are independent of the actual energy delivered in events and are found from the difference to normal operations. These assumptions are summarised in Table 2. The analysis shows that the frequency response services FFR and FCDM deliver only a small amount of energy, due to the short nature of their responses, which is also reflected in the payment structure. The energy delivered from STOR in a given year varies by how often the provider is called, which is itself a function of the price of the energy offered as National Grid uses a strict merit order system to call response energy [1].

### 3. Optimal pump scheduling

The optimisation of pump schedules for WDS operation is a difficult computational problem as the description of pump states and flow in pipes involve binary variables and some of the underlying fundamental system equations are non-linear. The integer problem can be solved through heuristic optimisation methods such as a genetic algorithm (GA) to optimise pump scheduling, with a separate solver for the non-linear hydraulic simulation problem [23–25]. Other heuristic procedures that have also been applied include simulated annealing [26,27] and particle swarm optimisation [28].

In the mathematical optimisation framework, the scheduling problem can be posed as an mixed integer problem (MIP), solving the hydraulic model and scheduling simultaneously. This can be solved using iterative linear programming for local optimisation methods or using dynamic programming [30,31]. The scheduling problem can also be solved using branch and bound methods [32,33] or model predictive control [12]. A detailed review work on water distribution operation optimisation is provided by DAmbrosio et al. [34]. To compare different operational conditions and the resulting pump schedules, certifiable global optimality is required. This can be achieved through a branch and bound method [35]. A piecewise linear approach that approximates the problem such that it can be solved to global optimality is presented by Morsi et al. [36]. The optimal schedules are calculated using a piecewise linear

---

**Table 1** Summary of the range of monetary rewards for FFR and STOR service options [9,16,18,19]. Rewards for FCDM are based on custom agreements with National Grid [17].

<table>
<thead>
<tr>
<th>Type</th>
<th>FFR</th>
<th>STOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows available:</td>
<td>24 h</td>
<td>1st &amp; 2nd</td>
</tr>
<tr>
<td>Availability £/MW/h</td>
<td>80–150</td>
<td>7–9</td>
</tr>
<tr>
<td>Energy pay £/MW h</td>
<td>75–200</td>
<td></td>
</tr>
<tr>
<td>Total £/MW (R)</td>
<td>45–100 k</td>
<td>30–45 k</td>
</tr>
</tbody>
</table>

* For modelling simplicity we neglect nomination fees.
approximation of the hydraulic constraints based on methods outlined in Menke et al. [13]. By comparing the operating cost and GHG emissions of a WDS participating in a demand response scheme to one under normal operations, we compute the optimal level of demand response capacity to provide, i.e. one that minimises the operating cost.

3.1. Optimisation problem formulation

The optimisation problem for scheduling WDS pumps for DR can be formulated as:

$$\text{minimise : } \text{Pump operation cost} - \text{DR revenue}.$$  
$$\text{subject to : } \text{Hydraulic constraints of pumps and pipes}, \text{ mass balance of the system}, \text{ additional constraints from DR provision}.$$  

where $C_{ij}$ is the penalty for a single pump switch. The value of $C_{ij}$ is based on recommendations by Van Zyl et al. [24]. For further discussion of pump switching constraints and the cost associated with the dynamic response of a hydraulic system, see Section 6.

3.3. Hydraulic balance

The pressure delivered by a pump (i.e. the piezometric head difference across it) can be described by a set of linear constraints that define a convex set approximating the characteristic curve of a pump. For a given time step, a fixed speed pump $i_p$ connecting nodes $J_1$ and $J_2$ is constrained by:

$$h_{J_1} - h_{J_2} \leq \begin{cases} m_{i_p} q_{i_p} + c_{i_p} T_{i_p} & \text{if } T_{i_p} = 1 \\ \Delta h_{\text{hub}}, q_{i_p} = 0 & \text{if } T_{i_p} = 0 \end{cases}$$

where $m_{i_p,1} \cdots m_{i_p,5}$ and $c_{i_p,1} \cdots c_{i_p,5}$ are the linear coefficients for the five hyperplanes describing the convex set. $\Delta h_{\text{hub}}$ is an upper bound on the pressure head generated by the pump. These constraints are enforced using a big-M method as detailed in Menke et al. [13] and Gleixner et al. [32].

