Applied Energy 253 (2019) 113603

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage

Ian J. Scott[a,b,*], Pedro M.S. Carvalho[c], Audun Botterud[b,d], Carlos A. Silva[a]

[a] Center for Innovation, Technology and Policy Research – IN+, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[b] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Instituto Superior Técnico and INESC-ID, Universidade de Lisboa, Lisbon, Portugal
[d] Energy Systems Division, Argonne National Laboratory, Argonne, IL, USA

HIGHLIGHTS

- Demonstrate importance of incorporating storage and ramping dynamics in clustering.
- Improve expansion planning model accuracy by 61% with adjusted cluster weights.
- Weights allow accurate modelling of total energy, peak demand, and ramp dynamics.
- Incorporate ramping challenges to clustering approaches improving results.
- Demonstrate under-representation of energy storage without ramping challenges.

ARTICLE INFO

Keywords:
Generation expansion planning
Electricity system modelling
Optimization
Clustering
Representative input selection
load modeling

ABSTRACT

Decision makers rely on models to make important regulatory, policy, and investment decisions. For power systems, these models must capture (i) the future challenges introduced by the integration of large quantities of variable renewable energy sources and (ii) the role that energy storage technologies should play. In this paper, we explore several different approaches to selecting representative days for generation expansion planning models, focusing on capturing these dynamics. Further, we propose a new methodology for adjusting the outputs of clustering algorithms that provides three advantages: the targeted level of net demand is captured, the underlying net demand shapes that define ramping challenges are accurately represented, and the relationship between annual energy and peak demand is captured. This weighting methodology reduces the magnitude of the error in the representative day based generation expansion planning models estimation of costs by 61% on average. The results also demonstrate the importance of carefully performing the clustering of representative days for both system costs and technology mix. In most cases improvements to the total cost of different representative day based expansion plans are realised where conventional generation capacity is substituted for energy storage. Based on the energy storage technology selected we conclude this capacity is being used to address ramping challenges as opposed to shifting renewable generation from off to on peak periods, reinforcing the importance of capturing detailed intraday dynamics in the representative day selection process.

1. Introduction

Power systems are changing at an accelerated pace with the introduction of ever greater proportions of renewable energy, smart grid technologies, electric vehicles, and the wider availability of affordable energy storage (ES). These changes, particularly to the types of technologies used in the production of power, are forecast to introduce a number of new challenges particularly around the area of flexibility [1]. Decision makers, such as energy companies, power system operators, and government policy makers, are then faced with an important problem: how to make decisions about the direction the power system should evolve to meet these challenges without having perfect knowledge of the scope of the challenge or the costs and benefits of potential future technologies.

To make these decisions we turn to the use of models, simplified representations of the system, to forecast and assess different future investment options. However, the future challenges are predicted to be different and more complicated than those seen historically [2]. For example, the problem of integrating variable renewable energy, and the demand for

* Corresponding author at: Center for Innovation, Technology and Policy Research – IN+, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.
E-mail address: ian.scott@tecnico.ulisboa.pt (I.J. Scott).

https://doi.org/10.1016/j.apenergy.2019.113603
Received 14 December 2018; Received in revised form 18 July 2019; Accepted 20 July 2019
Available online 01 August 2019

Acronyms

General

CCGT	combined cycle gas turbines
ES	energy storage
GEP	generation expansion planning
ND	net demand
NRMSD	normalised root mean square deviation
OCGT	open cycle gas turbines
PAM	partitioning around medoids
RDDC	representative day duration curve
UC	unit commitment

Results

DC	duration curve
DWS	demand, wind, solar series
GMMD	greatest min-max difference
GMMD6	greatest min–max difference over 6 h
MCRD	maximum consecutive positive change down
MCRU	maximum consecutive positive change up
MinV	minimum value
MRD	maximum single period negative change (peak ramp down)
MRU	maximum single period positive change (peak ramp up)
NDHF	historical and future net demand
RDC	ramp duration curve
S	data series (ordered)

flexibility it places on the system, increases with the proportion of renewable energy in the system and will therefore be significantly higher in the future [1]. This flexibility requirement has important implications for the optimal build and operation of all units [3], affecting costs, emissions [4], and sustainability [5]. The inclusion of the dynamics resulting from renewable integration, and the role that ES [6] and demand side response can play in addressing them, are classified as important current challenges for the long term modelling of power systems and the generation expansion planning (GEP) problem [7].

The study of generation expansion with high levels of renewable energy is a particularly active area of study (see [2,6,8] for recent reviews). In [6] the authors find optimisation models to be both the most common approach and the most suited to capturing the level of technical detail required to represent flexibility challenges. However, when applying a model, particularly an optimisation model, we are limited in our ability to add detail, such as future flexibility requirements, before the model becomes unwieldy. An important modelling task is then to select which areas of the real system we can simplify or abstract out. As the future system challenges require further detail in some areas of the model, reducing the scope in other areas may be required. A common area to reduce detail is the representation of time (the level of temporal granularity) included in the model to allow for the more complex incorporation of other factors such as operational details.

A trade-off then exists between model detail or model complexity and the level of temporal granularity included. This trade-off is an active area of research. In [9], the authors explored the trade-off between spatial and temporal granularity in a planning model for electricity expansion in California. The authors found the largest change in cost was associated with changes in spatial detail, which suggested a preference for spatial detail over temporal detail in model formulation. On the other hand, [10] examines the trade-off between the level of temporal detail and the level of operational detail in energy system planning models and finds that temporal detail and the method of selecting the temporal resolution to be important for results in the high renewable penetration systems of the future. Further, an increasingly important trend is the use of stochastic optimisation to attempt to deal with uncertainty in planning model inputs. To facilitate this, [11] explores the trade-off between the level of stochastic and temporal variability and concludes that temporal detail should be sacrificed to improve the level of stochastic representation.

The amount of temporal detail we can include in any model is then severely limited by our need to include detail in other areas. The challenge is to ensure that sufficient temporal detail is included so that the model is still representative. In a GEP model this is complicated by the fact that many of the important challenges of the future, such as flexibility provision (see [3;12]), demand shifting and storage operation [9] are dependent on the representation of time. For example, a detailed representation of flexibility in the model operational detail (such as integer unit commitment combined with ramping and start-up constraints) is important for results [4] yet, it cannot be expected to produce accurate results without the inclusion of the temporal detail that generates the flexibility challenges, as demonstrated by [13].

The task of selecting the level of temporal detail is an important one and is increasingly attracting academic attention. A common initial approach was to use a number of representative individual periods (see, for example, [12,14,15]), often calculated from a load duration curve, which minimises the computational effort [15]. This approach is particularly common in energy system models [16] (although work has been done to extend this approach in this context [17]). The problem with this approach is that the importance of the transition between time periods is lost; on the short-term ramping scale, the daily storage operation scale, and on the longer scales of weekdays vs. weekend or summer vs. winter.

To preserve these relationships, a number of studies then instead select larger blocks of representative time periods; for example, either representative days (as in [3,17,18,19,20]), or representative weeks (as in [21,22]). To incorporate seasonality, representations can be selected from periods of less than a year (monthly or seasonally) as in [20], for example, or alternative approaches, such as [23], which retain seasonal chronology, can be considered.

An increasingly popular means to select which periods should be included as representative is that of clustering, where observed historical data is grouped and a representative period (hour, day, or week) is selected to represent each group. A number of different algorithms are used, including k-means, k-medoids, principle component analysis, and hierarchal clustering (see [24], for an overview, and [25,26], for comparisons in the application to this task).

This paper extends previous work on the selection of representative days for power system expansion models where renewable build is treated exogenously (as is often the case with renewable integration studies). The work seeks to ensure that the future challenges of providing the flexibility requirement for the incorporation of variable renewable energy sources and the role of ES is accurately represented for the purpose of long-term planning. Towards this end, the main contributions of this work are that we:

- Develop a new methodology for adjusting the weights calculated by clustering algorithms for representative days before using them in the GEP model. The new approach achieves three important goals: it ensures forecast annual energy requirements are modelled, the relationship between annual energy and the peak value is preserved, and the inter-period ramping dynamics that drive flexibility requirements are accurately captured. This generic approach can be easily applied to any existing clustering methodology and is shown here to result in a significant improvement to the GEP model's ability to estimate costs.

- Explore different approaches to representing historically observed data (i.e. separate time series for load and renewables or a combined time series for net load, the means for determining net demand, and the use of duration curves or ordered series) in the clustering algorithm and find large differences in the quality of GEP modelling results based on this choice.
- Test the inclusion of a number of potential 'extreme points' or 'days of interest' to improve the representation of future flexibility requirements and determine the benefits for accurately modelling the role of ES.
- Compare and assess the technology mix in generation expansion plan results based on the emphasis of different features of the data to explore the implications for future technology selection based on the accurate inclusion of future power system requirements (and the potential biases where these are ignored).

In the remainder of this article we outline the clustering methodology used for representative days selection (Section 2.1), the different means of including input data and extreme points (Section 2.2), our proposed weighting methodology (Section 2.3), and the test GEP model and dataset (Section 2.4). We then follow with results (Section 3) and draw conclusions for the best approach to selecting representative days for GEP (Section 4).

2. Methodology

2.1. Clustering methodology

We make use of clustering techniques, as it is increasingly common for the task of selecting representative input data for GEP. A number of possible clustering algorithms that could be applied to our data exist, as reviewed by [24], with 'k means' being perhaps the most common. An important consideration in the selection of algorithm is which output will be used in the GEP model. One possibility would be to use an average of all data in the cluster (a k-means approach is based on minimising the distance to this average). However, an average day load or wind profile will necessarily reduce the hour-to-hour volatility by smoothing [27]. Instead, we use an observed day in the GEP model and so use a k-medoids approach where we select an actual observed day that minimises the distance of the selected day to the other members of its cluster. The specific algorithm is based upon Partitioning Around Medoids [28]. This approach is validated by the finding that k-medoids clustering can reduce error in long-term planning studies [25,29].

Typically, of more importance than the clustering algorithm itself, is the treatment and representation of the data to be clustered [26]. We test a number of different approaches to represent the data in the clustering; more specifically, we broadly test across two dimensions:

- **Representation of data** - The data series chosen to represent a day in the clustering algorithm (for example net demand vs. separate wind, solar and demand series).
- **Extreme points -** The potential inclusion in the clustering output of extreme points or 'days of interest' assumed to be important to the accuracy of the generation expansion planning model.

Allowing for the manipulation of these dimensions has implications on the implementation of the clustering algorithm as described in Section 2.1.1. The data selection, clustering, and weighting process is summarised in Fig. 1.

2.1.1. Algorithm

The clustering algorithm developed for this analysis is based on the Partitioning Around Medoids (PAM) algorithm [28]. For k representative clusters with the inclusion of p preselected medoids, the algorithm can be written in seven steps, as follows:

1. **Normalise data** - The data is a collection of days each with one or more time series of observations (e.g. 48 half hourly net demand observations). Each observation in each series is normalised with the series mean and standard deviation.
2. **Force medoids -** Determine p preselected days and set these as medoids for the first p clusters.
3. **Initialise remaining medoids -** Randomly select the remaining $k - p$ medoids from the dataset as initial medoids for the remaining clusters.
4. **Determine clusters -** For each day determine the distance (using the below described distance metric, see Section 2.1.1.1) to each of the k medoids and find the closest medoid. Groups of days with the same closest medoid are considered 'clusters'.
5. **Update medoids -** For each of the k - p clusters without a pre-selected medoid calculate the distance between each day and all other days (see Section 2.1.1.1) – select the day with the minimum total distance to all other days in the same cluster as the new medoid for that cluster.
6. **Converge -** Repeat steps 4–5 until the there is no change in the closest medoids for each day (or a maximum number of convergence iterations reached).

