Holistic modelling techniques for the operational optimisation of multi-vector energy systems

Jonathan Reynolds*, Muhammad Waseem Ahmad, Yacine Rezgui

BRE Centre for Sustainable Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom

A R T I C L E   I N F O

Article history:
Received 11 December 2017
Revised 22 March 2018
Accepted 26 March 2018
Available online 5 April 2018

Keywords:
Energy modelling
Multi-vector energy systems
Power-to-Gas
Building energy modelling
Urban energy systems
Energy management
Optimisation

A B S T R A C T

Modern district energy systems are highly complex with several controllable and uncontrollable variables. To effectively manage a multi-vector district requires a holistic perspective in terms of both modelling and optimisation. Current district optimisation strategies found in the literature often consider very simple models for energy generation and conversion technologies. To improve upon the state of the art, more realistic and accurate models must be produced whilst remaining computationally and mathematically simple enough to calculate within short periods. Therefore, this paper provides a comprehensive review of modelling techniques for common district energy conversion technologies including Power-to-Gas. In addition, dynamic building modelling techniques are reviewed, as buildings must be considered active and flexible participants in a district energy system. In both cases, a specific focus is placed on artificial intelligence-based models suitable for implementation in the real-time operational optimisation of multi-vector systems. Future research directions identified from this review include the need to integrate simplified models of energy conversion units, energy distribution networks, dynamic building models and energy storage into a holistic district optimisation framework. Finally, a future district energy management solution is proposed. It leverages semantic modelling to allow interoperability of heterogeneous data sources to provide added value inferencing from contextually enriched information.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Given that the building sector contributes around 40% of EU greenhouse gas emissions and energy consumption [1], increased focus on improving energy efficiency is vital to meeting national and international obligations. A prominent trend in achieving this is the increased decentralisation of energy infrastructure. This is, in part, enforced by the users who both consume energy and produce it, often using small-scale renewable generation like solar PV panels. Furthermore, decentralising energy production leads to other possible benefits; transmission losses are reduced, and cogeneration or trigeneration units could be utilised. Cogeneration can be achieved using combined heat and power units, CHP, that effectively capture the waste heat from electricity production and supply nearby demand with it. Often, these are facilitated by a district heating system, which also has the benefits of being able to accept various forms of heating energy input such as excess heat from industry, waste incineration, CHP, geothermal or heat pumps [2].

Whilst the use of stochastic renewable resources such as wind power and solar energy are undoubtedly necessary to reduce greenhouse gas emissions; they also introduce a level of uncertainty into the energy supply system. This necessitates a transition from a demand-led energy system to one in which both supply and demand are partially controlled. This new stress on energy grids is one of the driving forces behind research relating to the smart grid. To achieve the full potential of the smart grid, increased data interoperability, better forecasting and better energy management systems are required [3]. Buildings must be seen as an active participant in a district energy system, providing demand flexibility through bi-directional communication with the energy network [4].

Pivotal to the success of a decentralised, district energy system is the ability to manage it holistically. Energy networks such as heat, electricity and gas that were previously controlled independently must now be managed and controlled in an integrated manner as they become more coupled. For example, CHPs produce both electricity and heat often from gas; heat pumps use electricity to produce heat, electricity can be converted to gas, stored for later use or used to generate heat through ‘Power-to-Gas’ technology. Therefore, optimisation of just a single energy network may lead
to an overall sub-optimal result if other networks are not considered. To develop an effective optimisation, accurate, yet simplified internal models of all aspects of a district energy system. To aid in overcoming this challenge, this paper will review modelling approaches of various energy generation and conversion components often found in district energy systems. This paper will also provide an agenda-setting perspective on the eventual goal of integrating them into a holistic district energy model that can be used for operational optimisation of multi-vector energy systems.

The rest of the paper is organised as follows: Section 2 details why improved simplified models of district energy components need to be developed and provides the methodology and contribution of this review. Section 3 reviews modelling techniques from the supply side of a district energy system. Conversely, Section 4 discusses modelling of the demand side of a district. Recommendations for district energy modelling based on the reviewed literature are made in Section 5 alongside a future vision for a district energy management platform and the key components this requires. Finally, Section 6 provides the conclusions.

2. Motivation and methodology

Modelling and optimisation of entire district energy systems have already been attempted in several academic publications and scientific projects. The leading approach in the literature to achieve this is the Energy Hub modelling concept [5]; which simplifies complex urban energy systems to a series of input-output energy hubs. The inputs are in the form of primary energy sources, and the outputs are the produced electricity, heat and/or gas. The ‘Hub’ itself contains the mathematical modelling of the conversion process and technologies (Fig. 1). However, this type of modelling often simplifies energy conversion units to simple constant efficiencies, failing to take into account part load characteristics, warming up periods and other energy losses.

The energy hub concept has been utilised in several papers studying the optimal layout and design of district energy systems [6–10]. This includes selection and sizing of the energy production units and consideration of which energy hubs should be connected. This work is aimed at the design stage or future scenario evaluation and is based on steady-state analysis of known (or assumed) peak demand. Therefore, the assessed temporal scale is years of assumed behaviour rather than day to day optimisation at a sub-hourly resolution.

Operational optimisation of energy hubs can also be found in the literature, often using Model Predictive Control (MPC), [11–13]. In [14], the energy demand was determined from EnergyPlus building simulation models; then the potential, uncontrollable, renewable supply was assumed forecast and finally the energy hub then matched supply and demand in an optimal way using linear programming techniques. A dynamic particle swarm optimisation study was carried out on a Canadian case study in [15], using known hourly, heating, cooling, electricity and transportation loads. Maroufshah et al. [16] also built on the energy hub concept to create a generic smart energy network model for operational optimisation. This paper includes detailed modelling of energy storage, which was included in the energy hub modelling. Game theory has been applied to smart energy hubs in [17,18] for demand response interactions with an electricity and natural gas utility company. The authors’ argued that previously demand response measures could only be aimed at consumers with load flexibility, however, if a multi-source energy hub is available consumers can participate by changing the source of their electricity supply.

Considering a network of energy hubs is shown in the literature to be an effective way of optimising energy management at a district level. However, all of the studies in this section made a number of simplifications. The buildings are often simplified models or using design stage assumptions rather than using accurately calibrated building energy models. The building energy demand is also assumed perfectly forecasted and inflexible with no consideration of demand-side management or demand response measures. The efficiency of the energy conversion units is often oversimplified. They assume a thermal and/or electrical efficiency to be constant and therefore does not include part load factors and warming up characteristics, which are vital for a realistic day-ahead optimisation [19].
There is potential to improve upon the energy hub model with more detailed component models in place of a conversion matrix with static efficiencies. These models could be mathematically derived or could use machine learning or artificial intelligence. Machine learning techniques are able to capture and predict complex non-linear relationships in several fields including energy management and buildings [20]. These methods require large amounts of training data from which an algorithm can learn outcomes from past trends and, once trained, can use that knowledge to produce predictions for future scenarios [21]. Deep learning is a specific branch of machine learning that extends existing techniques to achieve a greater understanding of the data. These models are comprised of a more complex architecture and contrasting functions. Deep learning has been successfully applied to fields such as image classification, pattern recognition and speech recognition [22,23]. Regardless of the modelling techniques applied, it is essential that the resulting model remains mathematically and computationally simple enough for near real-time optimisation.

2.1. Previous reviews

State of the art reviews exist for the broader topic of district energy systems and district energy modelling. Allegrini et al. [24], reviews software modelling tools for district energy systems. While the topics reviewed are similar to this review (energy generation technology and multi-vector district energy systems) the scope of the review is focussed on physics-based software tools. Indeed, one of the outcomes of the review is the need for more low-level models with reduced complexity which is the main focus of our review. Keirstead et al. [25], reviewed urban energy systems from several aspects including urban climate, building design and transport. This review takes a broader view of urban energy systems with the long-term aim of an entirely integrated, smart city, model encompassing all the described aspects. However, component level energy generation or demand modelling is not included. A key identified future challenge is the access to and integration of vast amounts of data from several sources leveraging cloud computing advances. A potential approach to achieving this vision is discussed in Section 5 of this review.

A review of district energy modelling for energy planning optimisation has been carried out in [26]. The authors reviewed modelling and optimisation techniques at different district scales to result in the optimal selection and layout of energy generation technology. In contrast, our review focuses on modelling for use in real-time operational optimisation applied to existing district energy systems. Connolly et al. [27] gives a detailed review of 37 available computational simulation software packages. It studies each tool, gives the advantages and disadvantages, and discussed what aspects of a district energy system they can effectively model. A review by Baños et al. [28], discusses optimisation techniques applied to district energy systems. While this covers much of the same area as our review, the paper looks at the applicability of different optimisation methods applied to district energy systems rather than the modelling techniques those optimisations require.

Furthermore, existing, detailed, reviews for several subsections of this review can be found in the literature, these will be referenced and acknowledged where relevant. These reviews tend to be particularly in-depth and cover several aspects of modelling that are not required or suitable for modelling for operational control. This review aims to provide a more holistic review for researchers and practitioners that require a general understanding of modelling techniques for each component of a district energy system. The authors would encourage readers with a specific interest in modelling one component to also consult the specific reviews where applicable.

2.2. Scope, limitations and contribution

The purpose of the review paper is to provide an overview of existing modelling techniques of components within urban and district energy systems including the emerging technology of power to gas. In contrast to existing work, this review intends to provide a wider, holistic, summary of modelling a district energy system specifically for operational optimisation. It will attempt to gather academics’ and practitioners’ attention towards currently available methods, along with their performance, usefulness and limitations for online or near real-time optimisation applications. It will discuss not only the well-known physics-based modelling software but also include newer computational intelligence and machine learning techniques for modelling individual components as well as deep learning approaches. Indeed, this review will vote an increased focus on these approaches as they are likely to be more suitable for real-time, operational optimisation but are largely neglected in current urban energy modelling reviews. Furthermore, demand-side energy modelling will be reviewed, as forecasting future energy consumption is essential for any advanced district energy management strategy. This review does not cover urban design and planning tools as the focus is more on component modelling techniques to allow more optimal control of an existing district system.