Similarly, the energy balance for flows in pipes is modelled using a piecewise linear approximation of the head loss formulae given by either the Hazen–William or the flow dependent Darcy–Weisbach equations [39]. For a given time step, the head loss across pipe $P_{2}$ connecting nodes $J_3$ and $J_4$, for example, can be approximated using a piece of wise linear equations given by:

$$q_{p_{1}} m_{P_{1}} q_{p_{1}} + c_{P_{1}} T_{p_{1}} \text{, if } q_{p_{1}} \leq q_{p_{2}} \leq q_{p_{3}},$$
$$q_{p_{2}} m_{P_{2}} q_{p_{2}} + c_{P_{2}} T_{p_{2}} \text{, if } q_{p_{2}} \leq q_{p_{3}} \leq q_{p_{4}},$$
$$\vdots$$
$$q_{p_{n}} m_{P_{n}} q_{p_{n}} + c_{P_{n}} T_{p_{n}} \text{, if } q_{p_{n}} \leq q_{p_{n}} \leq q_{p_{n}}.$$
For a tank $j$ with flows $q_{in}$ and $q_{out}$, the mass balance for time steps $f = 1 \ldots N - 1$ is given by:

$$q_{in,j} + q_{out,j} = (h_{j,f+1} - h_{j,f}) \times A_j,$$  

where the surface area of the tank is given by $A_j$. Since demand patterns are similar from day to day, we ensure that schedules are repeatable (reasonably similar) by enforcing the constraint that final levels in tanks do not differ much from their initial conditions:

$$\begin{align*}
(h_{j,1} - h_{j,N}) \times A_j & \leq \delta_V, \\
(h_{j,1} - h_{j,N}) \times A_j & \geq \delta_V,
\end{align*}$$

where $\delta_V$ defines the volumetric difference. This relaxes the approach where the final or initial tank levels would be input data, which would limit the feasible search space and potentially lead to a sub-optimal final solution. A similar initial and final constraint relaxation is used in [29], while [40] includes a penalty for final water levels below the initial level or for ones away from a specified target at the end of the operating period.

### 3.5. Demand response cost function and constraints

The revenue from providing demand response is proportional to the power committed to the scheme. This is represented by:

$$f_j(\cdot) := -P_{dr} \times R$$

where $P_{dr}$ is the power committed to DR and $R$ is the expected revenue per MW committed as summarised in Table 1. However, to be able to provide this power as demand response, the WDS must certify that it would always consume this offered power unless it answers a DR call. The power consumed by the pumps is given by:

$$\sum_{j=1}^{m_p} P_{ij} T_{ij} \geq P_{dr}$$

When providing demand response, the WDS must be able to satisfy the expected water demand from consumers for the maximum possible duration of the DR event. This is enforced in the optimisation model by specifying a minimum fill level for tanks, which depends on the maximum DR time offered. This increases the lower bound in the tank dependent on the demand level by:

$$h_{\text{min},i,j} = h_{\text{min},i,j} + \frac{V_{d,j,i}}{A_j}$$

where $h_{\text{min},i,j}$ is the lower bound of the tank level range and $V_{d,j,i}$ is the volume deficit caused in tank $j$ by operating the network without pumps for the maximum DR call duration at time step $t$. $V_{d,j,i}$ is calculated using a hydraulic simulation before the optimisation and ensures the WDS can provide its water demand requirements during the demand response event.

The constraints for demand response provision are only enforced for time steps within the times for which demand response capability should be provided depending on the type of service being FFR where it is continuous or for STOR when it is during the first or both of the availability windows.