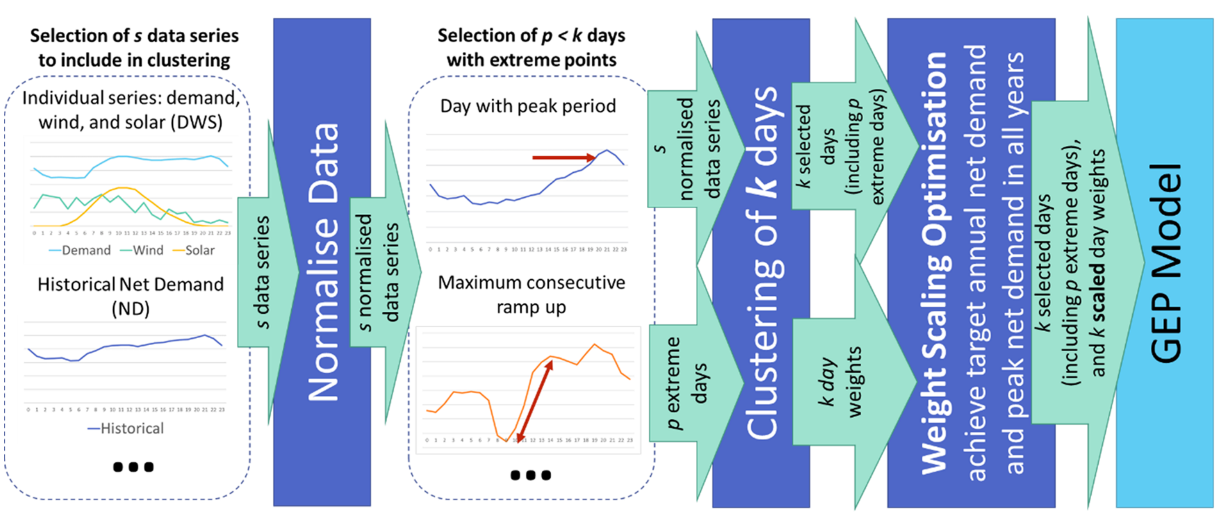

Fig. 1. Diagram of data selection and clustering process for selection and weighting of representative days for GEP model.

7. **Report -** For each cluster of days, report the selected medoid and the cluster size (to be used as a weight in GEP) as well as the total distance of all days in the cluster to their cluster medoid.

The quality of the solution found may depend on the random selection of initial medoids. The algorithm is typically repeated for several different randomly selected starts, and the result with the lowest total distance metric is kept. In this case, we repeat the algorithm 10,000 times.

2.1.1.1. Distance metric. The unit of selection here is a day, which we test with different levels of data, for example either a net demand series with 48 observations or two separate wind and demand series with 48 observations each. The distance metric used is an average squared Euclidean distance, which helps ensure consistency across different numbers of series and different series sizes (for example two data series per day vs. one):

$$Distance(d, d') = \frac{\sum_{s=1}^{n_s} \frac{\sum_{t=1}^{n_t}(d_{s,t} - d'_{s,t})^2}{n_t}}{n_s} \tag{1}$$

where d and d' are two selected days, each day is represented by n_s data series. With each data series containing n_t observations (here 48 half hour periods). $d_{s,t}$ is then the values of series s at time t on day d.

2.2. Representation of input data

2.2.1. Selection of series for clustering algorithm

Starting with historically observed data for a power system, there are a number of different ways we could consider using this information for the clustering algorithm.

Individual series: historical demand, wind capacity factor, and solar capacity factor - The data that determines net demand for each day can be represented as individual series (one series each for load, wind, and solar), as in [12]. Here demand is in MW per period, wind

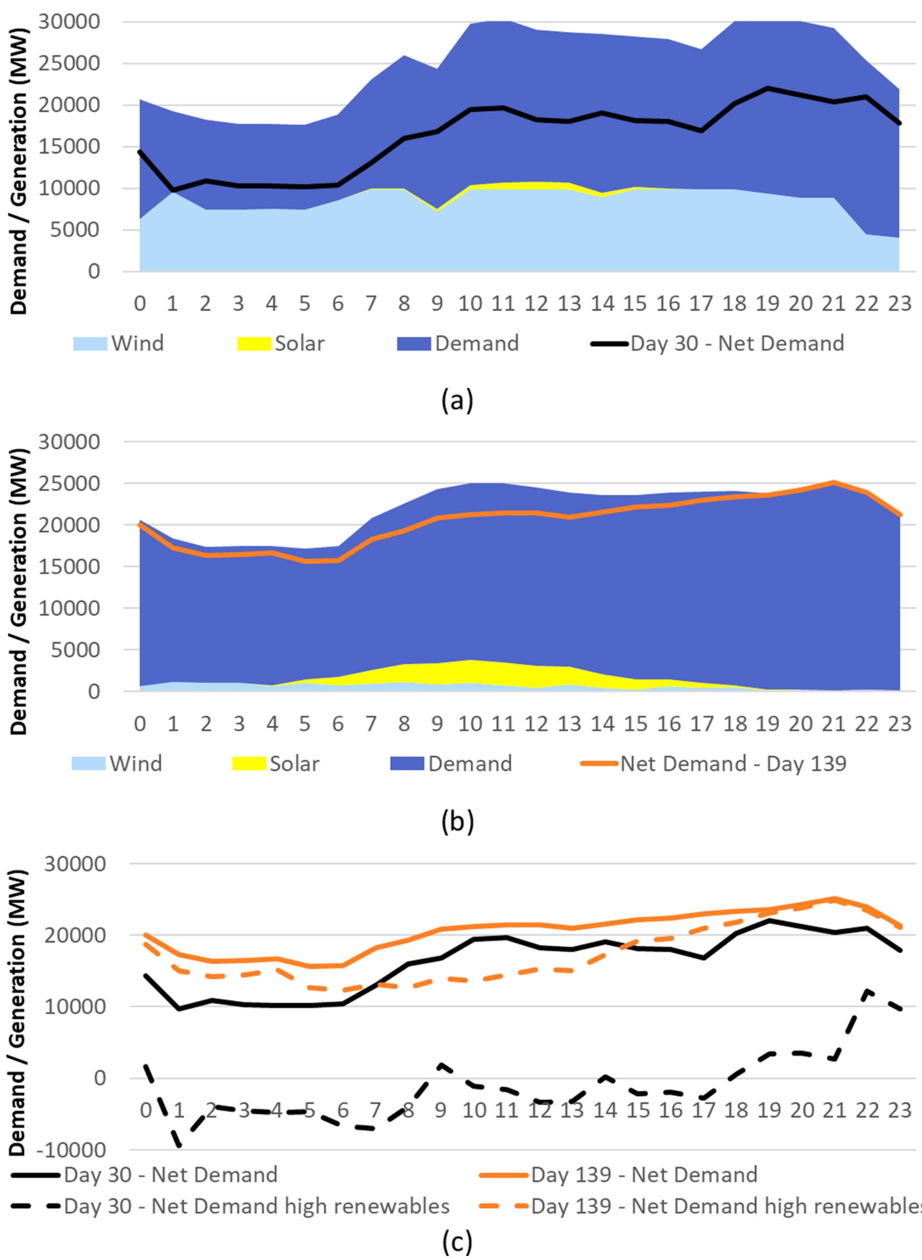

Fig. 2. (a): Determination of historical net demand day 1; (b) Determination of historical net demand day 2; (c) Comparison of net demand with different levels of future renewable uptake.

and solar in average fleet capacity factor for the period.

Net demand series – the data for wind, solar, and demand can be combined to determine net load series, as is in [22]. The derivation of the net load curve would be different for different levels of future wind and solar capacity (see Section 2.2.2, for discussion). The advantage of this approach is that it considers the relative importance of the underlying series (demand, wind, solar, etc.).

Net demand duration curve - An alternative common representation for a net demand series is that of a duration curve. Here the net demand is ordered from highest to lowest value for each day. The advantage of this approach for clustering is the highest demand period (and second highest, and third highest, etc.) for each day is compared to the highest of every other day, which would not happen with the chronological series, if they happen to fall in different periods. The disadvantage of the duration curve is that it does not represent the chronological nature of the time series. That is, it loses all information about the size of the transition between periods, which can drive flexibility requirements.

Net demand ramp duration curve – [26] introduces the concept of a ramp duration curve. First the ramp rates are calculated (by differencing the net demand series) and then ordered to form a duration curve. This allows an additional series for the algorithm to cluster on what represents the ramping requirements of each day.

We test these representations in different combinations to attempt to find a representation of the input series that allows the algorithm to cluster against the dynamics that are important for the GEP model.

2.2.2. Historical and future net demand

Clustering is often performed on historical observed net demand. The issue with this is demonstrated in Fig. 2, where two days are compared: day 30 in winter (a) and day 139 in spring (b). In chart (c) with historical levels of renewable capacity the two net demand levels are reasonably close (solid lines). However, at future levels of renewable uptake the net demand of the two days is drastically different (dotted lines).

This presents a potential problem when clustering days based on historical net demand as the clustering algorithm will see little difference between the two days (and cluster them together) when, in fact, for the purpose of the GEP model, they are very different.

For this reason, we expect that clustering based on a series derived from historical renewable deployment may not produce a net demand series which is representative for the whole period of the expansion planning problem. We therefore compare two representations of net demand:

- One where each day has a historical net demand series (**ND**).
- One where each day has two net demand series; one based upon historical renewable deployment and another based upon the future deployment of the modelled year (**NDHF**).

It is important to emphasise that knowledge of future net demand is required to include this series in the clustering, which is only the case where we have an idea of future renewable deployment. However, this is often the case as renewable build is still driven largely by government policy and usually determined outside of a strict cost minimisation model. Renewable integration (or investment given renewable build) is an active area of research to which this methodology is targeted [2].

2.2.3. Inclusion of 'days of interest' (forced inclusion of selected medoids)

For the purpose of modelling power systems, clustering can under-represent extreme points which are important for driving GEP results [14,30]. Forcing the inclusion of specific observations based on the modellers prior knowledge of the problem improves the quality of results for generation expansion planning. An obvious example is the inclusion of the 'Peak' period of the year as this will be a considerable driver of the expansion of capacity and is unlikely to be central to a cluster (as it is an extreme point) and, therefore, unlikely to be otherwise selected. The peak day is, therefore, often "manually" added to the selection as an additional data point or set as a cluster centroid [30,27]. Without this forced inclusion, the clustering will take a central point of the cluster that the day with the peak period is within to represent that period and the generation expansion planning model will not contain the exact peak demand required to be met in the future (based on historical peak demand).

As clustering tends to under-represent extreme points by selecting central points or averages [14,25,31], and as these can be important in driving expansion results [14,30], we examine the inclusion of a number of 'days of interest'. These 'days of interest' contain specific observations that may help to ensure the selected representative days capture the future power system challenges relating to flexibility and ES operation, which we are interested in. The extreme points tested here are described in Table 1.

The additional medoids are chosen based on similar reasoning to that of the peak day – that results can be driven by extreme values. We test two representations of future ramping requirements to ensure that the system builds the flexibility to deal with the greatest flexibility challenge. We test both the greatest ramp (up and down) between two individual periods and the greatest consecutive ramp (up and down).

To test properties that may be important to the deployment of ES devices that operate on a daily cycle, we test two 'days of interest'. The extreme min-max differences are tested to attempt to capture the greatest value a storage can achieve in a single hour and the greatest min max distance over six hours as an estimate of the value that can be achieved by using the full storage capacity.

Finally, minimum values are included as they are likely to be important for representing future renewable curtailment challenges (either through curtailing renewables or forcing conventional generation off and incurring future start-up costs).

Where the input data includes multiple series (e.g. net demand data for current and future renewable penetration), the selected observation (e.g. day with peak period) for each of series is included.

Table 1
Description of days of interest.

	Days of interest (day with:)	Description
	Maximum value (peak period) – **Always included**	Day with the series highest recorded value
GMMD	Greatest min-max difference	Day with the greatest difference between the highest and lowest value
GMMD6	Greatest min-max difference over 6 h	Day with the greatest difference between the 6 h with the highest values and the 6 h with the lowest values
MCRU	Maximum consecutive positive change up	Day with the highest positive increase over consecutive periods (largest consecutive ramp up)
MCRD	Maximum consecutive positive change down	Day with the highest negative change over consecutive periods (largest consecutive ramp down)
MRU	Maximum single period change positive (peak period ramp up)	Day with the highest positive difference between two periods
MRD	Maximum single period change negative (peak period ramp down)	Day with the highest negative difference between two periods
MinV	Minimum value	Day with the series with the lowest recorded value

2.3. Weighting of representative days for expansion planning model

The output of the clustering algorithm consists of several selected typical days (cluster medoids) and their corresponding cluster sizes (that quantify the number of actual days that are being represented by each selected medoid day). However, the summed product of these is not guaranteed to equal the total energy (or renewable production) of the observed data, which is an important factor in the GEP model.

