District energy system optimisation under uncertain demand: Handling data-driven stochastic profiles

B. Pickering*, R. Choudhary

Department of Engineering, University of Cambridge, Cambridge, United Kingdom

HIGHLIGHTS

- We present a new 3-step method for handling demand uncertainty in district energy optimisation.
- Temporally autocorrelated demand scenarios are generated using real metered demand.
- We apply the 3-step methodology to two case study districts: Bangalore, India and Cambridge, UK.
- Cost optimal districts favour building-scale technologies to centralised energy generation.
- Unmet demand in out-of-sample scenarios is reduced 25 to 50 times when using scenario optimisation.

ARTICLE INFO

ABSTRACT

Current district energy optimisation depends on perfect foresight. However, we rarely know how the future will transpire when undertaking infrastructure planning. A key uncertainty that has yet to be studied in this context is building-level energy demand. Energy demand varies stochastically on a daily basis, owing to activities and weather. Yet, most current district optimisation models consider only the average demand. Studies that incorporate demand uncertainty ignore the temporal autocorrelation of energy demand, or require a detailed engineering model for which there is no validation against real consumption data. In this paper, we propose a new 3-step methodology for handling demand uncertainty in mixed integer linear programming models of district energy systems. The three steps are: scenario generation, scenario reduction, and scenario optimisation. Our proposed framework is data-centric, based on sampling of historic demand data using multidimensional search spaces. 500 scenarios are generated from the historical demand of multiple buildings, requiring historical data to be nonparametrically sampled whilst maintaining interdependence of hourly demand in a day. Using scenario reduction, we are able to select a subset of scenarios that best represent the probability distribution of our large number of initial scenarios. The scenario optimisation step constitutes minimising the cost of technology investment and operation, where all realisations of demand from the reduced scenarios are probabilistically weighted in the objective function. We applied these three steps to a real district development in Cambridge, UK, and an illustrative district in Bangalore, India.

Our results show that the technology investment portfolios derived from our 3-step methodology are more robust in meeting large possible variations in demand than any model optimised independently with a single demand scenario. This increased robustness comes at a higher monetary cost of investment. However, the high investment cost is lower than the highest possible cost when each of the initial 500 scenarios is optimised independently. In both our case studies, building level energy systems are always more robust than district level ones, a result which disagrees with many existing studies. The outcomes enable better examination of district energy systems. In addition, our methodology is compiled as an open-source code that can be applied to optimise existing and future energy masterplans of districts.

1. Introduction

District energy systems are becoming more prevalent as a method of meeting building energy demand. Past work has shown that a district energy network gives better overall utility from energy systems than building-level solutions. Depending on the district under study,

---

* Corresponding author.
  E-mail addresses: bp325@cam.ac.uk (B. Pickering), rc488@cam.ac.uk (R. Choudhary).

calculated savings vary. Carbon emissions could fall by 23% [1], 44% [2], or 50% [3]. They could also more than double, albeit with a reduction of 74% in system cost, as reported by Jennings et al. [4] and Mehleri et al. [5]. Some studies disagree with this trend, however, suggesting only minor savings of a few percent in cost [6] or even a slight increase [7].

The positive and negative impacts of district systems as a replacement for building-level technologies are case specific and must be modelled accordingly. Finding an optimal district system depends on how the problem is mathematically formulated, including realistic representation of the demand under which the energy system will operate. Indeed, optimal design of district energy systems requires a good knowledge of the spatio-temporal variations of the energy demand of buildings by their end-use. If the district consists of existing buildings, historical demand data may be used directly [6,4]. Where no historical data exists, such as for new buildings, demand can be simulated [1,2]. If insufficient data is available, demand can be treated as exogenous. Archetypal or reference demand profiles from relevant buildings are thus used as representative of future energy demand for the district [8–10,3,5]. In all such cases, possible future variations in demand are not considered. Meanwhile, differences between demand assumed at the modelling stage and actual annual energy demand have been reported to be anywhere between 16% and 500% in UK commercial properties [11]. In the case of district energy systems, unexpected variations can result in failure to meet demand.

In Mixed integer linear programming (MILP), uncertainty is commonly expressed by distinct scenarios which are either optimised independently (e.g. sensitivity analysis) or optimised in parallel using scenario optimisation[1] (SO). When making investment decisions under operational uncertainty, SO allows a modeller to optimise technology investment such that the system is more robust to each scenario realisation.

There are numerous ways through which a modeller can generate distinct energy demand scenarios: physics-led, data-driven, or assuming some level of variance from average demand. The latter is the easiest to incorporate. Studies which have considered demand variations from the average include variations of ±10% [12] to ±30% [13]. Understandably, the choice of variance has been shown to significantly influence optimisation results [14], and thus cannot be applied without sufficient validation. On the other hand, physics-led (or engineering) models require a detailed understanding of the building fabric, external conditions, and occupancy to simulate demand from the bottom-up. Future scenarios can be generated through the physics-led model by Monte Carlo simulations. In each simulation, input parameters such as material properties, ventilation rates, appliance energy consumption, and building occupancy are varied stochastically [15,16]. However, specifying the uncertainty surrounding each model parameter can be onerous [17], especially when a number of different types of buildings are included in a district [18].