To achieve this stated aim, the body of literature was queried through searching well established and respected databases of academic publications such as IEEE Xplore, ScienceDirect, Scopus and Google Scholar. The extensive list of sources was further filtered with weighting given to rigorously peer-reviewed studies and based on their impact in the wider research field. A specific emphasis was placed on simplified modelling techniques with hourly to sub-hourly granularity where applicable and available.

3. Component modelling

This review intends to take a bottom-up approach to district energy modelling by focussing on modelling techniques developed for each specific component commonly found within district energy systems. This section will review modelling techniques found in the literature applied to combined heat and power (CHP), Boilers, Solar thermal and photovoltaic systems, Wind Generation, Power-to-Gas and Heat Pumps.

3.1. Combined heat and power

Combined heat and power, CHP, is becoming a favoured technology during the transition from a fossil fuel energy infrastructure to a low carbon future. While they still frequently use fossil fuels, namely natural gas, they can achieve greatly improved efficiencies. This is as a result of utilising the heat by-product from electricity generation in a local heating system and thus also reducing transmission losses. Total efficiencies of around 80–90% have been achieved as opposed to the 30–40% figure achieved in traditional, large-scale, fossil fuel electrical power plants [29]. There is a range of CHP types based mainly on the type of prime mover, typical examples include internal combustion, fuel cell and Stirling engine. Furthermore, during summer the heat produced by the CHP can be used to drive cooling cycles forming trigeneration cycles (heating, cooling, and electricity). The main three cooling technologies driven by heat are absorption, adsorption and ejector cycles. An ejector cooling cycle, in particular, was modelled in [29], based on the heat from a CHP.

Best et al. [30], developed a district energy modelling tool with a modular design. In particular, the authors focused on the mathematical modelling of CHPs and chillers. The CHP model used manufacturers rated capacity and adjusts this for altitude, outdoor tem-
perature, and part load ratio using statistical regression equations. The resulting model allowed the fuel consumption, cost, and CO2 emissions to be calculated based on the energy demand. Wang et al. [31], aimed to optimise the operation of several CHP units and thermal energy storage for a district heating network. Their CHP model was based on a convex, feasible operating region based on characteristic points. However, for 2 of the 3 CHPs included in the case study, they only had two characteristic points at maximum and minimum operation. The authors included ramp rate constraints, which were modelled as a percentage the CHP output can increase or decrease from one hour to the next. Maintenance periods were also considered in this optimisation problem.

Detailed thermodynamic modelling of micro-CHP, residential scale devices has been developed as part of an IEA project in [32]. A grey box approach to modelling sub-components of 4 types of CHP has been taken. The model reflected partial physical processes but also required empirical constants to be determined based on the measurements obtained from real units. Each sub-component within the device was modelled as a separate control volume to which fundamental conservation laws can be applied. These models have been integrated into four different modelling platforms, namely ESP-r, TRNSYS, EnergyPlus and IEA-ICE. Validation of these models was provided in [33], which showed excellent agreement between simulation and measurement of a Solid Oxide Fuel Cell (SOFC) CHP. Average errors of 1.2%, 8.3%, and 5.4% were reported for electrical, thermal and total efficiencies. For more information on the detail of the modelling techniques see [34] for internal combustion engine and Stirling engine CHP’s and [35] for information on solid oxide fuel cell CHP’s.

Savola and Keppo [36] aimed to generate multiple linear regression models to calculate the power production of several CHP at part loads. While CHP power output at high loads is almost linear, as the part load decreases the power decreases non-linearly due to a rapid decrease in turbine isentropic efficiency. Therefore, this work proposed multiple linear regression models depending on the part load factor of the CHP. These can be described mathematically using the following equation:

\[ P(Q, T_h, T_c) = a \cdot Q + b \cdot T_h + c \cdot T_c + d \]  

Where \( P \) is power production (W), \( Q \) is the part load factor (-), \( T_h \) is the outgoing fluid temperature (°C), \( T_c \) is the incoming fluid temperature (°C) and \( a, b, c \) and \( d \) are regression coefficients. Using three separate regression lines for different sections of the part load curve was shown to be accurate versus a simulation model and yet remains a linear equation simple enough to be included in optimisation strategies.

An analytical approach to assess the characteristics of a cogeneration gas turbine unit was carried out in [37]. Using this approach, curves relating several parameter ratios (such as thermal efficiency over design thermal efficiency) could be related to the part load ratio. This work amongst others, is used in [38] to create best-fit curves to calculate part load thermal efficiency and part load fuel consumption as a function of the part load percentage. These curves were compared to experimental data of three gas turbine CHPs and showed excellent consistency. The equation for this curve is given in (2):

\[ \eta_{th,PL} = -0.00000634(PL)^2 + 0.0137(PL) + 0.262 \]  

Where \( \eta_{th, PL} \) is the part load thermal efficiency, \( \eta_{th, Nom} \) is the nominal thermal efficiency, and PL is the part load percentage where all variables are dimensionless.

For wider district energy optimisation, the authors believe that multiple linear regression equations or non-linear regression curves are best suited for real-time operational optimisation and management. They provide an accurate representation of the behaviour of a CHP while requiring minimal computational effort to calculate due to their relative simplicity. This approach provides more realistic modelling than the constant efficiencies used in the state-of-the-art energy hub formulations.

3.2. Boilers

Typically, district heating plant rooms are comprised of multiple energy conversion technologies. Due to the decrease in efficiency in part load conditions and fluctuating electricity demand, CHPs are often sized to provide the baseload and operate continuously where possible. Additional heating load flexibility will be provided by more traditional boilers, which can more ably modulate their output based on instantaneous demand. Typically, these boilers will have very high thermal efficiencies and have a wider operating range than the more inflexible CHPs. The most commonly found fuel source for district-level boilers is natural gas however biomass is becoming increasingly popular due to governmental policy schemes.

A thermodynamically derived, mathematical model of a steam boiler was presented in [39]. The model included factors for various sources of energy loss such as heat losses to the environment through each component and combustion losses. This allowed each source of energy loss to be analysed and potentially reduced. From the mathematical model, a part load efficiency curve was produced consisting of three distinct zones. From 0–40% load, a hyperbolic relationship between load and efficiency existed, from 40 to 80% there was a near linear relationship and above 80% resulted in near constant efficiency. The model was verified through comparison with experimental measurements. A similar method of model development was applied to domestic condensing boilers in [40]. The resulting model calculated outlet water and gas temperatures and thermal efficiency based on the inlet temperatures, flow rates and static boiler parameters. Petrocelli and Lezai [41] analytically modelled a wood pellet boiler and analysed the effect of storage tank size and control strategy on the boiler emissions. The authors found that increasing the size of the storage tank decreased emissions due to less frequent startup and shut down times.

A numerical Computational Fluid Dynamics (CFD), software, ANSYS Fluent, was used to provide a more complete analysis of boiler behaviour in [42]. The verified model allowed analysis of flow conditions and flame behaviour as well as NOx output. As a result, NOx reduction strategies could be trialled before implementation. However, this level of detail does come at the cost of computational complexity as the model contains 6.8 million meshing cells and significant computational time. Similar CFD analysis of a biomass boiler was carried out in [43]. This study combined a 1D model of the fuel bed to provide inputs to a full 3D CFD simulation of the whole boiler.

A simplified grey box model was derived in [44]. The authors aimed to make a generic boiler model consisting of three phases; the combustion chamber, heat exchanger and thermal storage. Where possible empirical relationships were used to ensure the resulting model required as few input parameters as possible, most of which can be found on standard boiler specification sheets. A generic boiler simulation model was also developed in [45]. Several different combustibles including oil, gas, pellet and wood chips were modelled and several flue gas temperature modelling techniques were used. The model was developed to be integrated into the TRNSYS simulation platform and claims a thermal efficiency prediction accuracy of ± 1%.

A combined, hybrid model for determining the behaviour of a large coal-fired, steam boiler can be found in [46]. A neural network was used to provide a simple calculation of flue gas temperature which was an input for an analytical model to calculate the thermal efficiency. The resulting model was therefore computationally simple enough to be used for real-time control applica-
tions. A simplified, non-linear, 3rd order state space model of a biomass boiler was used in [47] for a model based control application. The model-based control contributed a significant reduction in CO and particulate emissions and resulted in an improved thermal efficiency.