### 3.6. Problem summary

The variables of the optimisation problem are the binary variable $T$ for the ON–OFF status of the pumps and the continuous variables $h$ for the hydraulic head at a node and $q$ for the volumetric flow rate in a connection. Although not described here, a set of binary indicator variables are used to enforce the piecewise linear approximations for head loss across pipes [13]. The power committed to demand response is given by $P_{dr}$. The pump schedule optimisation problem for a fixed electricity price tariff giving the normal operation schedule is given by:

$$\begin{align*}
\text{min} : f_1(\cdot) + f_2(\cdot) \\
s.t. : (4), (5), (6), (7)
\end{align*}$$

For the demand response case where the provide level of demand response power is left to be found by the solver the above formulations are modified as:

$$\begin{align*}
\text{min} : f_1(\cdot) + f_2(\cdot) + f_3(\cdot) \\
s.t. : (4), (5), (6), (7), (10), (11)
\end{align*}$$

The mixed integer quadratic program formulated in MATLAB and solved with CPLEX [41]. Since tighter bounds on variables can improve the computational speed of the branch and bound algorithm in CPLEX and we want to reduce the computational effort required to solve the problem to an adequate optimality, the bounds on the variables and the value of the big-M are chosen as tight as possible while including all hydraulically feasible solutions [13].

To demonstrate demand response from water distribution systems, the Van Zyl benchmark network shown in Fig. 2 is used as a case study. It is analysed under a range of pump utilisation rates, a range of overall rewards for the provision of DR and varying cost of energy through a selection of real electricity price tariffs. To enable these comparisons, the water demand is described in terms of the pump capacity of the network. In general, water demand at a node is modelled as the product of a time dependent term, the demand pattern shown in Fig. 3, and a constant term, the base demand $d_b$ [24]. The network model has only one source reservoir and all water demand consumed in the network must flow through the pump station containing pumps $\text{main}_1$ and $\text{main}_2$. We express the pump utilisation of the network as a function of the best efficiency point (BEP) flow rate of one of these identical parallel pumps. To achieve this, the water demand was modified from the version available in Van Zyl et al. [24] by changing the base water demand $d_b$ to the BEP flow rate of $\text{main}_1$ and modifying the pattern such that it had a mean of one. In simulating different levels of consumption compared to the capacity of each pump, the demand rate is modified, to a value $d_k$ and $\frac{d_k}{\text{base}}$ is defined as the pump utilisation rate. A low pump utilisation factor can be interpreted as a large pump supplying a network, while a higher factor indicates several smaller pumps supplying the network.

Larger commercial electricity consumers often utilise electricity tariffs with a range of prices across the day, with high peak prices during peak power consumption times. The electricity tariff used for this analysis is one used by a UK water utility and is shown in Fig. 4. The peak price makes pump operations in this period particularly expensive. The changes in the electricity supply, due to the introduction of renewables, are expected to lead to a change in peak prices. However, the experiences and estimates from Australia and Germany [43,44] show that the direction and scale of change in peak prices can vary significantly. To analyse the effect of peak price on the schedules and the financial viability of demand response services, we scale the tariff in Fig. 4 by altering the minimum and maximum prices over a large range, while maintaining the same mean price for the tariffs to allow a fair comparison between different tariffs. Each electricity tariff used in our optimisation is thus referred to using the ratio of the maximum and minimum prices, $\text{Pricemax} / \text{Pricemin}$. The tariff used by the utility in Fig. 4 has a ratio of 3.2. The time of the availability windows used for STOR services by National Grid vary from year to year and across seasons of the year; Fig. 4 shows the STOR windows that are considered for this analysis highlighted in red, representing an aggregation of the
window times offered by National Grid. In this analysis the availability window descriptions were simplified to the constraints that STOR service providers must be available to provide STOR services in both STOR windows or in just the first STOR window. A DR event can start only inside an availability window but may continue to outside of the window [16]; we model this explicitly. After participating in an event, there is a recovery period in which the STOR provider does not need to be available to provide another response. The likelihood of an event occurring at a specific time in the STOR window is considered uniform since the usage of historic statistics provided by National Grid [22] shows that there is no clear trend.