In contrast, top-down data-driven models use historical consumption data to identify inherent variability in energy demand. They implicitly consider all influencing parameters (predictors) such as weather, building fabric, and occupancy. If a sufficient amount of data is available, it is possible to understand the impact of each predictor by casting the data into statistical models [19,20,17]. If there is insufficient data, the reliability of statistical models to represent the relationships between predictors and energy demand can be limited. Alternatively, it is possible to sample historical demand data directly. In such cases, the emphasis is less on what *causes* variations in demand and more in capturing the variations in energy demand across different

days. With measured data from an office building, Gamou et al. [21] stated that energy demand in any given hour can be described by a normal distribution, where there is a 95% probability of samples remaining within ±20% of the mean. This distribution has since been used to generate demand scenarios for optimising energy systems of hospitals [22,23].

A common shortcoming across current data-driven approaches for scenario generation is that temporal autocorrelation of energy demand is not considered. It is evident that energy demand at any particular hour is influenced by the demand in other hours of the day. The uncertainty described by scenarios without considering the temporal autocorrelation of demand can thus be misleading, and result in incorrect system design. Mavromatidis et al. [15] identified this as a key issue, which was solved in their study by use of a bottom-up physics-led model for generating demand scenarios. Furthermore, all current approaches consider uncertainty at any given time to be normally distributed around the mean which can result in under or overestimation of demand uncertainty.

In this paper, we present a new 3-step methodology to handle demand uncertainty in district-scale energy optimisation. Our proposed methodology is data-driven and hence, unlike detailed bottom-up engineering models, scalable for application to large districts. We overcome the shortcoming of current data-driven models by sampling multivariate nonparametric representations of historical demand, thus accounting for temporal autocorrelation *and* skewness of demand data around the mean value at any given hour. The methodology we introduce is made openly available online, and therefore, in addition to being novel, it is also reproducible and extendable by others to practical district-scale planning problems.

In the following section we describe the principal steps of our proposed method: (1) Data-driven scenario generation, (2) Scenario reduction, and (3) Scenario optimisation. In the rest of the paper, we demonstrate this through two illustrative studies: one in Bangalore, India, and the second in Cambridge, UK. Through these, we examine the impact of demand uncertainty on the design of district energy systems. Finally, we test the robustness of our systems to unseen realisations of future energy demand, using 'out of sample' scenarios.

2. Methodology

2.1. Scenario generation

Particularly at a masterplanning level, little may be known about buildings within a district other than their intended use and floor area. Using high-resolution historical data of other buildings representing similar use, a multidimensional search space can be created to describe the possible demand profiles for the district. To create a search space from the available data, we borrow from machine learning by considering 'features' and 'observations'. Features are the individual measurable properties that are being observed. Observations are the existing data describing distinct instances of those features. Features are the consumption values in each hour of a day, which we have observed for all historical days for which we have data. Over one year, there would be 24 features and 365 observations.

Clustering of observations into independent search spaces can help ensure our samples are more realistic. For example, we know that weekend and weekday electricity consumption will differ in an office space. Seasons, academic term times, and months are all subjective typical day clusters that could be chosen. There may be other, unknown metrics by which observations can be clustered programmatically, such as by K-means and hierarchical clustering [24]. The advantage of subjective clustering is the ability to map those clusters into future years, which we cannot readily do with clustering algorithms. Typical days are commonly used in mixed integer linear programming (MILP) optimisation, reducing the length of the time dimension from its full scale (e.g. one year - 8760 h) to anything from three [5] to six [3], or

---

[1] 'Scenario' and 'stochastic' are often used interchangeably to describe the same optimisation procedure. As stochastic optimisation may refer to other methods of handling stochastic variables, we make the clear distinction by describing our method as scenario optimisation.
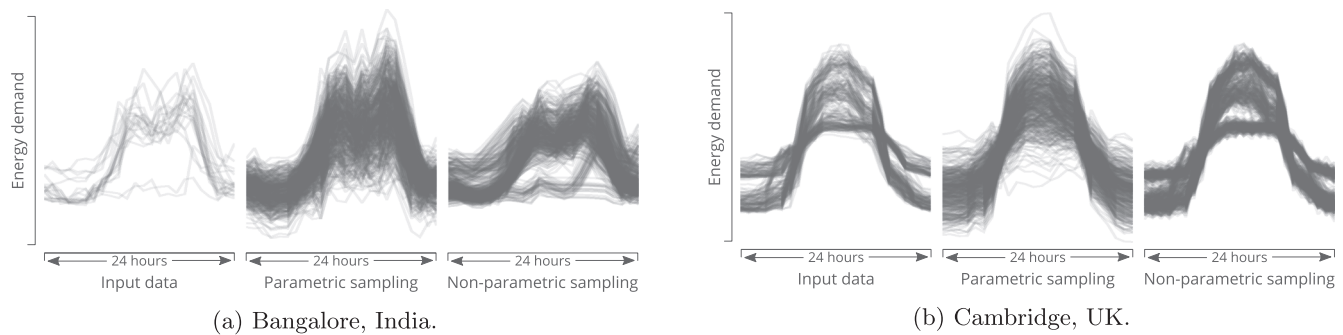
Fig. 1. Comparison of parametric and nonparametric sampling methods as a means to produce daily profiles for demand. Profiles are at an hourly resolution and 500 profiles of the input data were drawn from each of the sampling methods.