This section has shown several detailed, numerical modelling studies of the behaviour of boilers under various conditions and using various combustibles. However, for the purposes of real-time, operational control of a district, this level of computational complexity and simulation time is infeasible and unnecessary. Many of the modelling procedures described in this section are based on specific types of boilers. Therefore, in the authors’ opinion, appropriate modelling of a boiler in a district configuration can be achieved through experimentally finding the empirical relationship between fuel input or part load factor and the heat power output similar to that found in Section 3.1. Efforts should be made to account for start-up and shut-down periods which can display distinct behaviour and are likely to effect real-time optimisation strategies. Table 1 summarises the reviewed literature related to the modelling of CHP and Boilers.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Method</th>
<th>Input parameters</th>
<th>Output parameters</th>
<th>Model accuracy</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>[30]</td>
<td>Regression</td>
<td>Ambient temp, altitude, part load</td>
<td>Efficiency</td>
<td>-</td>
<td>CHP</td>
</tr>
<tr>
<td>[31]</td>
<td>Convex operating regions</td>
<td>Operating point</td>
<td>Power, heat and cost output</td>
<td>-</td>
<td>CHP</td>
</tr>
<tr>
<td>[32,33]</td>
<td>Grey box modelling</td>
<td>Empirical coefficients, operating strategy</td>
<td>Electrical, thermal and overall efficiency</td>
<td>1.2%, 8.5%, 5.3% Average error</td>
<td>CHP</td>
</tr>
<tr>
<td>[34]</td>
<td>Grey box modelling</td>
<td>Empirical coefficients, control signal</td>
<td>Fuel flow rate, electrical output, heat recovery rate, outlet temp</td>
<td>$R^2 = 0.991, 0.991, 0.991$</td>
<td>CHP</td>
</tr>
<tr>
<td>[36]</td>
<td>Multiple linear regression</td>
<td>Part load, output temp</td>
<td>Power production</td>
<td>&lt;0.01 Squared error</td>
<td>CHP</td>
</tr>
<tr>
<td>[38]</td>
<td>Regression Thermodynamic principles</td>
<td>Part load</td>
<td>Relative efficiency</td>
<td>-</td>
<td>CHP</td>
</tr>
<tr>
<td>[39]</td>
<td>Thermodynamic principles</td>
<td>Boiler static data, operating strategy</td>
<td>Thermal efficiency</td>
<td>0.35% Mean error</td>
<td>Gas boiler</td>
</tr>
<tr>
<td>[40]</td>
<td>Thermodynamic principles</td>
<td>Boiler static data, ambient conditions, fuel and water mass flow rate, water temp</td>
<td>Outlet water and gas temp, heat output, efficiency</td>
<td>0.2-2.5% Relative error (Efficiency)</td>
<td>Gas boiler</td>
</tr>
<tr>
<td>[42]</td>
<td>CFD</td>
<td>Geometry, boundary conditions, operating strategy</td>
<td>NOx Emissions, boiler temps, flow velocities</td>
<td>&lt;16% (NOx)</td>
<td>Biogas Boiler</td>
</tr>
<tr>
<td>[44]</td>
<td>Grey box modelling</td>
<td>Empirical coefficients, static manufacturer data, operating strategy</td>
<td>Hot water supply temp, flue gas temp</td>
<td>&lt;1% to &lt;8% Relative error</td>
<td>Gas boiler</td>
</tr>
<tr>
<td>[46]</td>
<td>ANN + Analytical model</td>
<td>Feed water temp, oxygen content, thermal power, heat flux to preheater, air temp, fuel lower heat value</td>
<td>Blending gas temp</td>
<td>$R^2 = 95%$</td>
<td>Coal-Fired Steam Boiler</td>
</tr>
<tr>
<td>[47]</td>
<td>State space model</td>
<td>Biomass flow rate, primary air mass flow, sum of primary and secondary air mass</td>
<td>Residual oxygen content, feed temp</td>
<td>-</td>
<td>Biomass boiler</td>
</tr>
</tbody>
</table>

3.3. Solar energy

Power systems’ operation and planning is being performed according to the Smart-grid (SG) vision [48]. With more renewable technologies being integrated into existing and new energy supply infrastructure, especially the non-predictable ones (wind and solar), it would be challenging to maintain balance between supply and demand. A continuous balance always needs to be maintained between supply and demand at any moment by continuously controlling demand and adjusting energy generation [49]. The stochastic nature of solar energy generation introduces exigent issues for the optimal operation and planning of SG. Predictive analytics will play a significant role towards optimal real-time management, secure operation and maintaining a balance between energy supply and demand. Solar energy generation is dependent on several factors such as orientation, shading, cloud cover, air temperature and solar irradiation. Therefore, prediction of solar energy output is often dependent on the prediction or measurement of these parameters. Whilst the field of solar energy systems is expanding to include building integrated solar systems this review will only consider the most common and developed solar energy technologies namely photovoltaic panels and solar thermal collectors.

3.3.1. Photovoltaics (PV)

The textbook approach to calculating the electrical power generated by a solar cell is defined as:

$$P = I \cdot \eta \cdot A \quad (3)$$

Where $P$ is the power produced (W), $I$ is the total solar radiation on the PV surface (W/m²), $\eta$ is the total system efficiency (-), and $A$ is the area of the PV panel (m²). However, making this calculation is dependent on knowledge of potentially difficult to obtain parameters such as solar radiation, shading, ambient temperature and solar cell efficiency which may not be constant. Durisch et al. [50], emphasised the need for more detailed information than that provided by a manufacturer datasheet at standard test conditions. It empirically modelled PV efficiency as a function of solar cell temperature, global irradiation and relative air mass. From ambient temperature and global radiation forecasting the cell temperature was determined through an empirical relationship. Then the cell efficiency was calculated using a further empirical relationship and hence cell power output could be produced using Eq. (3). The authors argued that their PV efficiency model could aid planners when selecting the type of PV cell to deploy in different regions based on typical ambient temperature and global irradiance. However, they did not foresee the model being used for short term power prediction. The developed model has been further validated in both [51,52], where the model was adapted and applied to real test sites in Algeria and Bulgaria respectively to assess the performance under different operating conditions. Additional development and refinement of the Durisch model was conducted in [53] by including wind speed as an input. This produced an alternate method of calculating PV cell temperature, as a function of ambient temperature, global irradiance and wind speed, which then impacted the resulting estimate of cell efficiency. A more simplified model was produced in [54] which does not require a large number of input parameters. However, due to its simplified nature, the model outputted the daily energy performance of a PV solar cell which is not suitable for use in operational control.

PV panels can also be modelled using a simple electrical circuit composed of a current generator wired in parallel with one or several diodes and resistors. Ma et al. [55] reviewed the various
Fig. 2. Solar cell equivalent circuits. Source: Ma et al. [55].

configurations found in the literature. Modelling an ideal solar PV cell consists of just a single diode although this lacks accuracy due to its simplicity. Introducing the additional resistors and diodes shown in Fig. 2 increases the accuracy of the PV model but also increases the complexity and hence computation time. The most commonly used model is the 5-parameter model with 1 diode and 2 resistors as shown in Fig. 2. However, this requires calibration procedures to determine the 5 parameters. Examples of procedures to determine the 5 parameters can be found in [56–58] along with validation of the models against measured performance. The modelling of PV arrays under partial shading was presented in [59]. The model’s inputs are the PV panel’s characteristics (maximum power, current, and voltage at the maximum power point, short circuit current, open circuit voltage) the shading patterns, solar insolation level, number of modules, working temperature and number of blocking diodes. The output of the simulation was the I-V characteristic and the maximum power point for each group of the PV panel. Despite the high accuracy of these models they still require weather parameters to be measured or predicted as inputs which can be difficult in practice.

Whilst solar cell equivalent circuits are the most common approach to modelling solar PV power output, advances in artificial intelligence and machine learning are beginning to emerge as contenders. A rural PV-Diesel hybrid system was modelled and optimised using neural networks in [60]. An ANN was developed to predict solar radiation based on more commonly available weather data. This was then used as an input to another ANN to predict the power output from a PV array. Using this information, optimal dispatch of solar power and diesel generator operation could be found. Kharb et al. [61] uses an ANFIS model to improve the efficiency of a solar panel by maximum power point tracking, MPPT. They use temperature and irradiance as inputs and from this predict the MPP which allowed the controller to react quickly to changing environmental conditions.

As equation (3) demonstrates, solar irradiance is directly proportional to the power output of a PV cell. Therefore, prediction of solar irradiance and solar power output are almost one and the same. Three different types of ANN model were trialled in [62] to forecast ground level solar insulation and ambient temperature which were then used to calculate PV panel power output. The models were trained using the previous 16 days meteorological data. The inputs to the model included the previous 24 h insolation, temperature and atmospheric insulation as well as forecast atmospheric insulation and relative humidity. There was a small difference between the three types of ANN, each using different learning algorithms, and this was likely to be influenced by the ANN parameter values. The mean absolute percentage error comparing the model output and actual values was around 15–20% throughout the year which translated to a similar accuracy in predicting the power output. Similarly, Mellit and Pavan [63], developed and ANN-based, 24 h ahead, solar irradiance prediction method. Inputs to the model included mean irradiance value, air temperature and day of the month and very good prediction accuracy was achieved, particularly on sunny days. Day-ahead solar irradiance predictions were then used to calculate predicted solar power output, and this was compared to a real facility in Italy. An $R^2$ value of 0.9 and a mean absolute error of less than 5% was achieved. Deep learning techniques were applied to model the power generation of 21 different solar farms in Germany in [64], Techniques trialled include Long Short-Term Memory (LSTM), Deep Belief Network (DBN) and Auto-Encoder LSTM. These were compared to a physical modelling approach as well as a ‘shallow’ Multi-Layer Perceptron (MLP) model. It was shown that whilst all machine learning models significantly outperformed the physical model, the deep learning methods only provided a small improvement over the MLP.

3.3.2. Solar thermal

Whilst PV technology uses solar energy to generate electricity; solar thermal collectors aim to convert the same solar energy into useful heat often in combination with a hot water storage tank. Theoretical solutions and standards for calculating the efficiency and useful heat energy conversion of solar thermal collectors are widely available and were well explained in [65]. The analytical modelling of solar thermal collectors has been adapted to be included in building simulation platforms such as EnergyPlus and TRNSYS. However, this requires knowledge of several solar collector parameters in addition to many weather variables such as the solar irradiance, wind speed and ambient temperature. Therefore, like the case of solar PV, simplified models are required for wider scale, real-time, energy optimisation.

Several thermodynamically derived, mathematical modelling studies of solar thermal collectors can be found in the literature.
These tend to develop models for improvements or alterations to the standard flat plate solar collector. For instance, Dowson et al. [66] developed a model for a polymer air collector with an aerogel insulation layer. A model to calculate the efficiency, output temperature and component temperature of a novel counter flow v-groove solar collector can be found in [67]. Luo et al. [68] modelled the effect of using nanofluids to improve the system efficiencies of a solar thermal collector. Electrical circuit analogies can also be used for the modelling of solar thermal collectors as demonstrated in [69,70]. Electrical circuit models simplify the mathematics of modelling solar thermal systems but still retain some knowledge of the physical components.