Careful treatment of this problem is particularly important when comparing different representative day selections. For example, if two representative day selections have different levels of net demand to meet, this will influence the expansion results.

It is important to note that small percentage errors can be significant here as the error is applied to total annual net demand. As such a 1–2% error in our case study results in a significant difference in the energy needed to be procured from additional units and will result in a switch between peaking (or ES) and CCGT technologies. In the literature, estimates of the size of this error vary, particularly depending on the number of representative days, however, it is often significant. For example, in [26] the authors find errors of approximately 2–6%, when using a clustering methodology, even with 8 or more selected days. Additionally, [32] reports values in the 1.8–5% range again even when 8 or more days depending on methodology used.

Additionally, we can expect the effect on the estimation of costs to be greater than the underlying error in net demand. For example, if we introduced a 5% error in total net demand by simply increasing all weights by 5%, then we would expect roughly a 5% increase in operating costs. However, when the clustering outputs result in a 5% higher total net demand, this means the high net demand days (high cost) have higher weights and the low net demand days (low cost) have lower weights, resulting in a disproportionate change in the estimation of costs.

In some approaches, the underlying data is scaled[1] so that this error is removed (as seen, for example in [16,27,32,33]). In practice, the scaling to observed values is often done in a way so as to preserve the value of the peak period [16]. This means that some periods are scaled more than others (the peak period is not scaled, and no other period is scaled to be above the peak period). The scaling is then different for different periods and this will alter the change between periods (the required ramp rate) - a key potential metric for driving system costs. Ramp rate requirements around the peak period, a potentially extremely challenging period for the model as most capacity is already committed at high load, will potentially be altered the most. An illustrative example is included in Fig. 3, where a single representative day is selected, one with a net demand 10% lower than the average over the full dataset.

Here we propose an alternative methodology that allows us to preserve underlying data series and instead adjusts the weights applied to each day in the GEP. The methodology is as follows:

- Select representative days and determine cluster weightings for these days (as outlined in Section 2.1).
- For each year that is being modelled, determine the net demand associated with each day. The net demand will depend on future demand, renewable modelled capacity[2], expected change in renewable annual capacity factor (if different to historical), and the day's renewable capacity factor.
- Find a set of weights to be used in the GEP (a weight for each cluster medoid, for each year) as close as possible to the cluster weights that ensures total net demand in the GEP.

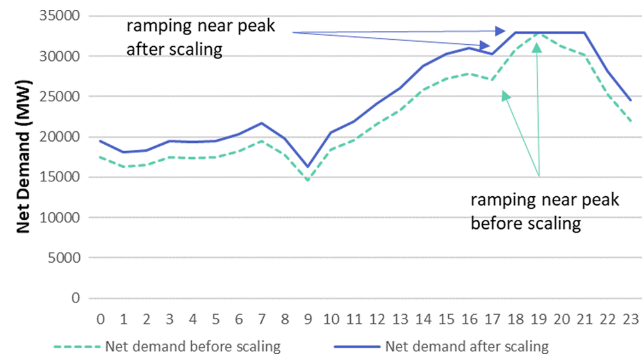

Fig. 3. Illustration of scaling of net demand series, maintaining peak value to increase average net demand by 10%, as in [16].

To determine the new weights, we solve a simple quadratic optimisation problem that calculates a scalar[3] for each cluster size that minimises the scalar squared distance from 1 (Eq. (2)) to achieve a final weight as close to the original cluster size as possible. The optimisation is constrained to ensure that the total net demand is achieved (Eq. (3)) and the sum of the weights remains unchanged (Eq. (4)). Additionally, the formulation requires that no weight can be scaled below 1 or above the total number of days (Eq. (5)).

The optimisation problem needs to be solved for each year in the GEP model (unless no changes in relative shares of wind, demand, or solar are expected as the level of Net Demand associated with a representative day will change with these shares). The mathematical formulation of the problem is given below.

$$\min_{s} \sum_{d\in D} (1 - s_d)^2 \tag{2}$$

$$\text{s.t.} \quad \sum_{d\in D} w_d \cdot ND_d \cdot s_d = ND^{Total} \tag{3}$$

$$\textstyle\sum_{d\in D} s_d \cdot w_d = D^{Count} \tag{4}$$

$$1 \le w_d \cdot s_d \le D^{Count} \qquad \forall\ d \in D \tag{5}$$

where s_d is the scalar to be applied to the weight w_d of each representative day $d \in D$ when used in the GEP model; ND_d is the total net demand of day d and ND^{Total} is the total net demand of the underlying dataset over all D^{Count} days.

While this formulation is not guaranteed to be feasible[4] in general, it was found to be so for all cases analysed here.

Here we focus on the case where the future net load forecast is an input to the model. This is often the case when renewable build is determined by factors exogenous to the model (e.g. government policy) and electricity demand is forecast exogenously. It is possible to formulate this problem so that not just net demand is achieved, but also capacity factors for wind and solar. However, this is a significantly more restricted problem which results in unreasonably large scalars being calculated (where the problem is feasible) in this case study, with the average absolute change to weights being greater than 50%.

This formulation, additionally, allows us to either treat renewable production as exogenous and simply use net demand in the GEP model,

[1] For example, if we had selected one representative day, we would scale the net demand values for each hour so the total net demand for the day multiplied by the number of days in the year would equal the full dataset annual total.

[2] Here we focus on the case where renewable capacities are known in advance (potentially as uptake is driven by policy targets as opposed to market dynamics).

[3] A formulation based on adding or subtracting to the weights was also considered. However, this approach tends to adjust the days with smaller weights to a relatively higher degree, which also tend to be the more extreme days, and results in more extreme final weighted representative days.

[4] For example, if all selected days had higher/lower average net demand than the average over the year.

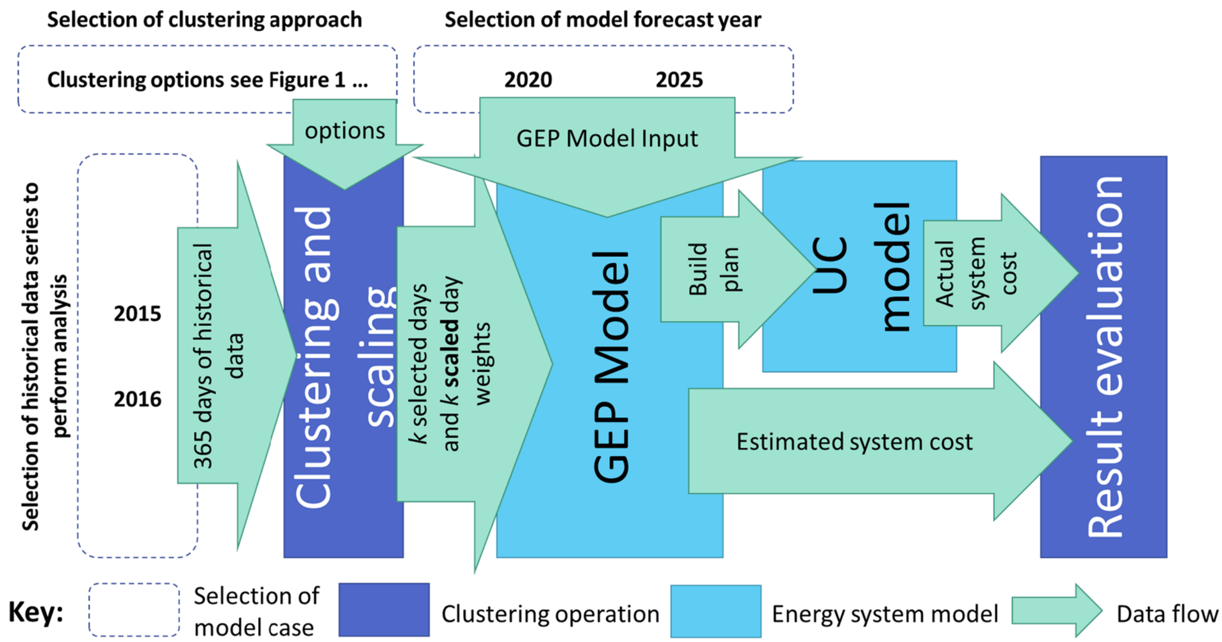

Fig. 4. Diagram of GEP modelling and result evaluation process.

or alternatively, have endogenous representations of renewable capacities and use the capacity factors associated with the representative days in the model.

2.4. Evaluation model

The aim of the representative day selection task is to determine the approach best suited to selecting inputs to a GEP model. To compare the different time period selections, we first simulate the selected represented days and weights in a generation expansion planning model to determine the set of expansion decisions and the estimated system costs.

To determine how well the selected days represent the full data set, the expansion plans are then tested in a daily unit commitment (UC) model. The daily UC model allows us to calculate how the expansion plan performs over the entire dataset, i.e. the full time period of the original data, which cannot be solved practically as a single GEP optimisation problem. The process is summarised in Fig. 4.

2.4.1. Generation expansion planning (GEP) model

The capacity expansion model optimises the investment decisions in new generation technologies accounting for the operational decisions of both new and existing units for the set of representative days. For simplicity, a single future year is included for the GEP model to make build decisions with a discount factor used to reflect the discounted difference between the one off build costs and other annual or variable costs (annual costs for the one modelled year are weighted to represent a discounted 15 years of operation). The objective function follows the following form:

$$\min_{b,o,g,p_t^{UE},p_t^{Dump},p_t^{SUViol},p_t^{SDViol}} \sum_{p\in P}\left(Cost_p^{Build}\bullet\Delta b_p + DF_{15Years}\bullet Cost_p^{Fixed}\right.$$
$$\bullet b_p + DF_{15Years}\bullet \sum_{t\in T} Weight_t\bullet(Cost_{t,p}^{Start} + Cost_{t,p}^{Oper}\bullet o_{t,p} + Cost_{t,p}^{Generation}$$
$$\left.\bullet g_{t,p} + Cost_{t,p}^{Curt})\right) + DF_{15Years}\bullet \sum_{t\in T} Weight_t\bullet(Cost^{UE}\bullet p_t^{UE} + Cost^{Dump}$$
$$\bullet p_t^{Dump} + Cost^{SUViol}\bullet p_t^{SDViol} + Cost^{SDViol}\bullet p_t^{SDViol}) \tag{6}$$

where for each unit $p \in P$ and model time period $t \in T$ the model optimises the build of unit b_p and the operation $o_{t,p}$ and generation $g_{t,p}$ at each point in time. The following costs are accounted for each unit p: build cost $Cost_p^{Build}$, annual fixed costs $Cost_p^{Fixed}$, start costs $Cost_{t,p}^{Start}$, operating cost $Cost_{t,p}^{Oper}$,generation cost $Cost_{t,p}^{Generation}$, and curtailment cost $Cost_{t,p}^{Curt}$ for renewable generators. The model can take penalty actions to incur unserved energy p_t^{UE}, dump energy p_t^{Dump}, spin up violations p_t^{SDViol}, and spin down violations p_t^{SDViol} all with associated costs. The factor $DF_{15Years}$ is applied to annual costs equivalent to 15 years of operation discounted to correctly weight the benefits of building a new unit. Finally, the weight$Weight_t$ is applied to each time period. This is the weight calculated in the weighting methodology and it will be the same for all time periods falling on the same representative day.