seven [25] days. Clustered observations form search spaces which represent 'typical' days in the year.

The shape of a search space depends not only on the set (or subset) of observations, but also the method by which we interpolate to create a continuous surface from discrete data. Previous studies have considered well-conditioned demand, which can be described by a multivariate normal curve [26,27]. However, daily demand profile sets generally do not fit a perfect Gaussian profile. This is especially the case when observations are acquired from multiple buildings within a single archetype. There can be sub-clusters of demand profiles with various local peaks in the distribution. Indeed, our initial analysis shows that demand is not well conditioned; i.e. it cannot be well described by a parametric distribution. From Fig. 1 we can see that, with parametric sampling (assuming multivariate normal distributions), clear clusters of profiles are lost. The symmetrical nature of multivariate normal sampling has also led to areas of high profile density to become the sample mean, not just the mode.

Nonparametric sampling can be more representative of the demand. However, as we make no assumptions about the shape of the input data, the search spaces resulting from nonparametric sampling are heavily dependent on the input data. One method to overcome this is to have large training and validation sets to tune the relevant hyperparameters. We discuss here two approaches in which a smaller data set may still be valid: multi-building data sets and functional principal component analysis (fPCA).

If we do not wish to replicate the demand profiles of one building when sampling a search space, we can sample from data describing multiple buildings within an archetype. Then, when sampling for an individual building, stochastic profiles will be unlikely to duplicate any one of the input profile sets. When little is known about a building other than its archetype, there is validity in this approach. The demand in an archetypal building could replicate the profiles in any of the input data buildings in that archetype. In fact, combining data on multiple buildings was used by the masterplanners when assigning archetypal demand to the Cambridge case study district [28]. As data from multiple buildings inform the search space for a single archetype, when two consecutive days are sampled from the profiles of different buildings in the same archetype, wildly varying demand may be observed from one day to the next. Much like the intraday temporal autocorrelation of demand, there clearly should also be some interday autocorrelation. To account for this, before stochastically sampling profiles, energy intensity for each input building was normalised by the maximum demand recorded for that building. After sampling, modelled buildings had their demand scaled by a randomly assigned normalisation factor from all those available for their archetype.

Without multiple buildings as input data, another approach must be taken to ensure samples are representative of the archetype and not only the input building. In fPCA, we can do this since demand for a typical day is treated as a function of time, as opposed to being discrete data points [29]. It provides a mathematical definition of the shape of
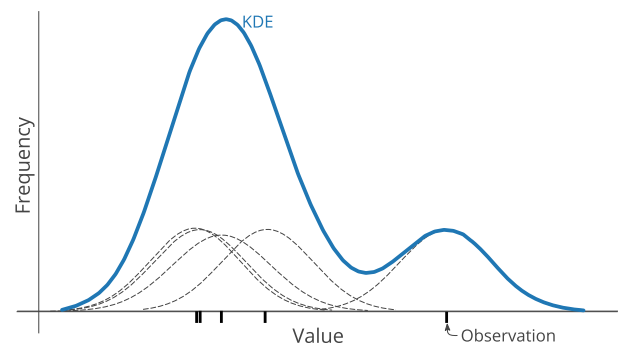

Fig. 2. Example of KDE, applied to a single feature observation set. A gaussian kernel has been applied to each observation in this case.

the curve in terms of a number of functional Principal Components (PCs) which are the same for all the data samples and describe particular features of the data. Associated with each PC which is a set of weightings, or 'scores', particular to each day of data and end-use. The scores describe mathematically the contribution of each PC to the overall day's demand profile, per end-use. A search space is then created for each typical day by using the set of scores associated with them [30–32]. Thus, we generate demand profiles that retain the primary *sense* of the input data without it being exactly reproduced.