When sufficient amounts of data are available, it is possible to model solar thermal collectors with an ANN, similar to the case of solar PV. For instance, the performance of a solar thermal system has been modelled using both ANFIS and ANN in [71] with comparable results. The model showed a mean relative error of 1% when predicting the stratification temperature, and 9% for the solar fraction. The results show a high level of accuracy and reliability using artificial intelligence methods, with a significant reduction in complexity compared to a full mathematical description of the system. However, the amount of data required (panel’s characteristic, orientation, tilt, and solar radiation every minute) can be difficult to collect in practice. Kalogirou et al. [72], also used an ANN to predict the output characteristics of a large-scale solar thermal system. It predicted the energy output and the storage tank temperature with accuracies of $R^2 > 0.95$. However, this study focussed on the total daily energy output rather than the finer timescales required for operational optimisation.

### 3.3.3. Discussion

This section has shown several mathematical and machine learning methods for predicting solar energy output, the reviewed literature has been summarised in Table 2. In the case of solar PV, the more simplified analytical models based on empirical relationships or equivalent electrical circuits may be suitable for use in operational control and optimisation due to their short calculation time. The analytical approaches used for solar thermal modelling are too complex for use in real-time optimisation. Accurate predictions of solar PV or solar thermal output will undoubtedly require relevant weather variables as inputs. Therefore, to predict future solar energy generation, accurate weather forecasts are required. In many cases, sufficiently accurate forecasts of variables on an appropriate temporal scale such as ambient outdoor temperature and relative humidity will be available from national meteorological services. The forecasting of global solar radiation has a higher associated uncertainty and is less commonly available publicly. Therefore, many of the machine learning methods reviewed in this section first aimed to predict solar irradiance and from that calculate the solar power output, offering a computationally efficient and simple approach. However, the common downsides associated with machine learning prediction models also apply for solar energy modelling. These include the requirement for a large amount of historical or simulated data and the inflexibility of the model to adapt to any changes made to the system. Furthermore, machine learning approaches can be susceptible to problems of overfitting. This occurs during the training process if the model fits too well to the training data set without learning the general trends. Then when applied to an unseen testing data set, the model performs poorly. Depending on the machine learning approach, different methods exit to prevent this. These include ‘pruning’ the trained model to remove any unnecessary links or stopping training early based on the performance of a validation dataset. Note that these drawbacks associated with machine learning are true of every application rather than just the reviewed studies presented here.

### 3.4. Wind power

Wind power generation relies on wind speed, which could be influenced by obstacle, terrain and height. Wind power generation is stochastic in nature, and therefore the reliability of wind
power generation is not satisfactory as it cannot produce and supply steady electricity to the electrical grid. The wind power penetration influences the power system operation. To tackle this challenge, the power system operators/decision makers must make a detailed schedule plan and set a reserve capacity for it [73]. Wind power may not frequently be considered a small-scale urban energy source as wind farms are often built on a large scale and in more remote locations. However, it is feasible that a wind farm may be first directly connected to an urban microgrid rather than the wider national grid. Also, given that wind power is one of the largest renewable generation sources currently deployed the authors believe that prediction of this power generation is worthy of discussion in this review. Two recent reviews [49,73], state that there are three broad methods for calculating wind speed or wind power generation. These include physical-based, white box, numerical models, more traditional statistical models such as ARIMA, and newer artificial intelligence-based models such as ANN, fuzzy logic and Support Vector Machine, SVM.

Typically, the power generated by a wind turbine can be defined as a function of wind speed. However, a wind turbine will have four operational zones which should be defined by the manufacturer of the turbine. Initially, at low wind speeds the turbine will remain stationary and produce no power until a cut in speed is reached. Then in the second zone, the output power is a cubic function of wind speed (shown in (4)) until the rated wind speed and power is reached. Where \( P \) is the generated power (W), \( C_p \) is the dimensionless power coefficient of the turbine, \( \rho \) is the density of air (kg/m\(^3\)), \( A \) is the swept area of the turbine (m\(^2\)) and \( U \) is the wind velocity (m/s).

\[
P = \frac{1}{2} C_p \rho A U^3 \tag{4}
\]

In the third zone, the power output will remain constant at the rated power regardless of wind speed. Finally, if the wind speed becomes too high, the turbine will shut down to prevent damaging loads. A typical wind power - wind speed curve is shown in Fig. 3. Therefore, the challenge of predicting wind speed and wind power are almost one and the same. However, errors in wind speed forecasting are exacerbated by the cubic relation between wind speed and power.

While the wind-power curve is typically provided by manufacturers, this relationship does not factor in the specific context of each site (e.g. turbulence) or the condition of the turbine (e.g. deterioration and wear) or the proximity to additional turbines [74]. A common method found within the literature aims to develop site specific wind-power curves to achieve greater accuracy. Jin and Tian [75], proposed a probabilistic method to model wind power generation by adding a term to Eq. (4) to reflect the stochasticity of the wind speed and power variation between wind turbines in the same wind farm. Lydia et al. [76], applied a range of techniques to generate a more accurate wind-power curve applied to 5 different datasets. These techniques included parametric modelling such as a linearized segmented model, four and five parameter logistic expressions as well as non-parametric modelling including neural networks, fuzzy clustering and data mining approaches. For the sake of brevity only the results from the best model (5-parameter logistic function) and for dataset 1 are included in Table 3. Wind power curve techniques may be necessary to understand more realistic site-specific conditions; however, the resulting curve still requires forecast wind speed as an input to predict power generation. Given that both recent reviews, [49,73], state that for short-term prediction (hourly to sub-hourly) artificial intelligence based models are most effective, the rest of this section will focus on this area.

Five different machine learning techniques were applied to the prediction of future wind speed and wind power generation in [77]. They considered predictions using different time steps and prediction horizons. For very short-term wind speed and power predictions, they found SVM models outperformed other data mining techniques. This used the previous hours’ time series data to predict up to an hour ahead in 10-min intervals. The authors also considered a slightly longer timeframe for predicting wind power up to 4 h ahead using the previous 4 h, mean power generation data. Multi-layer perceptron, MLP, was the most accurate method for this timeframe prediction. An ANN was used in [78] to make short-term forecasts of wind speed at a wind farm site in Mexico. The ANN was trained based on time series data and used the previous hours values of wind speed to predict the next hour. A method combining wavelet transformation and neural networks to predict short-term wind power generation at a national level in Portugal was developed in [79]. Adding the wavelet transformation to get a better representation of the input data provided an increase in accuracy compared to using an ANN alone in all four seasons.

Quan et al. [80] aimed to address the calculation of prediction uncertainties. They produced an ANN that outputted the lower and upper bound of electrical load and wind power generation rather than a specific prediction value. A Particle Swarm Optimisation (PSO) procedure was used to minimise the width between these bounds under the constraint of 90% prediction coverage. The proposed procedure provided a significant improvement over more traditional methods although the width between the bounds for wind power generation remained high due to the randomness and intermittent nature of wind power. Similarly, Men et al. [81] developed an ensemble mixture density neural network method to make a probabilistic forecast of wind speed and power. It provided not only a prediction but also confidence bounds for the predicted time series. It was found to outperform several other prediction methods regarding prediction accuracy and quality of the confidence bounds. An ensemble approach combined with wavelet transformation and a deep learning, Convolutional Neural Network (CNN) was proposed in [82]. The model required only recorded, time-series values of wind power as an input, from which it predicted wind power from 15 min to 8 h ahead. The proposed methodology was compared to a back-propagation and SVM approach and was shown to outperform these models in every test. Welch et al. [83], developed three neural networks using different methods to predict short-term wind speeds. The authors found that recurrent neural networks outperformed the multi-layer perceptron architecture. An alternative, Naïve Bayes decision tree prediction model is used in [84]. It aims to extract relationships between wind speed and additional weather data. Support Vector Machine (SVM) prediction models have been compared to ANN in [85] to predict mean daily wind speed. They find that the SVM model compares favourably against the ANN.
Table 3
Wind modelling literature summary.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Method</th>
<th>Input parameters</th>
<th>Output parameters</th>
<th>Model accuracy</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>[76]</td>
<td>Parametric modelling</td>
<td>Wind speed</td>
<td>Wind power</td>
<td>RMSE = 0.6408, MBE = 0.4874</td>
<td>Various</td>
</tr>
<tr>
<td>[77]</td>
<td>SVM, MLP, Decision tree</td>
<td>Previous wind speeds</td>
<td>Wind speed and power</td>
<td>Relative error = 15% and 23%</td>
<td></td>
</tr>
<tr>
<td>[78]</td>
<td>ANN</td>
<td>Previous wind speeds</td>
<td>Wind speed</td>
<td>MSE = 0.0016, MAE=0.0099</td>
<td>Mexico</td>
</tr>
<tr>
<td>[81]</td>
<td>Ensemble MDN</td>
<td>Forecast wind speed</td>
<td>Wind speed and power</td>
<td>RMSE = 1.9688 and 174.38</td>
<td>Taiwan</td>
</tr>
<tr>
<td>[82]</td>
<td>CNN</td>
<td>Previous wind power</td>
<td>Wind power</td>
<td>CRPS = 0.281–4.339</td>
<td>China</td>
</tr>
<tr>
<td>[83]</td>
<td>MLP, RNN, SRN</td>
<td>Current wind speed, air temp, humidity</td>
<td>Wind speed</td>
<td>Relative error: MLP=0.5038, RNN=0.4354, SRN=0.4544</td>
<td>USA</td>
</tr>
<tr>
<td>[84]</td>
<td>Decision trees</td>
<td>Time, atmospheric pressure, sea-level pressure, temp, humidity, wind speed and direction, insolation of previous hours wind speed</td>
<td>Wind speed</td>
<td>Classification error rate = 17.54-22.61</td>
<td>Japan</td>
</tr>
<tr>
<td>[85]</td>
<td>SVM and ANN</td>
<td>Previous hours wind speed</td>
<td>Wind speed</td>
<td>MSE=0.0090 (ANN), MSE=0.0078</td>
<td>Saudi Arabia</td>
</tr>
</tbody>
</table>

Note - MLP (Multi-Layer Perceptron), RNN (Recurrent Neural Network), SRN (Simultaneous Recurrent Neural Network), MDN (Mixture Density Network), MSE (Mean Squared Error), MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), CNN (Convolutional Neural Network), CRPS (Continuous Ranking Probability Score)

In summary, from the assessed literature, the authors agree that machine learning methods have the potential to provide the simplest and most accurate short-term prediction (up to 24 h ahead) of wind power generation. However, in comparison to the other generation technologies considered in this review, wind power generation forecasting appears to be the most difficult. This is due to the almost complete randomness in the wind speed profile. In comparison to solar energy prediction, which is also weather dependent, the daily, or seasonal patterns are very limited. This is reflected in the wide uncertainties reported from the reviewed literature. A summary of the literature reviewed in this section is found in Table 3.