The full formulation is included in the supplementary material for brevity. However, it is broadly similar to those found in [34,35] and features:

- Integer build decisions to reflect the reality that it is not possible to build partial units due to technical and economic constraints.
- The use of binary (for existing units) or integer (for new units) operating variables to capture starts and stops. The use of binary and integers in operational constraints is important in valuing flexibility, as discussed in [36].
- Period (half hourly) ramping constraints that ensure that sufficient flexible capacity is committed to meet challenges created by variable renewable technologies.
- The provision of a spinning reserve service; they pose a future challenge to the system, which is important for the integration of renewable energy [37], and one that ES can help resolve. Note that we assume that renewable generators cannot provide spinning reserves.
- The formulation of the operation of storage which features: 'full cycle efficiency', provision of reserves, ramping constraints, a constraint to ensure charge and generation are not used simultaneously as this is an unlikely market operation (to dump extra energy where needed). Additionally, the storage is formulated so that the stored energy at the end of the day is the same as that at the start of the day, to prevent transfer of energy between representative days (however, this quantity is freely chosen up to the storage limit).
- Penalties and costs for unserved or dump energy, violations of up and down spinning reserves, and costs to renewable generation curtailment.

2.4.2. Unit commitment (UC) model

To assess the build decisions on the full dataset, we restrict the model to a set of sequential single day problems that focus on the unit commitment decisions. To do so, we fix the build decisions to the optimal values from the GEP model and remove the investment and fixed costs (so they are not incurred every day).

The model is run sequentially for the full dataset (a full year) optimising one full 48 period day at a time with a 48 period look-ahead, as in [38]. That is, the model optimises decisions over a full 96 periods representing two days (to ensure naïve start or shutdown decisions are not made at the end of the first day, as the model must ensure units are online for a second day or incur start costs) but only the results of the first day are recorded. The model then simulates the following day with the initial commitment decisions and initial generation levels set from the end of the previous day (period 47).

2.4.3. Simulation system

All simulations are performed on a laptop with an Intel Core i7-6500U CPU @ 2.50 GHz and 8 GB of RAM. Models are implemented in C# and optimisations problems solved with the Gurobi solver. Optimality gap target of 0.1% is reached with an average time of 2013 s.

2.5. Case study

The United Kingdom (UK) has committed to decarbonisation and the electricity system is expected to feature large quantities of renewable energy in the near future. We make use of this system as a case study and examine two cases with differing levels of renewable penetration:

Medium term system (2020) - medium renewable penetration – the UK market as it currently exists with expected conventional retirements and expected increases in renewable capacities [39].

Future system (2025) – high renewable penetration – In this future scenario, there is a significant growth in wind and solar generation capacity and a decrease in conventional dispatchable capacity. From the capacity mix forecast for 2025 in [39], we add 2% growth to demand and allow the model to select the capacity to meet this gap. See Fig. 5, for capacity mixes.

For robustness, we perform the clustering exercise and all analysis on two separate historical datasets (one for 2015 and another for 2016) based on half-hourly demand, solar, and wind profiles[5].

2.5.1. Generation expansion planning

We examine the ability of a number of technologies to meet future system challenges. In particular, we compare two different theoretical battery storage technologies, one with 2 h and another with 5 h of storage capacity. While these technologies may not be representative of the current economical storage sizing, they allow us to distinguish the reason storage is being built. If the model is using storage to shift large quantities of load (overnight wind to peak or solar to mid evening) the larger storage capacity would be required. However, if storage is being used to meet ramping challenges it may not be the case that such a large storage is required, and a smaller cheaper storage may be selected by the model.

We compare these two storage technologies to the conventional technologies CCGT and to a Peaker (OCGT).

We include a supply curve for each technology by increasing the capital costs of each technology as more of that technology is built[6]. For our purposes, this assumption improves the robustness of our comparison as a small change cannot result in the optimal technology mix switching completely from one technology to another and we can be more confident of large technology changes when we observe them.

For more information on unit characteristics, see supplementary material.

3. Results

In this section, we look at the effect of changing the representation of the input series to the clustering (Section 3.1), before the inclusion of extreme points (Section 3.2), and assessing the scaling methodology (Section 3.3). For all results we examine the ability of the clustering to represent the underlying demand and, additionally, the ability to perform in a GEP model. For more details on the assumptions underlying this analysis, please see the supplementary material.

3.1. Representation of data input series

All cases report results for 9 selected days with the day including the peak value for each series forced as a centroid for inclusion and are described in Table 2.

3.1.1. Clustering results

Fig. 6 displays the load duration curves constructed from the different representative day selections and compared to the full data set (for the 2016 input data case). The peak value is the same across all selections based upon net demand, as the day including this value is enforced as the maximum of that series. For the selection based upon individual demand, wind, and solar inputs (DWS), this is no longer the case as while the peak demand day is captured, the peak net demand is not (as the peak net demand day would have lower renewable production). We resolve this difference with the DWS_NDP case which combines the clustering of DWS with the peak net demand days. As the weightings placed on this peak day are similar across all the selections, and no other days with high net demand values are selected, the high load portions of the load duration curves are similar across all selections. The ability of the selections to capture the low load portion of the net demand series differs more significantly. This effect becomes more pronounced as we look to model the net demand for the 2025 year, with higher levels of renewable penetration, as illustrated in Fig. 6(b). In Fig. 6(c) inclusion of future net demand in the clustering improves the 2025 net demand curve representation at the top and bottom of the load duration curve (highest 10% and lowest 10% of hours) compared to Fig. 6(b), however, the very lowest net demand hours where net demand is negative are still not well represented by any of the approaches. This result demonstrates how extreme points (such as extreme levels of low demand) are often not well captured by clustering

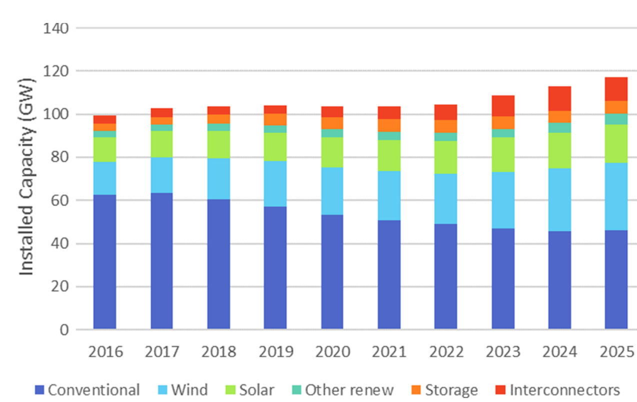

Fig. 5. Installed generation capacity adapted from slow progression scenario [39].

[5] While we would not recommend selecting on a single year of data to find a representative set of profiles for future expansion planning, this approach is sufficient for testing the ability of a clustering algorithm to represent an underlying dataset [26].

[6] This assumption can be thought to reflect either a preference for diversification or that as the optimal siting for new technologies is exploited, less optimal siting must be used and installation costs increase (or renewable production decreases). An alternative approach would to be allow for technology learning to create a supply curve, as in [40].

Table 2
Key for case abbreviations.

Abbreviation	Case description
ND_S	Historical net demand as original chronological series used in clustering
ND_DC	Historical net demand as duration curve used in clustering
ND_DC_RDC	Historical net demand as duration curve and historical net demand ramp duration curve used in clustering
ND_S_DC_RDC	Historical net demand as original series, as duration curve, and as ramp duration curve all included in clustering
DWS	Individual historical demand, wind capacity factor, and solar capacity factor included in clustering
DWS_NDP	Individual historical demand, wind capacity factor, and solar capacity factor included in clustering. We relax the inclusion of the peak wind day, peak demand day, and peak solar day and instead include the peak net demand day as a forced medoid
NDHF_S	Historical and future year net demand as original chronological series used in clustering (two data series total, per day)
NDHF_DC	Historical and future year net demand as duration curve used in clustering (two data series total, per day)
NDHF_DC_RDC	Historical and future year net demand as duration curve and historical net demand ramp duration curve used in clustering (four data series total, per day)
NDHF_S_DC_RDC	Historical and future year net demand as original series, as duration curve, and as ramp duration curve all included in clustering (six data series total, per day)

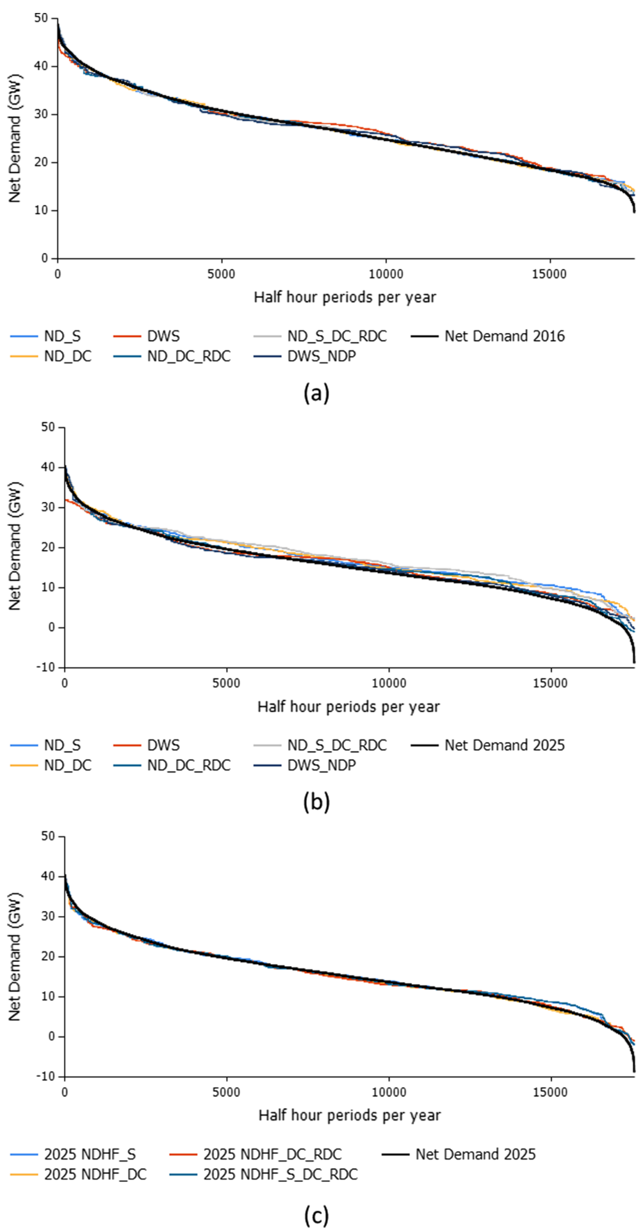

Fig. 6. Load duration curves for full year data and different representative day selections that differ in the representation of data used in clustering; (a) 2016 Net Demand; (b) 2025 Net Demand; (c) 2025 Net Demand including 2025 Net Demand in clustering.

methodologies and must be treated separately where they are considered important for the modelling (as we explore in Section 3.2).

Table 3 reports the clustering algorithm average squared Euclidean distance and a number of metrics relating to the fit of the selected representative days to the full data series as commonly used to asses selections [18,25]. To assess the difference between the full dataset duration curve and the selected day duration curve, we calculate the normalised root mean square deviation (NRMSD) from [18] for each point i on the representative day duration curve $RDDC$ and the full dataset duration curve DC.

$$NRMSD = \frac{\sqrt{\frac{1}{8760}\sum_{i=1}^{8760}(RDDC_i - DC_i)^2}}{\frac{1}{8760}\sum_{i=1}^{8760}DC_i}$$

The results for the average squared Euclidean distance, that the clustering algorithm attempts to minimise, demonstrate that the ordered duration curve data-based approaches (ND_DC, ND_DC_RDC) are significantly easier to cluster than the unordered net demand or individual series data. However, when we look at how well the selections fit the load duration curve as a whole (as measured by NRMSD) we find the values increase the further into the future we look. Again, this indicates that the representative days fit to the underlying data decreases for future years (as wind and solar capacity increase in the underlying net demand). The individual series clustering (DWS) is the least susceptible to this issue as it inherently places a high value on capturing wind and solar dynamics. This works well for our case study where wind and solar are important drivers of net demand, however, this may not be the case for all applications. The addition of the ramp duration curve to the net load duration curve does improve the represented days ability to fit the ramp duration curve both in the historical and future data as measured by the NRMSD. With the addition of future net demand to clustering series (NDHF), the Euclidean distance the clustering algorithm is attempting to minimise increases significantly as each day contains more data to compare between. However, the representation of the future duration curve improves significantly demonstrating that the Euclidean distance minimised by the clustering may not represent all information relevant for the clustering depending on the representation of the data that is used in the clustering.