In both methods of preparing the input data, we can use kernel density estimation (KDE) to create a probability density function (PDF). In KDE, a kernel (e.g. normal distribution) is applied to each feature in the observations, and the overlay of all these individual kernels represents the full data set. Fig. 2 represents this with one feature[2], which is a nonparametric distribution due to its lack of a single peak. There are two important hyper-parameters which dictate the efficacy of KDE: the kernel and the bandwidth. The kernel is the shape of the density function applied to each observation when constructing the full PDF. It is standard to use a Gaussian kernel (as used in Fig. 2), but 'top hat', 'triangular', and 'Epanechnikov' are among other kernel choices [33]. The bandwidth is the scale of smoothness applied to each kernel, akin to the standard deviation of a normal distribution. if bandwidth $= 0$, the resulting PDF will have non-zero values only at points corresponding to the input data. As the bandwidth increases to infinity, the PDF converges on a uniform distribution, with infinite variance. We do not wish to sample from either of these extreme cases. Instead, we choose the lowest possible bandwidth that fits a training data set and can reproduce an independent validation data set on sampling. If there are too few observations for training and validation, we can use *k*-fold

---

[2] The KDE of our 24 features, for each hour in the day, would form a 24-dimensional probability distribution function, which is impossible to visualise.

cross-validation in bandwidth and kernel selection. In *k*-fold cross-validation, the full data set is randomly partitioned into *k* subsets. One subset is retained for validation while the remaining $k-1$ sets are used for training. The process is repeated *k* times, such that all subsets are used for validation and training [34].

In Fig. 1 we can see the impact of multivariate nonparametric sampling using multi-building sets (Cambridge) and fPCA (Bangalore), both sampled from a KDE-generated PDF. fPCA samples were generated by Ward et al. [32] while multi-building sets were prepared in this study. The input data is better represented by nonparametric, rather than parametric sampling; The demand profiles cover a wider space beyond the range of the input data, while sub-clusters across the data are still apparent.

Using fPCA for the Bangalore study and multi-building sets for the Cambridge study, we generate 500 demand profiles per typical day, per end-use, and per building archetype in kWh/m^2. Annual hourly demand scenarios per building are sampled from these stochastic profiles for a reference year (Cambridge: 2015, Bangalore: 2016), such that no profile is duplicated between days in the year, or between buildings in the district. A summary of the 500 scenarios is given in Fig. 6 when discussing the results of applying this method.

2.2. Scenario reduction

A large number of probabilistic demand scenarios per building can result in intractability of a district energy optimisation model. Accordingly, the selection of the 'right' subset of scenarios becomes an important step, especially in scenario optimisation models. This subset of scenarios should be representative of the variations across the larger scenario set without requiring the whole set to be included in the optimisation model. Conejo et al. [35] proposed the use of the fast-forward algorithm to reduce the number of scenarios. Two primary variants of the methods are proposed, both of which aim to reduce the *Kantorovich distance*.[3] Both variants apply a cost metric to each scenario *s* in the scenario set *S*, from which a subset *S′* is chosen based on the minimisation of the difference in the probability distributions describing the costs in *S* and *S′*.

The two variants of scenario reduction proposed by Conejo et al. [35] differ on the cost metric applied to each scenario. In the first, a key performance indicator (KPI) describing the scenarios is selected. This might be the maximum hourly demand per scenario or the total demand over the entire year. The KPI is selected subjectively as a measure that has the biggest impact on the objective function. The first variant is considered less computationally intensive to apply and has been used for scenario reduction in existing SO studies [37,38]. The second variant requires that a non-probabilistic optimisation model is run for each scenario independently. These independent models are a formulation of the SO model which do not consider uncertainty. They are relatively fast to solve and can be run in parallel on a high-performance cluster in a matter of minutes. The objective function value calculated for each scenario is then used in the Kantorovich distance calculation.

A refinement of the second variant, proposed by Bruninx and Delarue [39], is used in this study. Conejo et al. [35] only considered optimisation of operation costs, fixing the investment cost for each of the independent models. Bruninx and Delarue [39] included investment costs in addition to operation costs. All decision variables are therefore part of the independent model optimisation, but binary and integer constraints are not included for computational efficiency. In our independent models, we keep the binary 'purchase' constraints applied to investment decisions. Because we use typical days as against the full time series of annual demand, each independent optimisation runs within a reasonable solution time (*O(100s)* on a high-performance computing cluster).

The process for scenario reduction can be thus summarised as follows:

1. Optimise the objective function for each scenario in parallel, minimising system cost (investment and operation) for each case independently.
2. Select 16 scenarios to represent the 500 input scenarios, by minimising the Kantorovich distance between the probability density of their objective function values to that of the full scenario set.
3. Assign each scenario in the full set to the closest (by probability distance) of the 16 scenarios in the reduced subset, weighting each reduced scenario by the number of scenarios it represents.

The result of this process is a scenario subset that can be used for tractable scenario optimisation. It is applied in the same manner for both Bangalore and Cambridge case studies. The selected reduced scenarios are detailed alongside results for SG and SO, in Section 4.