3.5. Power-to-Gas

The use of Power-to-Gas (P2G) (hydrogen or methane) technology is a relatively new concept for national energy systems. Due to plans for large expansions in stochastic renewable power generation, a technology is required to be able to effectively store or convert excess electricity at times when it cannot be dispatched. The power to gas technology can convert excess electricity into hydrogen, and subsequently, methane for later use. These gases could be integrated with other sectors such as the chemical industry or transportation if hydrogen powered vehicles have significant take-up. Alternatively, methane (or synthetic natural gas) could be directly injected into the existing gas network with some researchers also suggesting that pure hydrogen could be injected to the same network up to a defined threshold with minimal negative consequences. If appropriate economic and technological conditions prevail, P2G could become a significant technology in the context of multi-vector energy systems as they have consequences for electricity, gas and heat as shown in Fig. 4.

Both [86,87] provide a technical overview of the systems and economic analysis. Initially hydrogen is produced using water electrolysis requiring electricity as an input using one of three current methods; alkaline water electrolysis, proton exchange membrane electrolysis or high-temperature water electrolysis. Then a methanation stage converts the hydrogen to methane requiring a carbon source which could come from carbon capture at fossil fuel power plants, anaerobic digestion of biomass, or from the air. Whilst the technology is still largely at a pilot testing stage there is some concern at the high capital costs and relatively low conversion efficiencies of the technology.

Several national-level investigations into the economic feasibility of P2G have been carried out. Studies by [88,89] modelled the integration of hydrogen electrolysers and P2G at a national level based on UK gas and electricity networks. For a future scenario with high wind power generation capacity, the authors found that allowing hydrogen to be directly injected into the gas network could reduce costs and emissions due to the greater capture of wind resource. A similar national scale, energy storage study in a Dutch context was considered in [90]. A comparison of pumped hydro, compressed air, and power to gas energy storage was provided with varying capacity and different scenarios of wind power production. The study finds P2G to be the least cost-effective energy storage option due to relatively low cycle efficiencies. A future German scenario with 85% renewable energy was studied in [91]. This work aimed to consider the optimal amount of P2G capacity to deploy but also where to deploy it. In this scenario, P2G could lead to significant cost reductions, increased renewable share, and a reduction in CO2 emissions. Guandalini et al. [92], analysed the effect of adding hydrogen electrolysers and gas turbines to large wind farms to provide balancing services. Including these units allowed a more ‘aggressive’ declaration of production to the transmission system operator as inaccurate predictions could be mitigated. An economic analysis of the use of P2G was applied in a German context in [93]. This work found that for the current and near future energy landscape, P2G is not a profitable method of providing balancing services to the national grid. This is due to high capital costs and relatively low gas prices in relation to electricity prices.

All previously discussed studies model the electrolysers or power to gas systems as a constant efficiency and were interested in long-term economic effects over a large geographic scale. Thermodynamic analysis of electrolysers and power to gas plants was conducted in [94,95]. These studies assessed the energy demand for producing hydrogen at different pressures using different electrolysis pathways. However, these models were highly complex and would be problematic to integrate into real-time, operational, district optimisation. Despite their aims to account for thermodynamic irreversibility, these models have yet to be validated against real experimental data. Due to the fact that P2G technology is relatively new and still in an R&D phase, operational data is not widely available. This means that short-term, simplified, modelling of part load efficiencies is not covered in the state of the art literature and represents a significant research gap.

3.6. Heat pumps

Heat pumps have long been identified as a future clean energy source for meeting building heat demand providing they can utilise renewable electricity. They can be categorised as ground source or air source heat pumps and have the advantage that they can also provide cooling in warmer seasons. They have high energy efficiencies with a typical coefficient of performances (COP) of around 3–4, meaning for one unit of electrical energy input you get 3–4 units
of useful heating energy. Studies that consider heat pumps using the typical energy hub modelling procedure, outlined in Section 2, would model this COP as constant when in fact it is dependent on a number of factors including the part load percentage, outdoor air temperature, and ground temperature. Therefore, more realistic models must be developed to allow true optimal control of heat pumps within a multi-vector district energy system.

Several modelling approaches can be found in the literature. A thermodynamically derived, a dimensionless number relating borehole wall temperature to heat gain per unit length can be calculated. Commercial, numerical, heat transfer software can be used to model heat pumps with great accuracy. Artificial Neural Networks, ANN, have also been utilised as well as state space models [96]. Of these approaches, only ANN and state space models are simple enough to be utilised for real-time operational control, and thus only studies using these methods will be discussed in this section.

An Adaptive Neuro-Fuzzy Inference System, ANFIS, approach was used to calculate the COP of a ground source heat pump [97]. Compressor inlet and outlet temperature, as well as the ground temperature were used as inputs to the model. A number of different membership functions were trialled and the best of which achieved an accuracy with a maximum error of 0.25%. Gang and Wang [98] used an ANN to predict the output water temperature of a ground heat exchanger which allowed better control of a hybrid ground source heat pump with a cooling tower. An ANN was used in [99] to predict heating capacity and compressor work done (and hence calculated COP) of a direct expansion geothermal heat pump. Inputs to the model were the temperature and pressure of the evaporator at the inlet and outlet, condenser inlet cooling water temperature, and the discharge pressure. A formal method of varying heat pump parameter set points was utilised to allow generation of a complete training data set in a relatively short period.

Zhang et al. [100] used a Radial Basis Function Neural Network, RBFNN, to model the performance of a ground source heat pump. The model was then used in conjunction with a particle swarm optimisation, PSO, to minimise operational energy consumption of the heat pump given a known building demand. ANN and ANFIS models were compared in [101] for calculating the COP of a ground source heat pump. The inputs to the two types of model were the same; namely, the evaporator inlet and outlet temperature, condenser inlet and outlet temperature, and the load side inlet and outlet temperature. Good accuracy between experimental results and model predicted COP were reported with slightly better results from the ANFIS model. However, these models only allowed retrospective COP calculation as the temperature inputs needed to be measured first meaning this cannot be used for model predictive control applications.

Both a nonlinear autoregressive exogenous, NARX, model and a reduced order state space model were used in [102] for prediction of mean ground loop fluid temperature. These were then utilised in a dynamic programming optimisation and nonlinear MPC optimisation respectively. Both models achieved excellent prediction and allowed calculation of heat pump COP to minimise the cost of energy consumption for a hybrid ground source heat pump system. Ahmad et al. [103] and Ahmad [104] used a quadratic equation to model COP of a heat pump. The developed model was then used to develop nonlinear model predictive control for a solar thermal system combined with a heat pump. In [105], heat transfer and power of a heat pump was modelled using quadratic regression curves based on simulated data. Similarly, models of the pump, fan coil units, piping network, heat storage and building space temperature were created. Whilst several heat pump variables were accurately

**Fig. 4.** Schematic overview of the energy vector pathways of Power-to-Gas.
predicted the authors did not envisage the potential to use this model for a building set point temperature optimisation aiming to minimise the energy consumption from the heat pump.

In summary, simplified models for calculating heat pump parameters do exist within the literature. These are most commonly based on neural networks, state space models or regression curves. However, many of the examples discussed use very specific parameters as inputs that would not necessarily be metered or easily forecasted for the next 24 h. In an ideal case, for a holistic district energy model, the COP would be calculated based on the predicted energy demand, forecasted weather conditions and heat network temperatures. A summary of the literature reviewed in this section can be found in Table 4.

4. Dynamic building modelling

Buildings need to be considered as integral and active parts of an urban energy system and therefore need to be modelled accurately. Building loads (heating and cooling, hot water and electricity consumption) depend on a number of different factors e.g., weather conditions (solar radiation, dry-bulb air temperature, wind speed), thermal properties of building’s fabric, occupants’ behaviour, installed energy system, operational schedules, etc. These interdependencies increase the complexity of the problem, and therefore accurate prediction of building energy consumption can be a challenging task. However, several different building modelling techniques currently exist with different advantages and disadvantages. These modelling techniques can broadly be categorised as white box, grey box, or black box models [106].

4.1. White box

White box or Engineering methods are based on using physical principles to calculate thermal dynamics and energy behaviour of a building or system [107]. Engineering models can be divided into the following categories; detailed methods and simplified methods [107]. Simplified methods can include degree-day, bin methods, etc. and are steady-state models. These methods are predominately useful when the building energy consumption is more dependent on the building fabric. Detailed methods (e.g. TRNSYS, DOE-2, EnergyPlus) often enable users to evaluate design with reduced uncertainties, because of their multi-domain modelling capabilities [108]. Detailed simulation models can produce accurate results; however, they require an extensive amount of building and environmental data for modelling a building and its systems. Modern research efforts are targeting the use of 3D laser scanning and photogrammetry techniques to quickly realise an accurate as-built representation of building geometry on a district scale [109,110]. However, digitisation and subsequent generation of energy models remains a time-consuming task requiring significant manual intervention [111].