3.1.2. Expansion planning model results

In this section we compare how the capacity expansion planning model performs using the different sets of representative days. Fig. 7 displays the error in the estimate of total cost (as compared to a unit commitment model simulating the full dataset) for each of the expansion plans selected by the GEP model. For the most part, when using the net demand series to select representative days, the GEP model can be shown to overestimate the costs, with the exception of the 2025 model year with the 2016 load shapes. The individual series selection (DWS) greatly underestimates costs across all scenarios, potentially reflecting its lower coverage of the higher portion of the net demand load duration curve. To resolve this issue, we include the net demand peak day

Table 3
Clustering metrics - Euclidean distance and normalised root mean square deviation (NRMSD).

	Average Squared Euclidean Distance		Normalised root mean square deviation (NRMSD) of duration curves – average of 2015 and 2016								
	2015	2016	Historical Net Demand	2020 Net Demand	2025 Net Demand	Demand	Wind	Solar	Historical Ramp	2020 Ramp	2025 Ramp
ND_S	29.78	26.28	0.05	0.12	0.28	0.18	0.70	0.17	0.08	0.10	0.15
ND_DC	15.30	15.31	0.05	0.11	0.24	0.16	0.46	0.16	0.07	0.09	0.14
ND_DC_RDC	17.87	18.25	0.06	0.11	0.21	0.14	0.53	0.26	0.06	0.07	0.13
ND_S_DC_RDC	23.10	21.95	0.06	0.12	0.26	0.18	0.65	0.14	0.07	0.08	0.13
NDHF_S	46.70	39.48	0.07	0.09	0.10	0.10	0.35	0.16	0.09	0.12	0.14
NDHF_DC	25.08	23.54	0.06	0.07	0.07	0.11	0.32	0.25	0.07	0.08	0.13
NDHF_DC_RDC	26.62	24.51	0.06	0.08	0.09	0.09	0.25	0.17	0.06	0.07	0.09
NDHF_S_DC_RDC	35.71	31.22	0.06	0.08	0.09	0.07	0.31	0.15	0.06	0.07	0.10
DWS	72.69	77.67	0.13	0.14	0.15	0.12	0.14	0.11	0.11	0.12	0.16
DWS_NDP	75.84	80.86	0.13	0.13	0.14	0.13	0.15	0.11	0.12	0.12	0.17

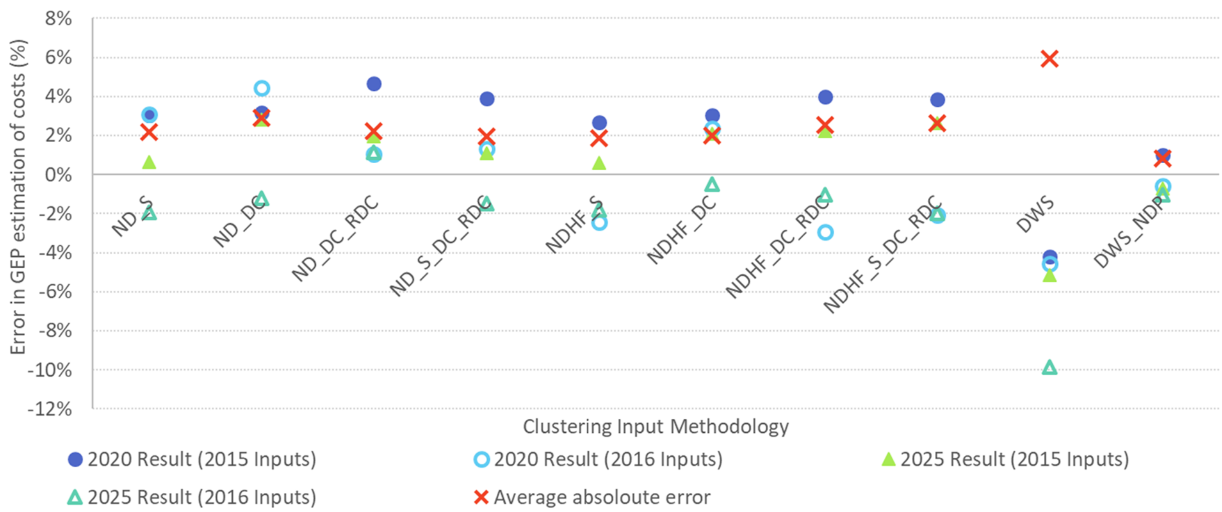

Fig. 7. Difference in GEP model estimation of system costs and full dataset UC estimate of costs (% error) for model year (2020 or 2025) and input data series year (2015 or 2016).

before clustering based upon individual series (DWS_NDP). This result demonstrates a potential difficulty of selecting representative days for a GEP model with endogenous renewable build. It is critical to include the peak net demand day in the inputs to the GEP model, however, this is a product of the renewable build decisions determined by the model and is not known in advance.

When using solely historical data, the addition of the ramp duration curve improves the results (ND_DC_RDC and ND_S_DC_RDC). However, a greater improvement is found by the addition of a future net demand series for both the ordered and the duration curve selections (NDHF_S and NDHF_DC). The magnitude of the error in cost estimation is reduced by 14% (from 2.2% to 1.9%) and 32% (from 2.9% to 2.0%) on average for the ordered series (_S) and duration curve approaches respectively (_DC). The addition of a future net demand series increases the GEP estimation errors where the ramp duration curve is included in the clustering (NDHF_DC_RDC and NDHF_S_DC_RDC).

In addition to the errors in cost estimation, we are interested in how close the solution found is to the optimal one for the entire represented year. While we are unable to solve the expansion problem for the full year, we can compare solutions found to the best solution found as assessed by the unit commitment model for that case (as an approximation of the optimal solution)[7]. Fig. 8 presents the costs above the best solution found (across all combinations tested here) for the different input datasets and GEP model years for each representative day selection method. It should be noted that these costs are small relative to the objective function of the model. This reflects the fact we are making small incremental changes to a large existing power system, as is typical of modelling in most applications. However, the value of poor selections of future technologies would still represent a large cost to society.

In Fig. 8, we see that most representative days perform relatively poorly in the 2020 model year when the 2015 input dataset is used[8]. Using the underlying data series with the peak net demand day (DWS_NDP) performs the best on average followed by the approaches that include the original net demand series (NDHF_S, ND_S_DC_RDS, and ND_S). Compared to a simple clustering on historical net demand duration curves, the DWS_NDP approach improves the accuracy of the estimation of costs by 71% (from 2.9% to 0.8%) and reduces the difference in cost to the optimum expansion plan by 57%. The importance of the unordered original series (ND_S) may reflect a number of factors: firstly, the even relative weighting between the duration curve and ramp duration curve may not reflect the relative importance of these features to the GEP model. Alternatively, the time of day at which ramping occurs (and its correlation with net demand) may be important

[7] Objective function values for best solutions found: £48.3b (2020, year 2015 input), £47.6b (2020, year 2016 shapes), £44.7b (2025, year 2015 input), and £43.7b (2025, year 2016 shapes)

[8] Despite different errors in cost estimation for this case, the final expansion plan for the model is very similar across all cases.

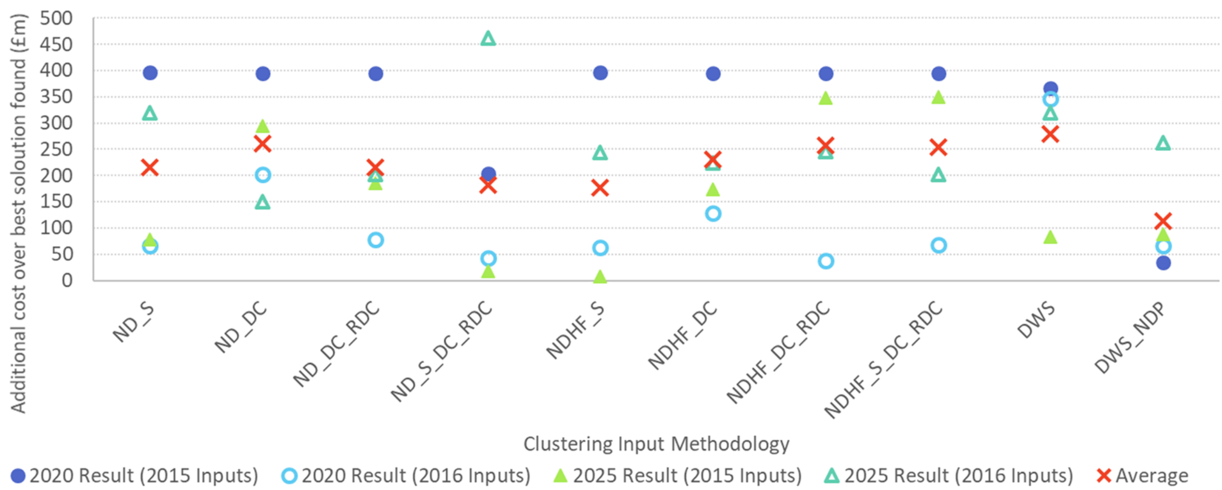

Fig. 8. Additional cost of expansion plan over best solution found for model year (2020 or 2025) and input data series year (2015 or 2016).

and oversimplified by the ramp duration curve. For example, a high ramp up may be easily met during an off-peak period but challenging close to the peak. However, without the unordered series, this difference is not captured in the clustering. Additionally, the number of consecutive ramping periods may cause different challenges for the GEP model and result in more starts than alternating ramp up and ramp down periods despite these being similar when converted to a duration curve. The addition of the future net demand series again significantly improves the results for the unordered series and duration curve series (NDHF_S and NDHF_DC), with reductions in the average additional cost over the best solutions found of 17.6% and 11.5% respectively. The differences between performances for the same model year with different datasets (2015 or 2016 data) demonstrates the importance of testing the selection against different inputs.

In Fig. 9, we compare the capacity expansion decisions for each technology depending on the representative day selection approach. The build plans for the 2020 model year are reasonably similar across net demand clustering approaches for the 2015 input data and, as shown in Fig. 8, perform poorly. A large improvement in cost is found in the DWS_NDP case where less capacity is built overall and there is a significantly higher reliance on storage than CCGT or peaking capacity. Looking at the 2020 model year, using the 2016 input data we see a similar result where reduction in overall capacity and increases in storage capacity tend to perform better. The addition of the ramp duration curve improves over the duration curve alone (ND_DC_RDS and ND_S_DC_RDS) in this way as does the inclusion of future net demand in the clustering (NDHF_S and NDHF_DC). The DWS approach leads to significantly lower build of capacity, in particular conventional capacity, and as seen, this results in a less optimal solution[9]. However, the inclusion of the peak net demand day (DWS_NDP) remedies this issue resulting in significantly higher capacity mostly in the form of storage. The expansion plan derived from this approach then has the highest reliance on storage and the lowest final costs.

In the 2025 model year, there is a more significant difference between the approaches. For the 2015 input data, the DWS and DWS_NDP series rely entirely on storage and perform relatively well in terms of cost as shown in Fig. 8, although the addition of a single CCGT in the NDHF_S case performs the best. The cases where a large amount of conventional capacity is built (with or without storage) perform poorly. Again, we see the DWS approaches have the highest reliance on storage technologies. With the 2016 input data there are relatively lower differences in the costs of the expansion plans, despite differences in the plans themselves indicating a relatively flat solution space where multiple different build approaches perform similarly.

3.2. Extreme point inclusion

In this section, we report results for the inclusion of additional 'days of interest' to attempt to capture the representation of future system challenges (see Table 4 for case descriptions) as opposed to the inclusion of ramp duration curves in the clustering. We focus on the three cases which provided the best results overall in the previous sections NDHF_S, NDHF_DC, and DWS_NDP.

3.2.1. Clustering results

The preselection of a cluster medoid to represent a cluster reduces the ability of the clustering algorithm to minimise the Euclidean distance (as it is now more constrained) although not greatly in most cases. For brevity these results are included as supplementary material.