2.3. Scenario optimisation

Once reduced scenarios are derived, the uncertainty described by these scenarios can be dealt with by scenario optimisation (SO). Our SO model has two stages. The first involves finding the optimal technologies and their capacities, irrespective of their ability to meet variability in demand. In the second stage, the optimal technologies are reassessed for their ability to meet variability in energy demand represented by the 16 reduced scenarios. If energy demand represented by any particular scenario is not met, a financial penalty is incurred. The impact of a single scenario is weighted by its probability of occurrence, such that low probability scenarios may have unmet demand without incurring a large penalty on the overall objective function. Having unmet demand is a risk, which we can choose to not allow (risk-neutral SO) or to monetise for direct application to the objective function (risk-averse SO). In this study, we consider a risk-neutral SO to examine the amount of unmet demand resulting from SO as against deterministic optimisation. As is standard for district energy optimisation models, we use MILP as the optimisation technique. However, other optimisation techniques can be applied within our proposed SO framework. The risk-neutral SO objective function is shown below in Eq. (1), where *S′* refers to the reduced scenario subset detailed in Section 2.2. It includes both initial investments as well as the operational cost of technologies and is a variant of objective function used by Maurovich-Horvat et al. [40].

$$\min \quad \mathbf{cost_{invest}} + \sum_{s' \in S'} P_{s'} \mathbf{cost_{operate}}_{s'} \tag{1}$$

2.4. Model formulation

The objective function in Eq. (1) is subject to various constraints typical to energy system models, all of which are formulated within the open-source MILP modelling framework Calliope v0.6.1 [41]. These are:

- Limit the maximum possible capacity of any technology;
- Limit the production of any technology to its capacity;
- Link production to consumption of energy carriers for any technology, considering efficiency losses in the process;
- Link locations by distribution lines, to allow carrier flow;
- Link the storage in one timestep to the storage in the previous timestep, accounting for standing losses; and
- Ensure all demand is met at all locations in the network.

Additionally, we consider:

- Binary purchase constraints, in which a technology has a fixed purchase cost associated with *any* non-zero capacity. This combines

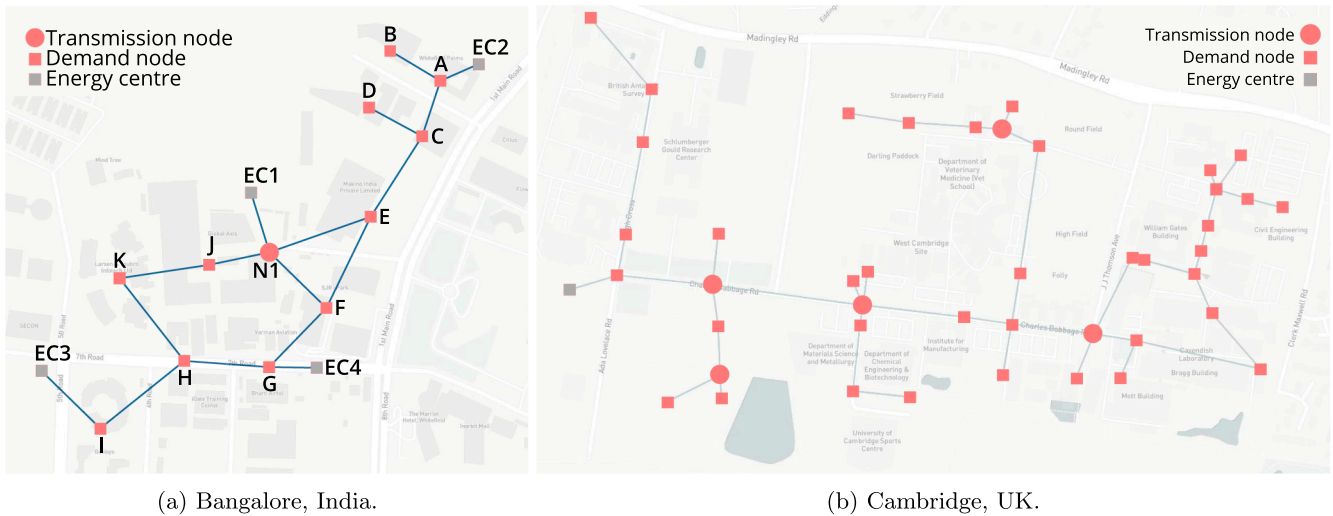[3] For a detailed mathematical formulation, readers are referred to [35,36].
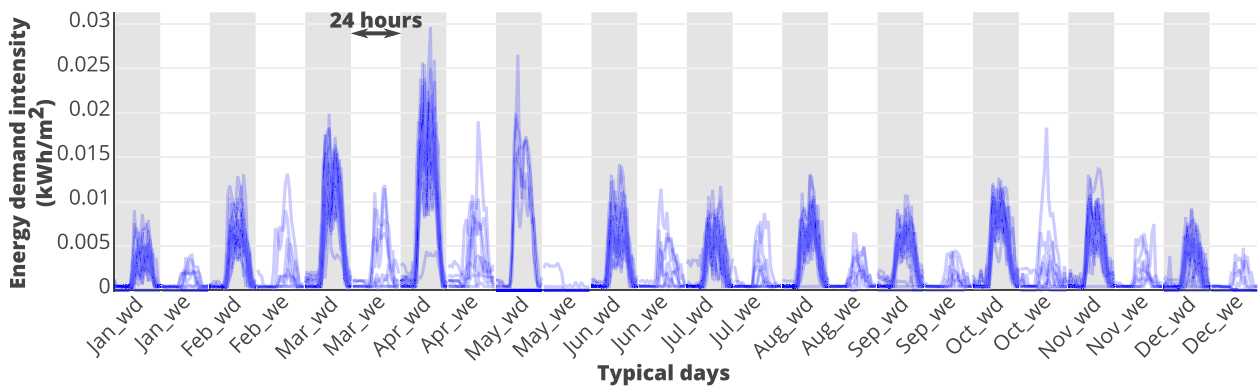
(a) Bangalore, India.