Furthermore, these initial building energy models do not tend to perform well in predicting energy consumption of occupied buildings as compared to the design stage prediction [21]. Extensive calibration efforts are often required during the operational phase to adjust the model to reflect reality. This requires widespread metering, categorised spatially and by end use at small time intervals. However, once a calibrated energy model has been completed it can output an exhaustive range of variables from building level total electricity consumption down to the air flow rate of a single zone. Detailed simulation models tend to be more computationally expensive and therefore, are generally considered not suitable for near real-time optimisation problems.

Once a basic energy model has been constructed using the known geometry, construction materials, energy systems and basic rule-of-thumb internal gains estimates; significant efforts are required to calibrate a model. While no agreed upon, universal, methodology has been achieved there are a number of literature reviews on the subject [112–114] and a number of proposed methodologies [115,116]. However, many of these methods are still manual, iterative and time consuming. They often involve identifying the most sensitive parameters that impact on energy consumption using probabilistic analysis such as a Monte Carlo simulation [117]. From this the modeller can allocate most effort to iteratively correcting these parameters [116]. Many of these methodologies aim to estimate a level of uncertainty associated with the resulting building model also [118]. A recent step has been made through the development of “Autotune” for Energy Plus models [119]. This method uses an evolutionary algorithm to tune selected important variables aiming to minimise the error between the Energy Plus output and measured data. However, given the number of “tunable” parameters in a typical building and given that a population-based optimisation method is used; this leads to a very large number of evaluations and hence simulations. To address this, the study uses several high-performance computing techniques and supercomputers, which make this method inaccessible to ordinary practitioners. The resulting calibrated model, when applied to a complex building, achieved an accuracy of CV/RMSE = 11.82% and MBE = −1.27%, equivalent to a manual calibration.
A calibration methodology was implemented in [120] applied to two simulated building and one actual building. Influential modelling parameters were first identified with best guess estimates inputted. This was followed by a course and fine grid Monte Carlo simulation to refine and improve calibration solutions. The resulting calibrated model achieved a CV(RMSE) value of 6–8% when comparing simulated vs actual monthly electricity consumption. Monetti et al. [121], used a particle swarm optimisation, PSO, to calibrate several parameters of an EnergyPlus building. The authors considered infiltration, equipment power, ground temperature, material properties and thicknesses as variables. Once calibrated, a CV(RMSE) of 0.19–0.40% was reported for hourly heating energy consumption comparison of several zones. A two-stage, building energy modelling procedure was carried out in [122]. The initial stage involved detailed inspection of as-built building documentation and surveys of internal loads. The second stage required a more thorough interrogation of key BMS data and occupant surveys. The completed model complied with ASHRAE Guideline 14 accuracy limits for modelling of heat pump electrical demands, heat pump thermal output, building electrical consumption, natural gas consumption, and indoor zone temperature.

White box simulations have been used in the literature as an evaluation engine for optimisation procedures. In [123], an EnergyPlus simulation model was used to optimise pre-cooling operation to minimise energy cost whilst ensuring temperature bounds were met. To automatically link EnergyPlus, other modelling tools, and optimisation procedures in environments such as MATLAB, the Building Controls Virtual Test Bed, BCVTB [124], can be used. EnergyPlus was used in conjunction with a Genetic Algorithm, GA, in [108]. The optimal management of window openings, window blinds and mechanical ventilation was considered with the objective of meeting the occupants thermal, visual and indoor air quality needs whilst minimising cost. Each individual solution in the GA was run in EnergyPlus allowing its’ fitness to be evaluated. However, both of these examples have to use very simplified building models to keep the simulation time within reasonable limits. To complete a metaheuristic optimisation requiring a large number of evaluations of a realistic, complex, building model is not feasible for real-time optimisation, where results are needed in the order of 15-min intervals.

4.2. Grey box

Grey box models are hybrid models; they use simplified physical descriptions to model building and/or building energy systems. The coefficients of the models are identified based on the operational data using parameter identification methods. A simple example of this type of models is the RC-model; in which an electrical analogy is used to model heat transfer through a wall. This method simplifies the problem through a linearization of the equation and hence reduces the computational time [125]. These models are mostly used as a good compromise between modelling accuracy and computational time.

A methodology to develop the simplest, yet suitably accurate, RC model for a single storey case study building in Denmark was explained in [126]. It aimed to model the indoor temperature as a function of solar irradiance and heating input. The final model achieved errors less than ± 0.1°C but from a district optimisation perspective, prediction of heat consumption as a function of set point temperature and weather would be more useful. Ahmad et al. [103] and Ahmad [104], developed an RC model for a two-room building. The model was used to output energy consumption of the building. The authors developed an MPC controller to save energy consumption while maintaining thermal comfort. Similarly, Berthou et al. [127] tested four different configurations of RC models each increasing in complexity. The authors found the 6 resistors, 2 capacitance model to be the best compromise between accuracy and complexity. TRNSYS data was used to tune the RC model parameters which used occupancy, ventilation, temperature set point and solar gain as inputs to predict indoor temperature and heating and cooling demand with resulting fit values of 88% and 89% respectively. Zhou et al. [128], developed not only a building load prediction model but also weather modules to provide the inputs to the building load RC model, hence developing an online, day-ahead, prediction service. Grey dynamic models were used to predict outdoor temperature and relative humidity which were then used to forecast the solar radiation. The predicted solar radiation was then used as an input to forecast building cooling demand with an eventual R² value of 0.91–0.93. However, the number of testing days included was quite limited, and weather forecasting errors had an impact on the eventual energy demand prediction.

Reynders et al. [129], derived several RC models to emulate a more complex, white box Modelica model. First to fifth order RC models were tested along with different training data sets, the addition of noisy data, and using alternative, more easily measured, inputs. The study found that using solar irradiance on vertical planes could effectively take the place of solar gain data and building electrical demand could be used as a proxy for internal gains data. However, the resulting grey box model is only validated against a white box model of a generic Belgian house rather than a real case study. A toolbox design for the streamlining and semi-automation of the development of RC models for model predictive control is outlined in [130]. The software aids the data handling, model selection and parameter estimation, however, achieved poor validation results in one case study due to inappropriate training data. A dynamic, thermal RC model was integrated with an existing stochastic, Markov-Chain, electrical demand and occupancy model in [131]. Building demand, hot water cylinder, gas boiler and heating control models are all integrated and receive active occupancy profiles based on a UK building use survey. However, this study was aimed at producing generalised, aggregated, probabilistic thermal demand of several building rather than specifically for real-time optimisation like the other studies in this section.

Afram and Janabi-Sharifi [132], developed a detailed grey box model of a residential HVAC system comprised of subcomponent models for an Energy Recovery Ventilator (ERV), Air Handling Unit (AHU), buffer tank, radiant floor panels, and a Ground Source Heat Pump (GSHP) based on energy balance equations. Once the model parameters had been identified, only zone and buffer tank set points as well as outdoor air temperature were required as inputs. The authors argue such a model would be prime for use in conjunction with MPC. An example where RC models were effectively applied is provided in a case study based on a Czech university building in [133,134]. The MPC strategy took weather and occupancy as inputs and aimed to minimise the energy cost whilst ensuring thermal comfort by controlling the set point temperature of the water supplied to the building. The RC model would output the predicted building temperature based on the given inputs.

4.3. Black box

Black box models are input-output models based purely on data with no representation of the underlying physical characteristics of a system. These can include purely statistical based regression models, Artificial Neural Networks (ANN), Neural Network Auto-Regressive model with exogenous inputs (NNARX), Support Vector Machine (SVM) or Random Forest models. Black box models have been used extensively in the literature to predict or calculate a wide range of variables key to building optimisation and control such as electricity demand, heating demand, indoor temperature, and predicted mean vote (PMV – a measure of thermal comfort).
Summaries of these types of computational intelligence techniques can be found in [20,135]. The above methods rely on a training period that uses extensive amounts of data. This means that historical data needs to be logged for an extended period or simulation models need to be used to produce substantial amounts of realistic data.

Much of the literature based on creating ANN to accurately predict building data emphasises the need to ensure the most appropriate inputs are used as well as the optimal architecture and internal function are selected. Ferreira and Ruano [136], uses a GA to find the optimal architecture of an ANN to predict the climate of a greenhouse, the resulting model can then be used for optimisation processes. A complete example of selecting functions between each layer can be found in [137]. The resulting model could predict electricity consumption, thermal energy consumption and PMV in a sports facility. From this the HVAC system could be optimised using a model predictive control technique. PMV is normally a complicated parameter to calculate requiring seven (often difficult to measure) variables to be used as inputs to Fanger’s equation. Both [138,139] produce ANN based solutions to calculate PMV without the need to solve Fanger’s equation.

Bagnasco et al. [140], uses an ANN to forecast the electricity demand of a hospital in Turin. Considered inputs include the day of the week, time of day, loads at the same timestep from the previous day and from seven days ago, outdoor temperature, and whether or not it is a weekday. Similarly, [141], forecasts day ahead electricity consumption at 15-min intervals using an ANN. It only considers five input variables, day type, time of day, operational condition, outdoor temperature and outdoor relative humidity but achieves very good prediction accuracy with CVRMSE in the order of 8–10%. A regression-based, data analysis approach was used in [142] to find a correlation between weather and occupancy variables to three electrical load types (appliance, ventilation, and cooling). They found that work hours, occupancy and outdoor temperature were the most important variables in calculating the electrical loads and using fewer predictor inputs resulted in lower errors. The use of ANN and Random Forest algorithms was compared in [21] for the prediction of HVAC electrical consumption of a hotel in Madrid. Considered inputs included weather variables, date and time variables, the number of guests and the number of rooms booked. The ANN was shown to marginally outperform the Random Forest model however the authors argued that Random Forest based methods are easier to tune. A comprehensive and systematic review of electrical load forecasting in buildings, [143], concluded that black box models such as ANN or SVM are well suited to the task.