3.2.2. Expansion planning model results

Fig. 10 compares the cost estimation errors with the forced inclusion of different 'days of interest' in the clustering, while Fig. 11 reports the average cost over the best solution found. Fig. 11 demonstrates the lowest costs are found where the DWS_NDP series has any of the days that include relatively low values of net demand (the minimum net demand value, and greatest differences between minimum and maximum net demand). However, this improvement does not seem robust as in these cases the ability of the GEP model to estimate costs is reduced, increasing the error in cost estimation two or more times, making it difficult to conclude that this change is an improvement. When examining the clustering based on ordered net demand series NDHF_S, we see both an improvement to cost estimation and to quality of the GEP solution over the full dataset in four cases. Three of these cases relate to ramping requirements ("max ramp down", "max consecutive ramp down", "max consecutive ramp up"). While these cases do not necessarily improve the representation of the ramp duration curve as a whole (as measured by the NRMSD) they all result in the inclusion of values near or at the low end of the ramp duration curve which is not well represented without these medoids. The other medoid that improves the solution is the min-max value distance which increases slightly the storage built in two of the four cases (and unchanged in the other two). This demonstrates the importance of extreme points, as while including them generally performs poorly on metrics relating to how close is the representative data to the underlying series (the distance metric and RMSD) they can drive more important dynamics in the GEP model. Effectively in this case, the

[9] This result may not hold in a model with an exogenous capacity target that is calculated completely independently of the modelled peak demand, as sufficient capacity would be built to meet this target.

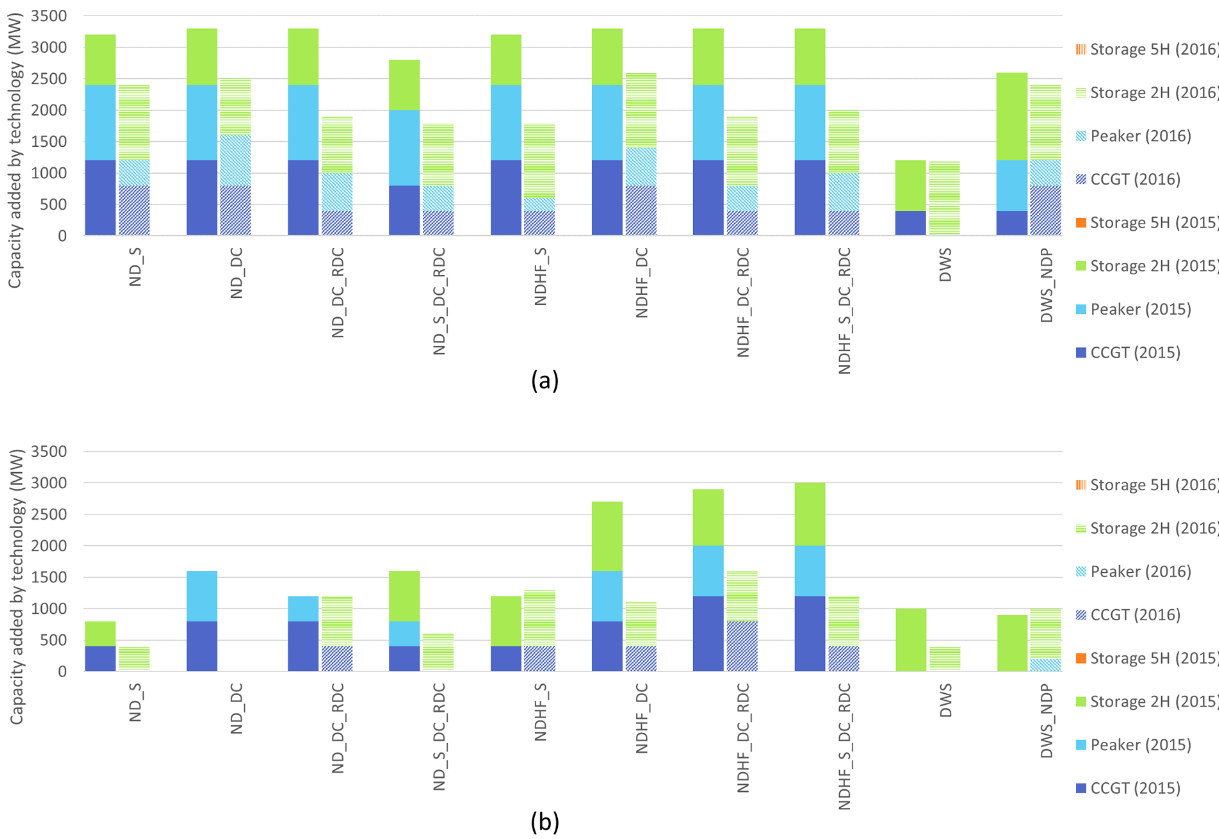

Fig. 9. Technology and capacity selected by GEP model depending on representative day selection approach and input dataset (2015 or 2016); (a) 2020 expansion model (b); 2025 expansion model.

Table 4
Key for case abbreviations.

Abbreviation	Data description
NDHF_S	Historical and future year net demand as original chronological series
NDHF_DC	Historical and future year net demand as duration curve used in clustering
DWS_NDP	Individual historical demand, wind capacity factor, and solar capacity factor included in clustering. We relax the inclusion of the peak wind day, peak demand day, and peak solar day and instead include the peak net demand day as a forced medoid
Abbreviation	**Forced inclusion of medoid description**
MinV	Forced inclusion of day with minimum value of net demand as medoid
GMMD	Forced inclusion of day with greatest difference between maximum and minimum net demand value as medoid
GMMD6H	Forced inclusion of day with greatest difference between maximum 6 h and minimum 6 h of net demand as medoid
MRU	Forced inclusion of day with maximum ramp up in net demand as medoid
MRD	Forced inclusion of day with maximum ramp down in net demand as medoid
MCRU	Forced inclusion of day with maximum consecutive ramp up in net demand as medoid
MCRD	Forced inclusion of day with maximum consecutive ramp down in net demand as medoid

inclusion of these dynamics is significantly more important for the GEP decisions than the clustering would consider.

Finally, we find one extreme point that improves the GEP model's ability to estimate costs and the cost of the expansion plans across all methods of representing input data, the inclusion of the maximum ramp down (MRD). While again this addition does not improve the ability of the selections to represent the ramp duration curve as a whole (as measured by the NRMSD) without this inclusion the extreme low ramps are particularly poorly represented.

In Fig. 12, we report the expansion plans where the maximum ramp down day is included in the clustering. As discussed, the greatest cost improvements are found in the two net demand data series (NDHF_S, NDHF_DC) in the 2020 model year with the 2015 dataset. In these cases, we see a relatively large 800 MW reduction in conventional capacity, replaced by 400 MW and 300 MW of storage respectively. In the individual series clustering (DWS) we see a less dramatic shift towards CGGT in the 2020 model year (400 MW increase in CCGT, 200 MW reduction in Peaker, 100 MW reduction in storage) and a slight reduction in CCGT capacity in 2025 (400 MW in one case), all associated with improvements of the cost of the solution. Overall, the best approach found (DWS_NDP MRD) on average builds less than half the conventional capacity and more than double the energy storage capacity compared to a model using representative days selected from historical net demand duration curves.

The maximum downwards ramp was not considered the most likely candidate to improve results and we explore the source of this result further. Firstly, taking the cases where the maximum downwards ramp is included, we relax the ramping restrictions in the GEP model and find that ramping constraints are tight enough to affect build decisions, particularly in the high renewable penetration 2025 year. Next, we

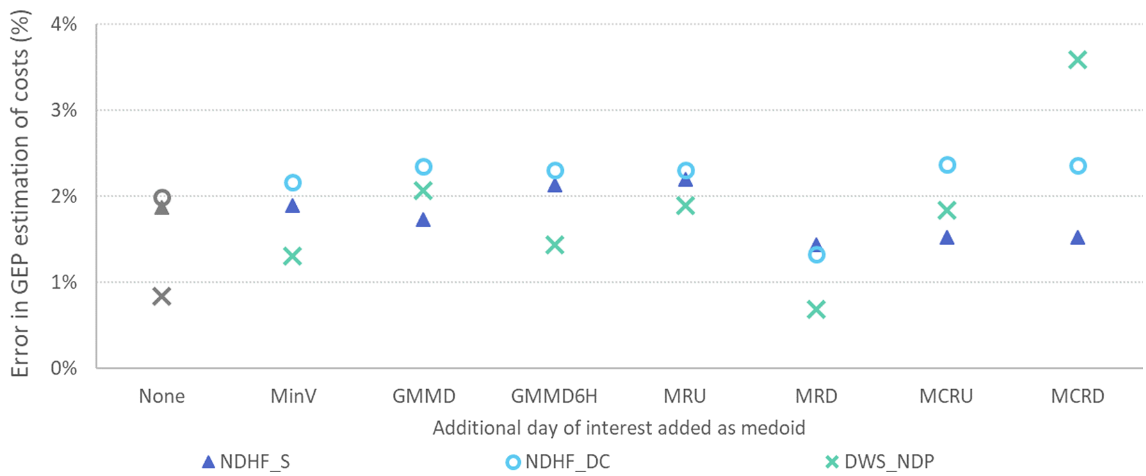

Fig. 10. Average difference in GEP model estimation of system costs and full dataset UC estimate of costs (% error), average over all input data series and model years.

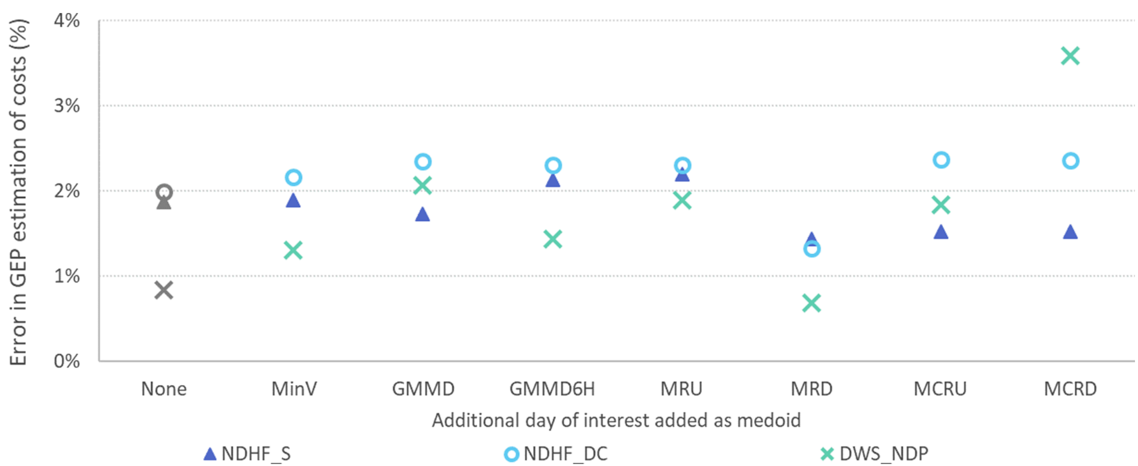

Fig. 11. Average additional cost of expansion plan over best solution found for model year (2020 or 2025), average over all input data series and model years.

assess the build plan found without ramping constraints in the full UC model (which feature lower levels of storage). We find that in the higher renewable penetration 2025 model years, the build plans determined considering ramping constraints perform significantly better in 5 of the 6 cases with an average reduction in additional costs relative to the best solution found of 35%. Finally, to ensure the changes in build are reflecting the need to meet ramping constraints, we run the full UC model without ramping constraints. We find, in this case, the build plans derived from the GEP model without ramp constraints to be lower cost, indicating the additional storage is only beneficial when ramping constraints are considered. Given the situational improvement of the inclusion of this medoid we recommend its inclusion is tested in any modelling exercise, as it may not be material where ramping and cycling challenges are less significant.

It is important to note that throughout this exercise the GEP model has selected the storage technology with a very small energy storage (two hours) and, in the few cases where the technology with more storage capacity was selected, the expansion plan performed poorly in the UC model over the full dataset. The fact that such a small storage capacity benefits the system so significantly indicates that the benefits are mostly due to its ability to address ramping challenges or peak period challenges as opposed to moving energy from off-peak to on-peak periods. This can additionally be seen when looking at an average generation and charge profile for the storage which typically discharges during the morning ramp up, charge during the middle of the day (where costs can be much higher than overnight), and again discharges for the evening peak. In some representative days the storage even discharges during the overnight periods which appears to help deal with volatility in wind production.