(b) Cambridge, UK.

Fig. 3. Case study districts.

Table 1
Bangalore district node details.

| Node	A	B	C	D	E	F	G	H	I	J	K	EC1	EC2	EC3	EC4	N1
GIA (m²)	5,440	36,586	12,650	22,400	17,184	78,086	46,582	93,064	23,846	178,496	39,504			N/A		N/A
Roof area (m²)	2,720	6,098	3,162	5,600	8,592	11,155	11,646	18,613	5,962	22,312	7,900					
Technologies						PV, AC, DG, GridE							ECh, CCHP, StoreT, GridE			
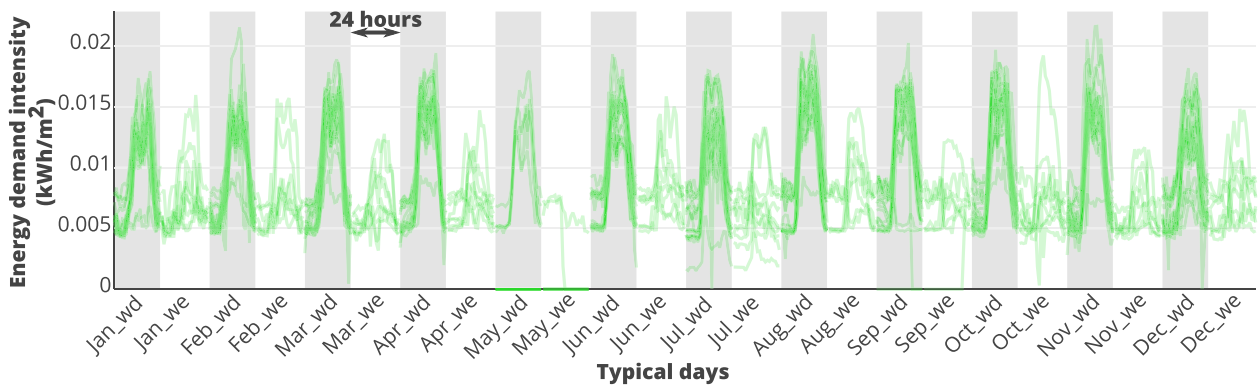
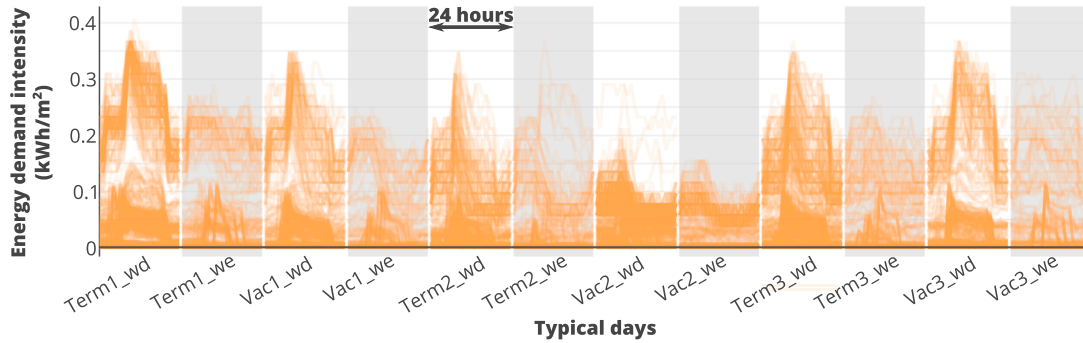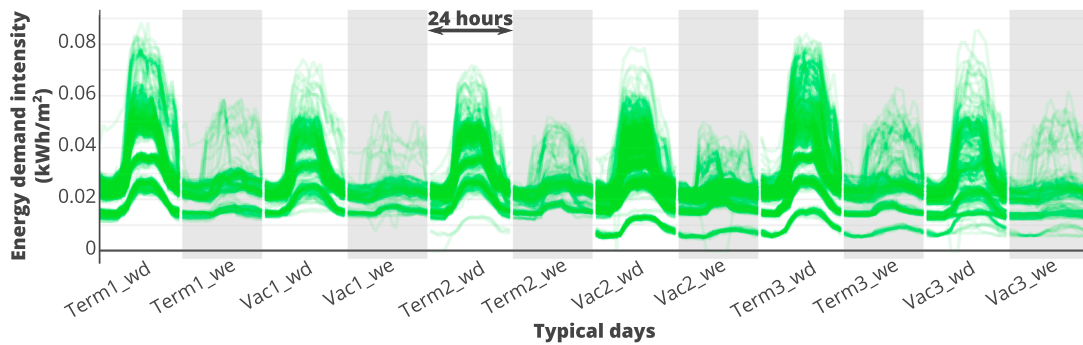(a) Energy type: electricity.