ARX and NNARX models were compared for their suitability to model indoor temperature in [144]. The model aimed to predict the indoor temperature of a building using previous indoor temperature, outdoor temperature, solar radiation and heating power as inputs. The NNARX model significantly outperformed the linear ARX model and once pruned using the optimal brain surgery algorithm achieved an SSE of 0.906. Royer et al. [145], used a second order state space model to predict the indoor temperature of zones also using outdoor temperature, solar radiation and HVAC operation as inputs. The model proved itself to be adaptable to different buildings but achieved poor results in colder climates.

In most cases, the purpose of the previously described models is to be used as an evaluation engine in optimisation strategies and utilised in conjunction with MPC. For example, in [146], ANN are used to predict the outdoor and indoor temperature for the next 8 h. These prediction models were used by a GA-fuzzy optimisation to control the fan coil operation reducing energy consumption by 35.8%. Similarly, Lee et al. [147], used a neural network based model to predict zone temperature and power consumption. This was used as part of an optimisation strategy that controlled the set point temperature, generation devices and deployment of storage. A multi-objective GA was used to simultaneously minimise energy consumption and predicted percentage dissatisfied (PPD - A measure of thermal comfort) in [148]. A combination of GA and ANN were applied to the same case study building in [149]. In this case a zone level optimisation approach was applied to reduce energy consumption from sporadically occupied zones. This required independent ANN to model the heating energy consumption and indoor temperature of each zone.

Deep learning techniques have been more widely applied to building energy consumption than in the other topics of this review paper. Deep learning methods are commonly based on extensions of a simpler ANN and are well suited to complex tasks such as image processing. Both [150,151], applied deep learning methods to the same dataset. The trialled methods included Long Short-Term Memory (LSTM), Conditional Restricted Boltzmann Machines (CRBM) and Factored Conditional Restricted Boltzmann Machines (FCRBM) with the aim of forecasting residential electricity consumption over varying time horizons. In most scenarios the deep learning models were able to outperform more traditional machine learning models. Fan et al. [152], tested different feature extraction methods combined with several modelling techniques ranging from multiple linear regression to machine learning techniques to Deep Neural Networks (DNN) to predict building cooling energy consumption. They found that application of a deep learning unsupervised feature extraction technique could improve model performance compared to more traditional methods. However, in this case study, it was concluded that a truly ‘deep’ model was not optimal, and the cooling load was best predicted by an Extreme Gradient Boosting (XGB) model. DNN were also used in [153] for forecast the electricity consumption of 40 commercial buildings in South Korea. The DNN were shown to consistently outperform shallow neural networks and a double seasonal Holt-Winters (DSHW) model across different building use categories.

4.4. Discussion

A summary of the reviewed literature can be found in Table 5. The authors of this paper are in agreement with previous reviews that detailed, white box simulation models are not suitable for sub-hourly real-time optimisation. The computational time is too great to be used as an evaluation engine, and they require an expert to create and then calibrate the model using vast amounts of static and dynamic building data. Both grey box and black box building models have been proven to be effective in the reviewed literature for modelling a wide range of building variables. For use in conjunction with district optimisation, it is assumed that building demand prediction and indoor temperature or thermal comfort would be the most useful model outputs. From this, the simplified building models could be used as an evaluation engine in the optimisation algorithm testing the building response to chosen control signals.

5. Discussion and future research directions

This paper has reviewed the broad topic of energy modelling for district energy systems. Due to the interdependencies and connectivity between previously distinct energy vectors, a more holistic energy management strategy and modelling approach must be provided. Several approaches can be found within the literature; however, conversion technologies are often modelled simplistically. They often assume constant conversion efficiencies and no warm up or cool down periods which could lead to overall infeasible or sub-optimal solutions. Therefore, this paper has reviewed modelling approaches for common energy generation and conversion
Table 5
Building modelling literature summary.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Method</th>
<th>Input parameters</th>
<th>Output parameters</th>
<th>Model accuracy</th>
<th>Real case study</th>
</tr>
</thead>
<tbody>
<tr>
<td>[119]</td>
<td>EnergyPlus autotune</td>
<td>Building geometry, material properties, occupancy, lighting, equipment and HVAC gains and schedules</td>
<td>Electricity consumption</td>
<td>CV = 11.82%, MBE = 1.27%</td>
<td>No</td>
</tr>
<tr>
<td>[120]</td>
<td>EnergyPlus manual tuning</td>
<td>Building geometry, material properties, occupancy, lighting, equipment and HVAC gains and schedules</td>
<td>Electrical and gas consumption</td>
<td>GOF = 5–7%</td>
<td>Yes</td>
</tr>
<tr>
<td>[121]</td>
<td>EnergyPlus manual tuning</td>
<td>Material properties, lighting, equipment and HVAC gains, and ventilation</td>
<td>Heating consumption</td>
<td>CV = 0.19–20.4%, MBE = 0.14–0.83%</td>
<td>Yes</td>
</tr>
<tr>
<td>[122]</td>
<td>EnergyPlus manual tuning</td>
<td>Building geometry, material properties, occupancy, lighting, equipment and HVAC gains and schedules, occupancy survey</td>
<td>HP electricity consumption, HP heat supply, indoor temp, electrical and gas consumption</td>
<td>CV = 7.3–25.1%, CV = 18.2–33.5%, CV = 12.4–28.7%, CV = 6.3–16.5%, CV = 4.5–14.1%</td>
<td>Yes</td>
</tr>
<tr>
<td>[126]</td>
<td>RC grey box</td>
<td>Solar irradiance, heating input, static parameters</td>
<td>Indoor temperature</td>
<td>Absolute error &lt; ± 0.1 °C</td>
<td>Yes</td>
</tr>
<tr>
<td>[128]</td>
<td>RC Grey Box</td>
<td>Outdoor temp, humidity, solar radiation, internal gains</td>
<td>Building load</td>
<td>R² = 0.91–0.93</td>
<td>Yes</td>
</tr>
<tr>
<td>[129]</td>
<td>RC grey box</td>
<td>Solar irradiance, electrical consumption, outdoor temp, set point temp</td>
<td>Heating demand, indoor temp</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>[130]</td>
<td>RC grey box</td>
<td>Outdoor temp, horizontal solar irradiance, heating load</td>
<td>Indoor temp</td>
<td>RMSE = 0.33K</td>
<td>Yes</td>
</tr>
<tr>
<td>[132]</td>
<td>RC grey box</td>
<td>Buffer tank and zone set point, outdoor temp</td>
<td>Ventilation &amp; AHU outlet temp, Tank &amp; Zone temp, radiant floor and gshp return temp</td>
<td>R² = 0.996–1.000, CV = 0.010–0.069</td>
<td>Yes</td>
</tr>
<tr>
<td>[137]</td>
<td>ANN</td>
<td>Minute, hour, day, month, occupancy, humidity, pool temp, indoor temp, outdoor temp, air temp &amp; humidity</td>
<td>Electrical and thermal energy consumption, and PMV</td>
<td>MSE = 0.0015%</td>
<td>Yes</td>
</tr>
<tr>
<td>[138]</td>
<td>ANN</td>
<td>Air temp, web bulb temp, globe temp, air velocity, clothing, activity</td>
<td>PMV</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>[139]</td>
<td>RBF ANN</td>
<td>Air temp, relative humidity, globe temp Previous days consumption, previous week consumption, day type, timestamp, outdoor temp</td>
<td>PMV</td>
<td>Absolute error &lt; ± 0.0075</td>
<td>Yes</td>
</tr>
<tr>
<td>[140]</td>
<td>ANN</td>
<td>Operational condition, time, day, outdoor temp and humidity</td>
<td>Electricity consumption</td>
<td>CV = 7.97–11.06%</td>
<td>Yes</td>
</tr>
<tr>
<td>[141]</td>
<td>Regression</td>
<td>Outdoor temp, day light, work hours, radiation, occupancy</td>
<td>Appliance load, ventilation load, cooling load</td>
<td>RMSE = 71.1–13%</td>
<td>Yes</td>
</tr>
<tr>
<td>[144]</td>
<td>ARX and NARX</td>
<td>Outdoor temp, solar radiation, heating input</td>
<td>Indoor temperature</td>
<td>SSE = 0.9060 (NARX), SSE = 0.9579 (ARX)</td>
<td>Yes</td>
</tr>
<tr>
<td>[145]</td>
<td>State Space Model</td>
<td>Outdoor temp, solar radiation, HVAC Operation</td>
<td>Indoor temperature</td>
<td>Fit = 92–84%</td>
<td>No</td>
</tr>
<tr>
<td>[146]</td>
<td>ANN</td>
<td>Outdoor temp, time, HVAC operation, convective transfer of windows</td>
<td>Indoor temperature</td>
<td>R² = 0.97, RMSE = 1.11K</td>
<td>Yes</td>
</tr>
<tr>
<td>[147]</td>
<td>NARX</td>
<td>Day of the week, time, outdoor temp, set point temp, AHU supply temp, AHU flow rate</td>
<td>Indoor temperature and HVAC consumption</td>
<td>CV = 0.007868, CV = 0.114</td>
<td>No</td>
</tr>
<tr>
<td>[148]</td>
<td>ANN</td>
<td>Outdoor temp, solar irradiance, humidity, hour, set point temp, previous indoor temp</td>
<td>Energy consumption, PPD, indoor temperature</td>
<td>R² = 0.9888, 0.9982, 0.9985</td>
<td>No</td>
</tr>
<tr>
<td>[152]</td>
<td>SVR, XGB, DNN</td>
<td>Time, Date, Outdoor temperature, Relative humidity</td>
<td>Cooling energy consumption</td>
<td>CV (RMSE); SVR = 19.0%, XGB = 17.85, DNN = 20.9%</td>
<td>Yes</td>
</tr>
<tr>
<td>[153]</td>
<td>DNN</td>
<td>Outdoor temperature, humidity, solar radiation, cloud cover, wind speed, date and time, previous consumption</td>
<td>Electricity consumption</td>
<td>Average MAPE = 8.85</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note - CV (Coefficient of Variation), MBE (Mean Bias Error), GOF (Goodness of Fit), HP (Heat Pump), RC (Resistor Capacitance), RMSE (Root Mean Squared Error), RBF (Radial Basis Function), PMV (Predicted Mean Vote), GSHP (Ground Source Heat Pump), MAPE (Mean Absolute Percentage Error), NARX (Nonlinear Autoregressive Network with Exogenous Inputs), SSE (Sum of Squared Error), SVR (Support Vector Regression), XGB (Extreme Gradient Boosting), DNN (Deep Neural Network).

technologies including CHP, boilers, solar PV, solar thermal, wind power, Power-to-Gas, and heat pumps.