3.3. Assessment of scaling methodology

The method of adjusting clustering weights described in Section 2.3 has been used throughout to ensure that total and peak net demand are accurately represented in the model. In this section we assess this methodology by comparing to three alternatives: using clustering outputs without adjustment, scaling clustering to achieve annual total net demand, scaling clustering to achieve annual total net demand while maintaining the peak annual net demand as described in Section 2.3.

Firstly, we report, in Fig. 13, the size of the error in total net demand that results from using the days and weights selected by the clustering directly without any adjustment. We compare this value with the size of the scalars calculated by the optimisation to remove this error. Fig. 13 demonstrates that the clustering can make significant errors when only clustering on historical net demand data (ND_) and these errors increase as the penetration of renewables increases with the error, more than doubling between the 2020 and 2025 model years. The series that also include future net demand in the clustering (NDHF_) perform significantly better as is to be expected based upon the discussion in Section 2.2.2. When the clustering is based upon the underlying series

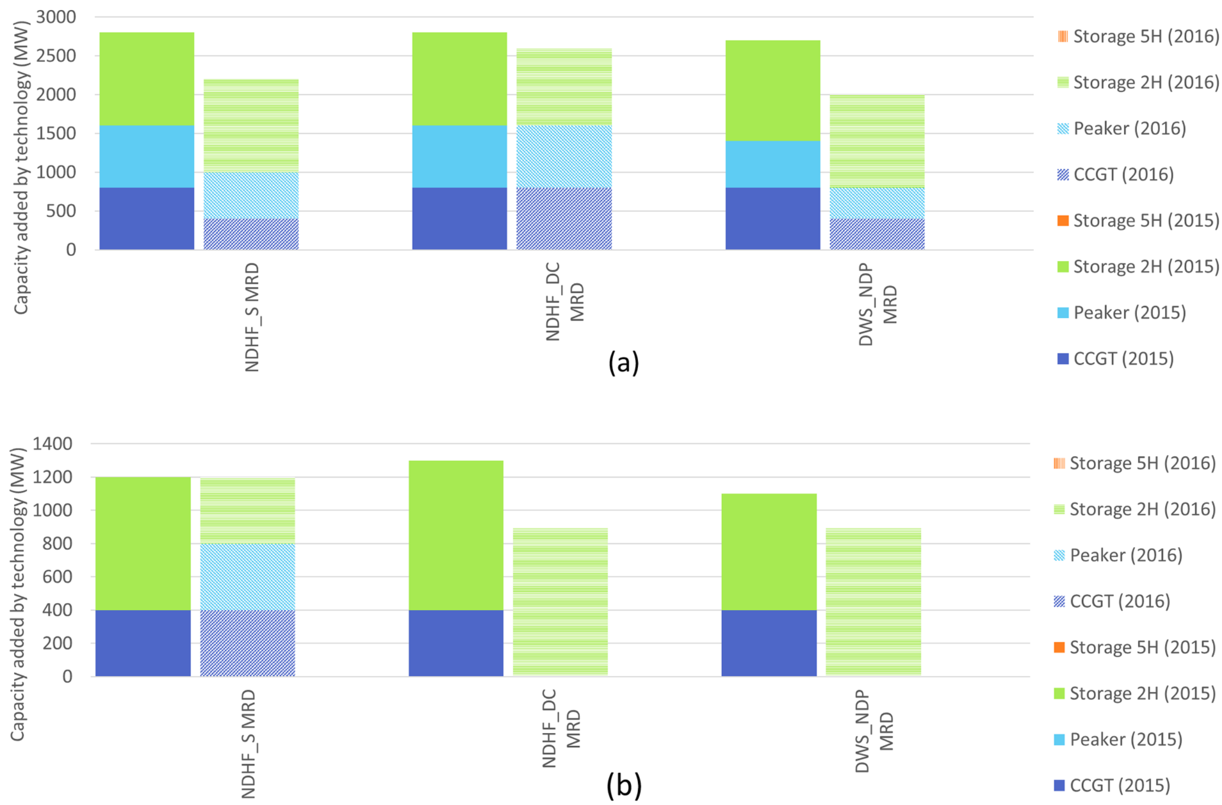

Fig. 12. Technology and capacity selected by GEP model depending on representative day selection approach, including identified days of interest, and input dataset (2015 or 2016); (a) 2020 expansion model; (b) 2025 expansion model.

(DWS and DWS_NDP) the errors are slightly higher than when future net demand is directly included (NDHF_) and are likely to be significant for the GEP model (ranging from 1.6% to 3.3%).

In Fig. 14 we report the results of using different approaches to remove the error in total net demand in the GEP model. In this assessment we fix the expansion plan decisions (the set of build decisions) for each case and assess the estimation of costs under each type of scaling and compare to costs estimated for that expansion plan over the full dataset. This method provides the most direct comparison between the different methodologies but likely underestimates the differences in using these methodologies in practice. The scaling methodology where the peak is not maintained is likely to lead to different build decisions, potentially without enough resources to meet peak demand, however that effect is not possible in this case where all build decisions are set. These differences would make it more difficult to compare between the approaches as they would be assessing the accuracy on different

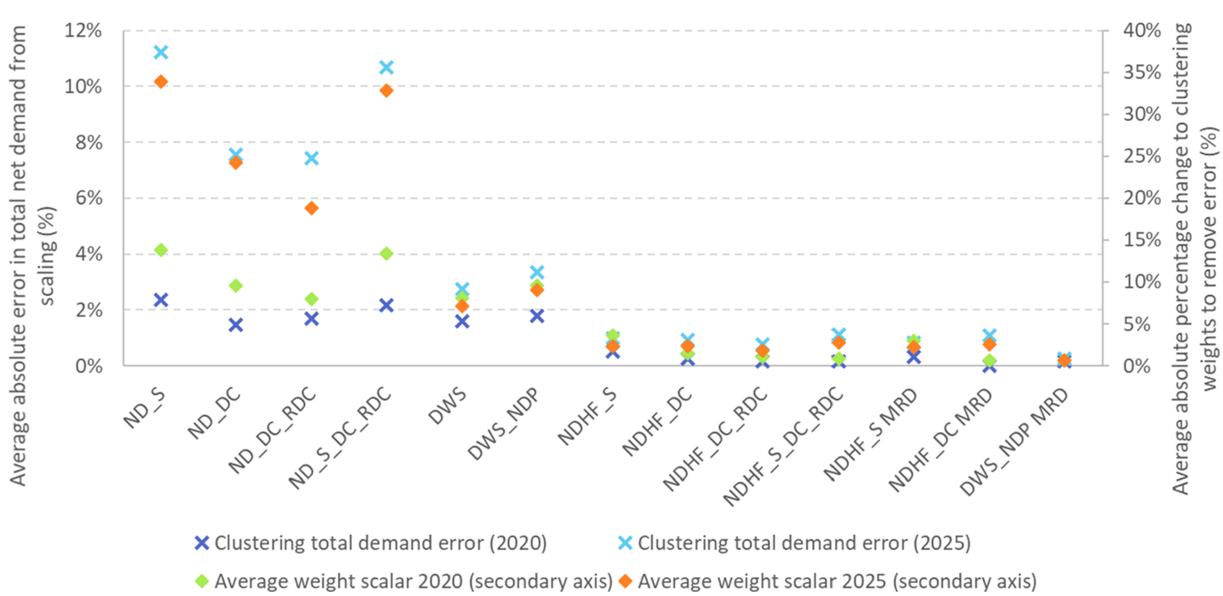

Fig. 13. Average absolute error in annual total net demand that results from using clustering outputs directly in model (%) and average absolute scalar calculated by weighting optimisation to remove error.

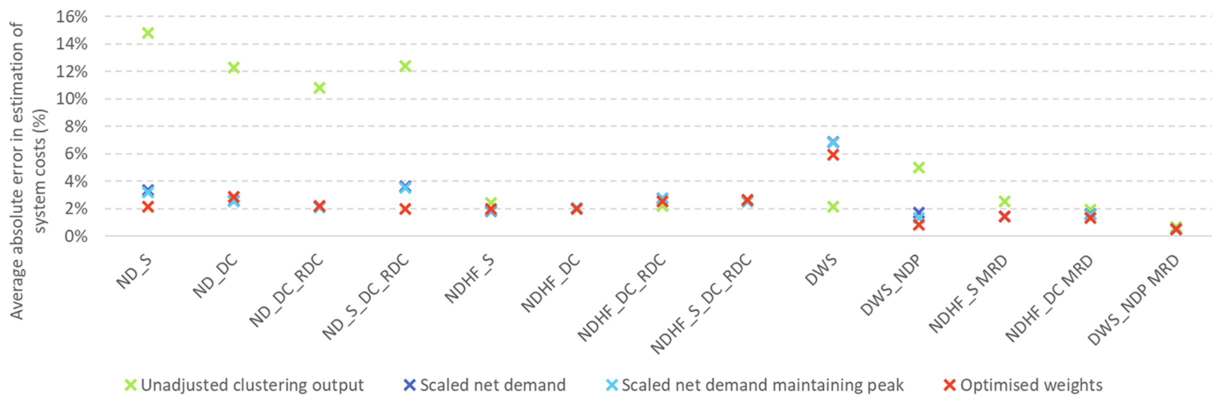

Fig. 14. Average absolute error in GEP model estimation of full year costs under different weighting methodologies.

systems. In this approach we are directly assessing the differences in the representation of net demand in a comparable system.

Fig. 14 demonstrates the importance of scaling with large increases in the GEP system cost estimation errors, where there are large errors in the clustering outputs as reported in Fig. 13. The weighting methodology provides a 60.5% improvement (from 5.5% to 2.2% on average) in the magnitude of errors made in the estimation of costs by the GEP model against using the clustering outputs directly. We find that the scaling maintaining peak method performs better than just scaling alone, however, that a 11.5% improvement over this approach is achieved with the weighting methodology on average. As expected, based on the discussion in Section 2.3, we find the error in GEP estimation of costs to be higher than the error in net demand. Additionally, we see that the benefits of scaling or adjusting weights are clearly proportional to the size of the error made by the clustering. Where the error in total demand is small, the benefits of removing this error will clearly be smaller. In the cases where future net demand is included in the clustering process (NDHF_) and errors in total demand are lower, the weighting methodology only performs slightly better than using the outputs directly, improving cost estimation errors 13.5% (from 2.30% to 1.99% average across cases) and is only 2.2% better than adjusting net demand while maintain peak. In the cases where we cluster on historical net demand and the error in total demand is high, the weighting methodology performs 81.7% better than the unadjusted inputs and 18.5% better than the scaling maintain peak. Importantly, the weighting methodology performs well compared to the other scaling methodologies in the cases where we cluster based upon the underlying data series (DWS_), improving cost estimation errors by 7.6% against the unadjusted cluster outputs and performing 15.6% better than the scaling methodology that maintains the peak. In the best performing case overall (DWS_NDP MRD) the weighting improves errors 28% against not scaling the outputs and 1% against scaling maintain peak, and in the best case without the inclusion of extreme points (DWS_NDP) it improves errors by 83% and 36%.

4. Conclusions

In this paper, we have demonstrated the importance of carefully performing the clustering of representative days for generation expansion planning. While representative day selections had the same peak load and total energy to be met in the generation expansion planning model, the generation expansion plans differed greatly among representative day selection methods in terms of both the total capacity to be added and the technology selected. The differences between the best and worst expansion plans range in cost between £350m and £690m (large in comparison to the modest investment in new units of £400m–£2bn).

Firstly, we addressed a typical issue with deriving representative days for generation expansion planning from a clustering process that results in the target level of net demand not being represented in the model. We presented a method for ensuring that the results from the clustering algorithm achieve this targeted level of net demand in the system without altering the underlying net demand shapes that are observed or predicted from historical data, an issue found with other approaches in the literature. This method resulted in an average error of the generation expansion planning model 61% lower compared to using the clustering outputs directly and a 12% improvement against alternative methodologies that scale net demand levels found in the literature.