(b) Energy type: cooling.

Fig. 4. Bangalore case study input demand profiles, grouped by typical day, and energy type. Profiles have low opacity such that darker sections indicate significant profile overlap. '_we' = weekend, '_wd' = weekday.
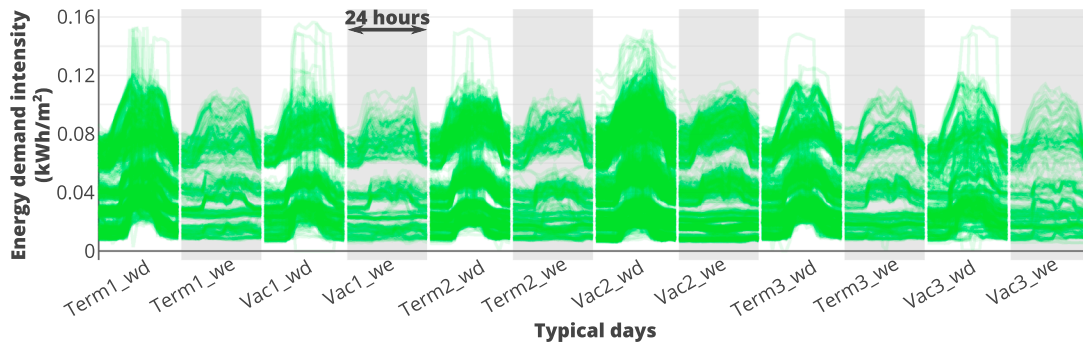
(a) Archetype: lab_research; energy type: electricity.

(b) Archetype: lab_research; energy type: gas.

(c) Archetype: desk_research; energy type: electricity.

(d) Archetype: desk_research; energy type: gas.