The scope of this review was to determine suitable modelling for use in real-time optimisation and therefore with short computational periods. For CHPs and boilers, this can be achieved using relatively simple polynomial regression curves relating the part load factor to the efficiency, or through using multiple linear regression equations. This either requires manufacturer data or a small amount of experimental data. Solar energy prediction (both PV and thermal) is highly dependent on the prediction of solar irradiance. Currently, leading methods in the literature use machine learning models to forecast this variable. Then either a further machine learning model or solar equivalent circuits can be used to calculate PV output. In the case of solar thermal, machine learning models are recommended. However, as is often the case with machine learning models, a significant amount of historical data is required.

Short-term, wind power forecasting remains a significant challenge within the literature. This is due to the inherent stochasticity in wind speed and the lack of a consistent daily profile in comparison to solar power. The modelling of P2G systems is relatively unexplored within the current body of literature due to their status as an emerging technology still in an R&D phase. Therefore, no recommendation can be made on the suitability of different modelling approaches. It is expected that when operational data becomes available, linear or polynomial regression curves relating
expected gas output to electricity input will be appropriate. Heat pumps are generally modelled by a COP or seasonal performance ratio; however, this is far from constant in reality. Many factors including part load, outdoor air temperature, and ground temperature can influence the conversion efficiency of a heat pump. From the reviewed literature, machine learning methods such as ANFIS or ANN could prove useful in modelling this behaviour.

A section on the modelling of building energy consumption is also included in this review. Many leading district energy optimisation studies fail to consider the demand-side of the district, often assuming a perfectly predicted, inflexible load. However, buildings must be considered an active participant in a district energy system, providing crucial flexibility. Building energy modelling can be placed into three broad categories, white box, grey box or black box models. From the reviewed literature, the authors do not recommend the use of white box models due to their computational complexity and input data requirements. Both grey box and black box models have been proven in the literature to be able to effectively model building energy demand or indoor temperatures as a function of weather conditions, occupancy and HVAC operation. From a district optimisation point of view, parameters such as the set point temperature of a zone or building could be utilised as a decision variable to allow flexibility at certain peak points.

Of the energy conversion and generation technologies reviewed, modelling techniques are well developed and accurate results have been reported. However, from the reviewed studies it appears that wind speed and power forecasting with short timesteps is the most difficult due to the erratic nature of wind. Therefore, improvements to current methods or development of new prediction methods are required. Furthermore, research remains to be completed on the simplified modelling of power to gas systems. Whilst this is a relatively young technology, it is vital that operational data can be made available to allow researchers to understand the inner relationships between unit inputs and outputs. Something that is largely neglected in the modelling studies reviewed in this paper is a clear representation of the uncertainty of the resulting model. Overall accuracy results are reported, but clear uncertainty bounds could be useful if provided to an optimisation strategy to prevent practically infeasible or undesirable solutions.

5.1. Requirements for holistic optimisation of multi-vector energy systems

Providing suitably accurate modelling of district energy generation and conversion technologies can be developed, the clear future research task is to integrate these models into a unified district energy management platform. To strive towards optimal operation in terms of minimal cost to consumers, reduction in primary energy consumption and reduction in greenhouse gas emissions, the authors believe any such platform must have a number of distinct modules set out in Fig. 5.

- Data Logging - There must be a direct interface to log key time-series data from currently underutilised, Building Management Systems (BMS) and other district-level generation and energy network sensors. Increased availability of sensor data through growth in the Internet of Things (IoT) technology should also be envisaged and embraced as a new opportunity for district energy management.

- Prediction - Forecasting models of pivotal energy management trends such as building energy demand and expected renewable energy generation should be created using the methods discussed in this review. This module should also be responsible for directly predicting expected weather conditions or retrieving this information from local weather data repositories. Exploration of newer, ensemble-based prediction methods such as Random Forest should be tested in this field.

- Optimisation - Information developed in the prediction module should be utilised to optimise the set point strategy of the controllable district energy generation over the period of 24 h. Internal simplified models, reviewed in this paper, will form essential components of this optimisation algorithms to evaluate potential solutions effectively. Efforts should also be made to include models to simulate the behaviour of the energy networks and energy storage solutions which were considered beyond the scope of this review but undoubtedly influence solution feasibility in multi-vector networks. By digitising and up-scaling energy management to an urban level, developments in cloud computing technology could be leveraged to provide increased computational resource.

- User display - Easy to read, Key Performance Indicators (KPI’s) must be effectively communicated to a facility manager in a graphical manner. Changes to the baseline scenario made by the optimisation should be clear alongside automatic fault detection. Communication through this module must be bi-directional, allowing a facility manager to provide their inputs and overrides due to user request. The development of recognised district scale, urban sustainability assessment criteria is required to provide more instructive feedback.

- Semantic modelling - The distinct modules that form the district energy management platform must be unified through higher order semantic models of the district. This underpins the entire platform structure, allowing data interoperability through common machine interpretable descriptions of district components. Furthermore, by exploiting semantic reasoning and inference of rich, contextualised data, automatic fault detection, adaptation to unforeseen circumstances and semi-automatic parameter selection for modelling can be achieved.

As discussed, the semantic description of the district energy system provides an additional level of robustness to energy management at an urban scale. A comprehensive and generic description allows future adaptation, scalability, and interoperability. For example, future expansion or retrofitting projects could be simulated in the design stage and then easily integrated upon project completion with no requirement for a complete model overhaul. Furthermore, due to the semantic basis of this energy management platform, the information could be mapped to provide inputs to wider regional scale modelling platforms through a higher level semantic model that encompasses factors such as transportation or wider energy policy planning.

Application of advanced deep learning techniques has not yet become commonplace within the building energy domain. However, the inclusion of an extensive store of semantically enriched historical data could also allow the exploitation of newer AI techniques such as deep learning, automatic feature extraction and unsupervised learning. This could potentially improve the prediction accuracy of many of the models detailed in this paper. Furthermore, these models could adapt to changes in system configuration providing they are continually re-training based on the most recent available data. Feature extraction techniques and unsupervised learning have the potential to determine the optimal input variables to a machine learning model without relying on the prior knowledge of facility managers.

A further, vital future research direction is the requirement for an optimisation methodology that leverages flexibility in building demand in addition to the flexibility provided in the supply-side by a multi-vector energy system. This can be achieved in several different ways that will likely depend on the configuration of the selected district, the number of different stakeholders and the building mix (residential or publicly owned). In the
6. Conclusion

This paper has aimed to provide a wide-ranging review of state of the art techniques for modelling multi-vector district energy systems. The key criterion of the review was modelling techniques for operational optimisation, meaning a sub-hourly time resolution requiring short computational times and simplified models. Therefore, throughout this paper, newer machine learning methods, largely neglected in previous reviews, have contributed a sizeable proportion of the paper. The conversion technologies considered in this review are CHP, boilers, solar PV, solar thermal, wind turbines, Power-to-Gas, and heat pumps. Of these technologies, most are well covered within the literature. However, wind speed and power prediction remain challenging, and Power-to-Gas modelling has yet to be explored by researchers due to its infancy.

Building modelling has also been included in this review as the authors felt that inclusion of demand-side modelling is essential to any district energy optimisation. The three broad categories are physics based white box models, data or machine learning based black box models and in between hybrid grey box models. This review agrees with previous studies that white box modelling techniques are too complex, and computationally time-consuming for use in real-time optimisation. However, both grey box and black box models have proven effective in modelling building energy demand and indoor thermal conditions.

Finally, this article has outlined the future research directions by illustrating the ideal potential district energy management platform. This requires several distinct modules; including data recording from BMS, a prediction module based on historical recorded
data or otherwise, an optimisation module to take advantage of the district energy predictions and generate set points for controllable supply, and a well-developed user-interface to illustrate to the facility manager the impact of the optimisation and the overall performance of the district. These modules will be supported by a higher order semantic model describing the district energy system to ensure interoperability and communication between management modules. Work to achieve this vision of a future district energy management platform is ongoing through the authors' work as part of the PENTAGON project and through the development of an integrated, semantic, computational urban sustainability platform.

Acknowledgements

The authors would like to acknowledge the financial support of EPSRC (Engineering and Physical Sciences Research Council) and BRE (Building Research Establishment) as well as the European Commission as part of the Horizon2020 THERMOS (Project Id: 723562) and PENTAGON (Project Id: 731125) projects.

References


[19] B. Jukaj, Mathematical modelling of a steam boiler room to research thermal