The likelihood of an event occurring at a specific time in the STOR window is computed following the procedure given below:

1. For a normal day operation (i.e. operation optimised for DR provision), solve (13) to get schedule $T_o$, and record the initial fill levels of the tanks $h_0$.
2. Solve the DR event simulation problem, which has additional constraints as described in (16).
3. The optimal schedule ($T_{event}$) for the operation with the event is computed.
4. Compute the operation cost and GHG emissions of the event and compare to the original cost and GHG emissions.

The number of events to model was chosen large enough to ensure convergence of the mean and standard deviation of the results. For example, Fig. 6 shows the convergence of the standard deviation and mean cost and GHG emissions computed indicating

4. DR events

4.1. Simulating a DR event

The operation during and in the 24 h after the begin of the event is investigated to provide an indication of the cost and GHG emissions that arise due to such an event. Compared to the original problem, which yielded $T^*$, $h^*$ the original pump settings and tank levels, the event optimisation problem is further constrained, by the power consumption requirements of the event and the resulting changes in tank levels and deviation from the originally optimal path resulting from the optimal schedule.

Given the original operating schedule and the required power consumption reduction during an event the operating schedule during the event is computed a priori and together with the initial tank levels further constrain the optimisation problem. The initial and final tank levels are constrained by:

\[
\min : f_1(\cdot) + f_2(\cdot)
\]

s.t.: (4), (5), (6), (7), (10), (11), (14), (15)

Further optimisation methods for a DR event are discussed in Section 6.

4.2. Monte Carlo simulation of demand response events

The cost and GHG emissions from response energy provision depend on a range of factors. The electricity tariff used here is the same for all work days of the year, which are the only days considered for DR. The cost of a DR event, therefore, is dependent only on the duration and start time of the event and not the date. The range of daily GHG emission variations can influence the overall emissions [46]. To ensure this is captured in the estimate of the GHG emissions from events, the date of the simulated events is varied and the corresponding historic GHG emission data used to compute the associated emissions. The traces of the date specific emission levels are shown in Fig. 5.

The cumulative probabilities of STOR event durations are summarised in Table 3. For each 24-h day, the modelled event durations are discretised into 30 min intervals. The starting time of the events are modelled with a uniform probability within the STOR window times offered by National Grid [22]. The number of events to model was chosen large enough to ensure convergence of the mean and standard deviation of the results. For example, Fig. 6 shows the convergence of the standard deviation and mean cost and GHG emissions computed indicating
that the analysis has been performed on a sufficiently large sample [47]. These results of the simulations are discussed in Section 5 and summarised in Table 4.

5. Results and discussion

We investigate three aspects of demand response from WDS, how optimal pump operations change to enable the provision of demand response before and during a DR event, requirements needed for the provision of DR through pump scheduling to be financially viable and the environmental aspects of DR from WDS and how it compares to other alternative response energy provision technologies.

The results and discussion are separated into sections focussing on the financial viability of providing DR from WDS in Section 5.1, the GHG emissions associated with the provision of DR in Section 5.2, and the optimal scheduling for DR provision and DR events in Section 5.3.

5.1. Financial viability

Fig. 7 shows the volume formed by the combinations of price ratios, reward size and pump utilisation rate that are financially viable to provide a FFR or FCDM service from the Van Zyl Network when assuming a maximum event duration of 30 min. It shows that a high pump utilisation rate, high reward and low price ratio benefit the financial viability of the fast response schemes. It also demonstrates that the ratio of the maximum and minimum prices of the electricity pattern as well as the pump utilisation rate have the strongest effect on the financial viability of the DR service. The reward level on the other hand has a lower impact on the financial viability. Demand response through FFR provides the highest amount of yearly revenue for committed power while requiring the least amount of energy provision. The revenue primarily stems from the nomination and availability payments for the power capacity provided and to a lesser extend from the energy provision. Assuming a WDS can fulfil the technical requirements with regards to size and pump switching speeds, it could provide a good opportunity for a profitable committed demand response provision if the pump utilisation rate and the electricity tariffs are moderately high. Otherwise, a bespoke FCDM agreement where the peak hours of the contract are spared out could provide a viable alternative.