For the clustering process, we found the largest improvement to the clustering when using net demand was gained by including both the future (high renewable penetration) and historical (low renewable penetration) net demand associated with each day in the clustering. When clustering solely on historical net demand, the algorithm is prone to group together days that are high load - high renewable production with days that are medium load - low renewable production days. While these are similar in the historical data set, as renewable capacity increases into the future, the days produce very different net load profiles. We found the clustering does not select days to represent the extreme low values of the net load duration curve without this approach.

We found additional gains by ensuring that days containing specific ramping features are included as cluster medoids. This approach performed better than attempting to add a ramp duration curve to the clustering. We found it is important to include the largest downwards ramp in all cases tested here. The large improvements to solution cost seen by the addition of this extreme point indicate that clustering alone cannot be relied upon to capture ramping dynamics and the contribution energy storage can make. We find, however, no consistent benefit to including several other potential extreme points such as the minimum value, which we expected to be important for driving curtailment costs.

Overall, we found the best results when clustering based on the individual demand, wind, and solar series as opposed to the chronological net demand series or the net demand duration curve based approach (or any combination of the two or ramp duration curves). This result only holds, however, when we include as medoids the days that feature the peak net demand period. This result is improved by the addition of the peak single period ramp down in net demand. In comparison to a simple clustering on historical net demand duration curves, this approach improves the accuracy of the estimation of costs by 71% on average (76% including the additional extreme point) and reduces the difference to the best solution found based on cost by 57% on average (65% including the additional extreme point).

We found important technology mix implications for the expansion plans selected by the GEP model based upon the different approaches to clustering input data. In most cases, improvements to solution optimality are gained where conventional capacity is substituted by storage capacity indicating that an oversimplification of future system challenges biases the model against energy storage. The best approach found on average builds less than half the conventional capacity and more than double the energy storage capacity, compared to a model using representative days selected from historical net demand duration curves. Based on the specific energy storage technology that was selected, we find it probable that this storage is beneficial for improving ramping costs as opposed to shifting large quantities of additional renewables between on- and off-peak periods. This finding is reinforced by the fact that the inclusion of medoids relating to ramping (particularly downwards ramping) improved the solution quality and led to a greater preference towards storage over conventional technologies.

The good overall performance of clustering based on underlying demand, wind, and solar series, combined with extreme points based on net demand indicates that there is promise in extending the methodology developed here to cases of endogenous renewable build. Future work of the authors intends to expand the weighting approach presented here to ensuring that renewable capacity factors are accurately modelled in representative days for endogenous renewable build modelling. In addition, we aim to explore how extreme points derived from the combination of underlying data series (demand, wind, solar) and intraday ramping dynamics can be represented in this case. Additionally, while this application focused on capturing the dynamics that would ensure that the role energy storage could play in meeting future power system challenges was accurately represented, this approach could be extended to ensuring that the representative days selected reflect the potential contribution of demand side response.

Acknowledgements

This work is supported by the Fundação para a Ciência e Tecnologia (FCT) through the MIT-Portugal Program and the FCT scholarship PD/BD/113715/2015 as well as the InnoEnergy PhD school. Thanks are also due to FCT through support from the IN+ Strategic Project UID/EEA/50009/2013. The authors are grateful to FCT and InnoEnergy for the financial support.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2019.113603.

References

[1] Frew BA, Becker S, Dvorak MJ, Andresen GB, Jacobson MZ. Flexibility mechanisms and pathways to a highly renewable US electricity future. Energy 2016;101. https://doi.org/10.1016/j.energy.2016.01.079.

[2] Oree V, Sayed Hassen SZ, Fleming PJ. Generation expansion planning optimisation with renewable energy integration: a review. Renew Sustain Energy Rev 2017;69:790–803. https://doi.org/10.1016/j.rser.2016.11.120.

[3] Koltsaklis NE, Georgiadis MC. A multi-period, multi-regional generation expansion planning model incorporating unit commitment constraints. Appl Energy 2015;158. https://doi.org/10.1016/j.apenergy.2015.08.054.

[4] Palmintier BS, Webster MD. Impact of operational flexibility on electricity generation planning with renewable and carbon targets. IEEE Trans Sustain Energy 2016;7. https://doi.org/10.1109/TSTE.2015.2498640.

[5] Pina A, Silva CA, Ferrão P. High-resolution modeling framework for planning electricity systems with high penetration of renewables. Appl Energy 2013;112. https://doi.org/10.1016/j.apenergy.2013.05.074.

[6] Haas J, Cebulla F, Cao K, Nowak W, Palma-behnke R, Rahmann C, et al. Challenges and trends of energy storage expansion planning for flexibility provision in low-carbon power systems – a review. Renew Sustain Energy Rev 2017;80:603–19. https://doi.org/10.1016/j.rser.2017.05.201.

[7] Koltsaklis NE, Dagoumas AS. State-of-the-art generation expansion planning: a review. Appl Energy 2018;230:563–89. https://doi.org/10.1016/j.apenergy.2018.08.087.

[8] Lund PD, Lindgren J, Mikkola J, Salpakari J. Review of energy system flexibility measures to enable high levels of variable renewable electricity. Renew Sustain Energy Rev 2015;45:785–807. https://doi.org/10.1016/j.rser.2015.01.057.

[9] Frew BA, Jacobson MZ. Temporal and spatial tradeoffs in power system modeling with assumptions about storage: an application of the POWER model. Energy 2016;117:198–213. https://doi.org/10.1016/j.energy.2016.10.074.

[10] Poncelet K, Delarue E, Six D, Duerinck J, D'haeseleer W. Impact of the level of temporal and operational detail in energy-system planning models. Appl Energy 2016;162. https://doi.org/10.1016/j.apenergy.2015.10.100.

[11] Jin S, Botterud A, Ryan SM. Temporal versus stochastic granularity in thermal generation capacity planning with wind power. IEEE Trans Power Syst 2014;29. https://doi.org/10.1109/TPWRS.2014.2299760.

[12] Merrick JH. On representation of temporal variability in electricity capacity planning models. Energy Econ 2016;59. https://doi.org/10.1016/j.eneco.2016.08.001.

[13] Pereira S, Ferreira P, Vaz AIF. Generation expansion planning with high share of renewables of variable output. Appl Energy 2017;190:1275–88. https://doi.org/10.1016/j.apenergy.2017.01.025.

[14] Blanford GJ, Merrick JH, Bistline JET, Young DT. Simulating annual variation in load, wind, and solar by representative hour selection. Energy J 2018;39:189–212.

[15] Ehrenmann a, Smeers Y. Generation capacity expansion in a risky environment: a stochastic equilibrium analysis. Oper Res 2011;59:1332–46. https://doi.org/10.1287/opre.1110.0992.

[16] Nahmmacher P, Schmid E, Hirth L, Knopf B. Carpe diem: A novel approach to select representative days for long-term power system modeling. Energy 2016;112:430–42. https://doi.org/10.1016/j.energy.2016.06.081.

[17] Pina A, Silva C, Ferrão P. Modeling hourly electricity dynamics for policy making in long-term scenarios. Energy Policy 2011;39. https://doi.org/10.1016/j.enpol.2011.06.062.

[18] Liu Y, Sioshansi R, Conejo AJ. Hierarchical clustering to find representative operating periods for capacity-expansion modeling. IEEE Trans Power Syst 2017;33:3029–39. https://doi.org/10.1109/TPWRS.2017.2746379.

[19] Ludig S, Haller M, Schmid E, Bauer N. Fluctuating renewables in a long-term climate change mitigation strategy. Energy 2011;36:6674–85. https://doi.org/10.1016/j.energy.2011.08.021.

[20] Koltsaklis NE, Dagoumas AS, Georgiadis MC, Papaioannou G, Dikaiakos C. A mid-term, market-based power systems planning model. Appl Energy 2016;179:17–35. https://doi.org/10.1016/j.apenergy.2016.06.070.

[21] Flores-Quiroz A, Palma-Behnke R, Zakeri G, Moreno R. A column generation approach for solving generation expansion planning problems with high renewable energy penetration. Electr Power Syst Res 2016;136. https://doi.org/10.1016/j.epsr.2016.02.011.

[22] Sisternes FJ De, Webster MD. Optimal selection of sample weeks for approximating the net load in generation planning problems. ESD Work Pap Ser; 2013.

[23] Pineda S, Morales JM. Chronological time-period clustering for optimal capacity expansion planning with storage. IEEE Trans Power Syst 2018;8950:1–9. https://doi.org/10.1109/TPWRS.2018.2842093.

[24] ElNozahy MS, Salama MMA, Seethapathy R. A probabilistic load modelling approach using clustering algorithms. 2013 IEEE Power Energy Soc Gen Meet 2013. p. 1–5. https://doi.org/10.1109/PESMG.2013.6672073.

[25] Härtel P, Kristiansen M, Korpås M. Assessing the impact of sampling and clustering techniques on offshore grid expansion planning. Energy Proc. 2017;137:152–61. https://doi.org/10.1016/j.egypro.2017.10.342.

[26] Poncelet K, Hoschle H, Delarue E, Virag A, D'haeseleer W. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. IEEE Trans Power Syst 2016;PP. https://doi.org/10.1109/TPWRS.2016.2596803.

[27] Green R, Staffell I, Vasilakos N. Divide and Conquer? k-means clustering of demand data allows rapid and accurate simulations of the British electricity system. IEEE Trans Eng Manage 2014;61:251–60. https://doi.org/10.1109/TEM.2013.2284386.

[28] Kaufman L, Rousseeuw PJ. Partitioning around medoids. Find Groups Data 1990:68–125. https://doi.org/10.1002/9780470316801.ch1.

[29] Kristiansen M, Korp M. Sensitivity analysis of sampling and clustering techniques in expansion planning models; 2017. p. 609795. https://doi.org/10.1109/EEEIC.2017.7977727.

[30] Kotzur L, Markewitz P, Robinius M, Stolten D. Impact of different time series aggregation methods on optimal energy system design. Renew energy 2018;117:474–87.

[31] Krad I, Gao DW, Member S. An assessment of flexibility reserves in stochastic modeling at multiple timescales. CSEE J Power Energy Syst 2017;3:84–92. https://doi.org/10.17775/CSEEJPES.2017.0011.

[32] Mallapragada DS, Papageorgiou DJ, Venkatesh A, Lara CL, Grossmann IE. Impact of model resolution on scenario outcomes for electricity sector system expansion. Energy 2018;163:1231–44. https://doi.org/10.1016/j.energy.2018.08.015.

[33] Pfenninger S. Dealing with multiple decades of hourly wind and PV time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Appl Energy 2017;197:1–13. https://doi.org/10.1016/j.apenergy.2017.03.051.

[34] Poncelet K, Van Stiphout A, Meus J, Delarue E, D'haeseleer W. LUSYM Invest: a generation expansion planning model with a high level of temporal and technical detail; 2018.

[35] Jenkins J, Sepulveda N. Enhanced decision support for a changing electricity landscape: the GenX configurable electricity resource capacity expansion model. MIT Energy Initiat Work Pap; 2017. p. 1–40.

[36] Palmintier B, Webster M. Impact of unit commitment constraints on generation expansion planning with renewables. IEEE Power Energy Soc Gen Meet 2011. p. 1–7. https://doi.org/10.1109/PES.2011.6038963.

[37] Wierzbowski M, Lyzwa W, Musial I. MILP model for long-term energy mix planning with consideration of power system reserves. Appl Energy 2016;169. https://doi.org/10.1016/j.apenergy.2016.02.003.

[38] Deane JP, Drayton G, Gallachóir BPÓ. The impact of sub-hourly modelling in power systems with significant levels of renewable generation. Appl Energy 2014;113:152–8. https://doi.org/10.1016/j.apenergy.2013.07.027.

[39] National Grid. UK future energy scenarios. Warwick, UK; 2016. http://fes.nationalgrid.com/media/1292/2016-fes.pdf.

[40] Heuberger CF, Rubin ES, Staffell I, Shah N, Mac Dowell N. Power capacity expansion planning considering endogenous technology cost learning. Appl Energy 2017;204:831–45. https://doi.org/10.1016/j.apenergy.2017.07.075.