The financial viability of STOR services is explored with an annual reward of £25,000/MW, which represents the lower bound estimate of the revenue from availability payments alone, based on approximately 3500 h of availability per year and £7–9/MW availability payments as detailed in Table 1. The optimal power level to commit to be available for STOR provision in both availability windows is shown in Fig. 8. Fig. 9 shows the same for provision in the first availability window only. The two figures show that for lower pump utilisation rates no STOR service is viable as the optimum power is 0 kW. For a small range of pump utilisation rates and electricity peak price ratios, STOR provision from the booster pump with 89 kW is the optimal power for DR while for a large range of pump utilisation rates \( P_1 \) a pump from the main pump station with 178 kW can be committed to STOR provision. With a reward of £25,000/MW, the additional revenue from DR provision, if viable, can be up to 4.6% of the normal operating cost.

The cost of scheduling for STOR relative to normal operations not only depends on the pump utilisation ratio and maximum price of the electricity tariff, but also on which availability windows STOR is offered in. If only the first STOR window is used, a wider range of conditions and a larger capacity of pump power can be committed to the provision of demand response. On the other hand, also providing STOR services in the second STOR window as well reduces the range of financially viable options as the peak electricity price becomes more relevant.

The cost of providing response energy in a STOR event are estimated by modelling a wide range of events in a Monte Carlo simulation, as shown in Fig. 10. Table 4 shows that the cost of energy when rescheduling due to an event reduces the income generated

---

### Table 4

Summary of Monte Carlo simulation results, showing mean and 95th percentile of the cost and carbon intensity of the simulated energy provision.

<table>
<thead>
<tr>
<th>DR service type</th>
<th>Cost (p/kW h) ( \bar{k} )</th>
<th>Cost (p/kW h) ( z_{0.05} )</th>
<th>GHG (gCO₂e/kW h) ( \bar{k} )</th>
<th>GHG (gCO₂e/kW h) ( z_{0.05} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFR</td>
<td>7.4</td>
<td>20.2</td>
<td>137</td>
<td>295</td>
</tr>
<tr>
<td>STOR</td>
<td>4.2</td>
<td>11.1</td>
<td>88</td>
<td>202</td>
</tr>
</tbody>
</table>

---

Fig. 5. GHG emissions for the UK grid for 2014 [45]. The emissions profile of each day is described by a trace with the mean highlighted in red. The emission intensity descriptions are discretised to 48 time steps as \( \text{G} \).

Fig. 6. Variation of the normalised mean and standard deviation of cost and GHG emissions of the Monte Carlo simulation for STOR events, showing their convergence well within 1000 simulated events.

Fig. 7. Reward and pump utilisation required for a given price pattern ratio for permanently committed demand response from FFR or FCDM.
through demand response event on average by 4.2 p/kWh and 95% of the events cost less than 11.1 p/kWh. This represents 25–50% of the income generated from participating in the event. Thus, the cost associated with an event are shown to be well below the payments received for reserve energy provision.

For National Grid, STOR from a WDS would be very attractive since a water utility would not require a minimum guaranteed STOR provision to provide demand response; this is because, unlike for the competing technologies, STOR is only an additional revenue stream and not the sole purpose for a water utility [14].

5.2. Environmental analysis

The GHG emissions for the UK grid in 2014 are shown in Fig. 5 and the carbon intensities of the competing technologies are summarised in Table 5. For the competing technologies providing STOR services, to be displaced by demand response from WDS, we consider the three most important technologies: Open Cycle Gas Turbines (OCGT), Pumped Hydro Storage (PHS) and Internal Combustion Diesel (ICD) engines, which together account for 82% of the STOR market [20]. When comparing the emissions of technologies, only the operational emissions are considered. For an OCGT plant these are provided by Seebregts et al. [48]. The PHS plant is assumed to fill its reservoirs in the early hours of the morning or at night, when average emissions are approximately 400 gCO₂/kW h, which is based on data from [45]. The GHG emissions from operating an internal combustion diesel (ICD) engine and generator unit are computed by considering full load operation emissions only and neglecting emissions other than CO₂ [50]. The carbon intensity range used here is more generous towards the competing technologies than the figures given by National Grid [21], but as performance of power plants varies this range will give a better insight into the potential for GHG emission abatement through their displacement by DR from WDS. The emission ranges found are similar to the range of values identified by Turconi et al. [51].

The carbon intensity of providing the availability for demand response services by a WDS varies significantly. They can increase or decrease the overall emissions as they are not considered in the optimisation of the schedule and vary throughout the day. However, the histogram in Fig. 11 shows, the additional GHG emissions from scheduling for DR are contained to a range of 50–50 gCO₂/kW h.

The GHG emissions caused by the provision of energy from the DR event are summarised in Fig. 12. The mean GHG emissions as modelled in a Monte Carlo simulation are 88.3 CO₂/kW h and 95% of emissions are below 202.1 CO₂/kW h. Table 4 shows that the 95th percentile of the Carbon Intensity of the events are much less than the emissions from competing frequency and reserve energy providers summarised in Table 5. The total GHG emissions associated with the provision of energy provided with less than 240 gCO₂/kW h. This performance is significantly better than the next best conventional alternatives, an efficient OCGT or pumped hydro storage.

The network configurations yielding higher GHG emissions in Fig. 5 are characterised by larger changes to operation schedule to facilitate a higher capacity of DR provision. However, there is no clear trend with regards to pump utilisation rate or electricity peak prices. Green house gas emissions from FFR are not considered here due to the small amount of energy displacement of the mechanism make the comparison to the values quoted in Table 5 misleading. However, it has been shown that frequency response services from PHS can reduce the GHG emissions compared to OCGT plants [3].

Generally the GHG emissions per unit of response energy provided by a WDS linearly depends on the total yearly response energy provided, in order to reduce the carbon intensity a lower
price in the offer tender is suitable while higher prices may improve the profits at the expense of emission reductions.

5.3. Scheduling for demand response

Scheduling for DR divides into two problems: finding an optimal schedule to operate when providing the readiness to reduce energy usage at request and the scheduling during and after a DR event. The scheduling for the readiness requires the guaranteed operation and consumption of the energy tendered as demand response capacity. To schedule during and after a DR event, more operational constraints need to be considered; these include the initial tank fill levels, the minimum tank levels and the desired final tank levels.

To enable a meaningful comparison between schedules generated for a range of operating conditions the use of an optimisation method that can guarantee global optimality is necessary. However, the difficulty in solving the MIPs can lead to convergence issues with the solver not reaching the required level of optimality in the given maximum solve time. In the rare cases when the additional revenue from demand response is smaller than the optimality gap of the solution, this can lead to small variations of the results [13].

Fig. 13 exemplifies the development of the levels of the storage tanks and the change in the 24 h operating schedules after a 90 min DR call at 13:30PM on the 21st January 2014. The figure highlights the resulting behaviour that occurs due to a DR call and the following deviation from the optimal pump schedule for normal operations without an event. The overall pump activity is increased with pumps operating in parallel more often. As a result, the overall energy consumption is increased.

Considering the minimisation of operating cost is only ever the second priority of a water utility, second to the guaranteed provision of water to the customers. This guarantee of supply needs to be also ensured in events and is verified through hydraulic simulations with the schedules. To ensure this feasibility, the approximation of the pump capacity underestimates the flow rate while the approximation of the head loss overestimates the head loss, ensuring the schedule provides sufficient energy to the system and the tank levels stay within the required bounds. For example, the capacity limit shown in red in Fig. 7 is derived from hydraulic simulations; it shows that, in the marginal cases, the approximations lead to an underestimation of the pump capacity in the case of the upper end of the capacity limit.

---

### Table 5

<table>
<thead>
<tr>
<th>Technology</th>
<th>Best</th>
<th>Worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Cycle Gas Turbine (OCGT)</td>
<td>480</td>
<td>575</td>
</tr>
<tr>
<td>Pumped Hydro Storage (PHS)</td>
<td>470</td>
<td>571</td>
</tr>
<tr>
<td>Internal Combustion Diesel (ICD)</td>
<td>520</td>
<td>700</td>
</tr>
</tbody>
</table>

---

1 For interpretation of colour in Figs. 4–9 and 13, the reader is referred to the web version of this article.
6. Conclusion and future work

Through the use of a global optimisation technique we compared the operating schedules of a WDS system minimising the operating cost alone and minimising the operating cost while participating in different demand response schemes in the UK. Through this analysis we show that for a wide range of electricity tariffs and water demands there exist demand response mechanisms which allow the WDS to provide demand response and reduce its cost and provide response energy at low GHG emissions. These new results should encourage water utilities to investigate the potential to schedule for demand response provision from their WDS. To be viable demand response from a WDS there would be no need for a no minimum response duration. Unlike competing STOR and demand response providers, demand response from a WDS does not have shut down (start up) costs that need to be recouped by a minimum STOR operation, which makes potential demand response provision from WDS particularly valuable to National Grid [9].

STOR when tendering only for the first window can be provided at little extra cost to the WDS compared to regular optimal operation, as the highest price tariff times can be excluded from the provision period. Due to the specific requirements this poses, it may be necessary to provide STOR service through an aggregator or combine the STOR provision with another energy asset. When tendering to both windows the maximum price of the electricity tariff, if charged during the operation window, limits the financial viability of STOR provision.

The provision of response energy in an DR event is shown to have limited additional cost and GHG emissions. The operation scheduling during and after the event was performed using the same optimal scheduling techniques used to obtain the global optimal operating schedule. Through optimisation with a receding time horizon considering the uncertainty of future events occurring to allow repeated provision of demand response could further reduce the cost of providing DR through a better schedule.

The environmental impact is dependent on a range of factors, but demand response from WDS can often be provided at very low carbon intensity per unit of response energy provided. The faster responding services – FFR and FCDM – provide small amounts of energy potentially leading to worse carbon intensity of the energy provided, however the custom nature of FCDM may enable the inclusion of such services with small changes in scheduling and thus small changes in GHG emissions. Shorter response events have lower carbon intensities as the originally optimal schedule is only perturbed a little.

Electrical power distribution losses and life cycle emissions were ignored in this analysis since they are similar for the different technologies considered, for which we assume the infrastructure to be already in place. With the additional grid regulation services that come with the introduction of more renewables to the grid, the usage of already built WDS to regulate demand could provide significant reductions in GHG emissions compared to newly built infrastructure.

The WDS was modelled using quasi steady state modelling. For the provision of FFR the power reduction ramping speed constraints need to be analysed. This will require a more detailed analysis of the transient response of the network [52]. Further work could consider the difference in business case of upgraded surge protection devices to enable faster pump ramp rates without causing pressure induced failures in the pipes. Upgraded pump controls or battery systems to enable a gradual shut down may also be considered [53].

Acknowledgements

Ruben Menke is funded by the Grantham Institute for Climate Change and the Environment, and Climate-KIC. Dr. Edo Abraham is supported by the NEC-Imperial “Big Data Technologies for Smart Water Networks” project. The work of the Panos Parpas was partially supported by an FP7 Marie Curie Career Integration Grant (PCCG11-GA-2012-321698 SOC-MP-ES) and EPSRC grants EP/M028240/1, EP/K040723/1.

References