Fig. 5. Cambridge case study input demand profiles, grouped by typical day, archetype, and energy type. Profiles have low opacity such that darker sections indicate significant profile overlap. See Table 3 for typical day dates and Table 2 for buildings per archetype. '_we' w̄eekend, '_wd' w̄eekday.
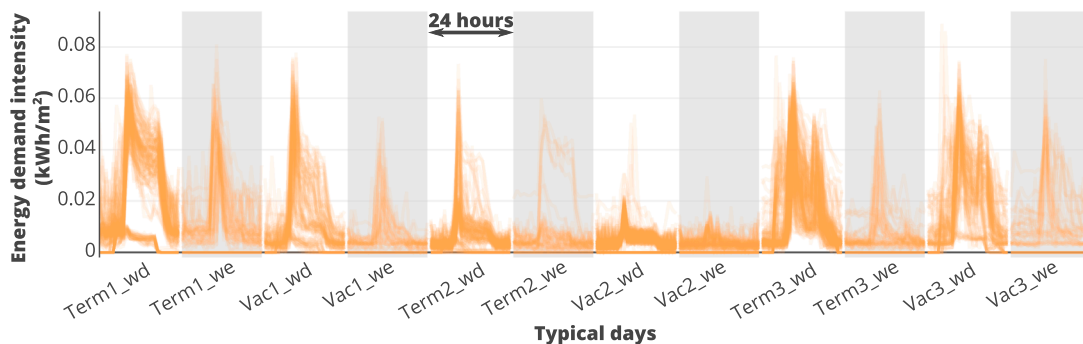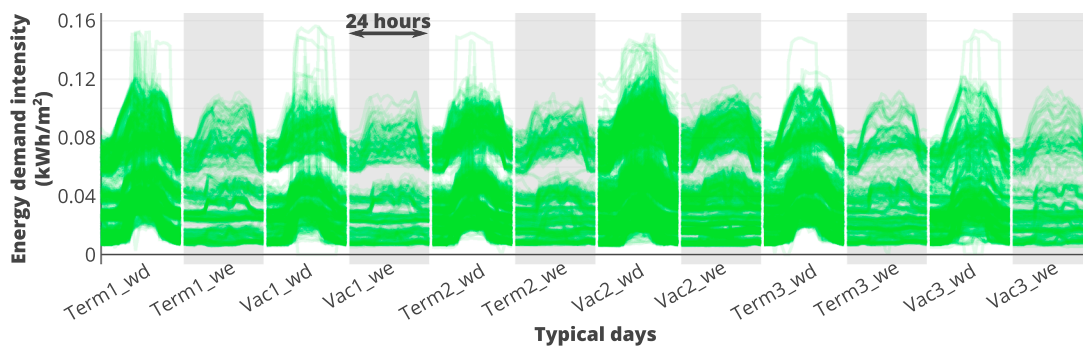
(e) Archetype: lab_commercial; energy type: electricity.
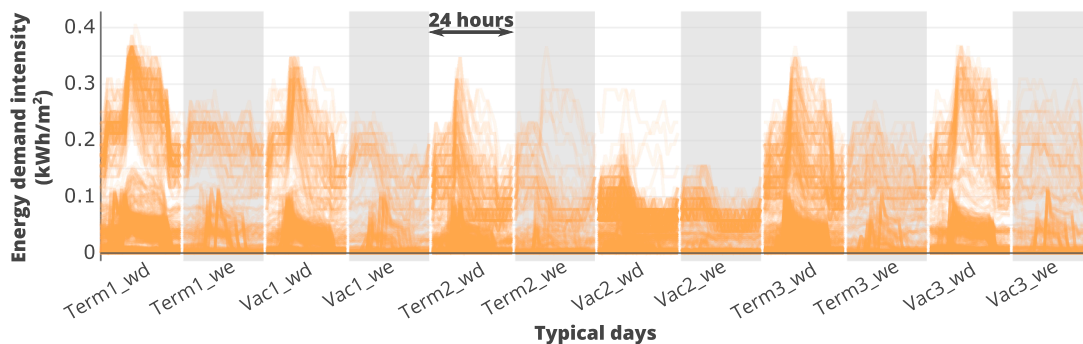
(f) Archetype: lab_commercial; energy type: gas.
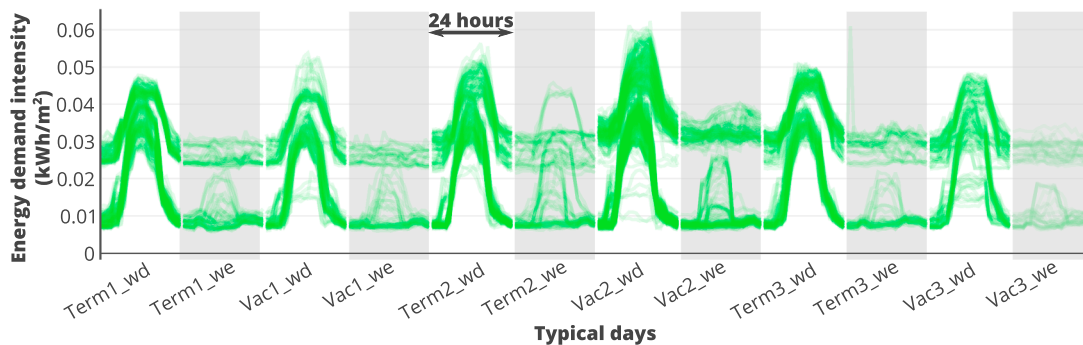
(g) Archetype: desk_commercial; energy type: electricity.

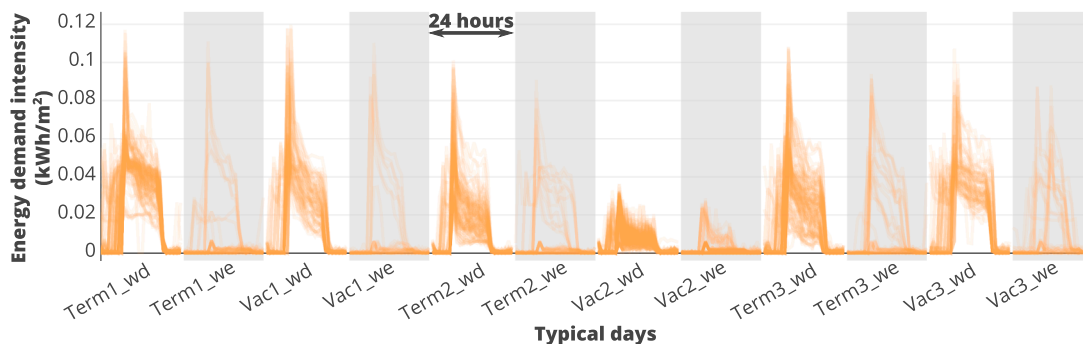(h) Archetype: desk_commercial; energy type: gas.

Fig. 5. (*continued*)

Table 2
Cambridge district node details.

Node	desk_research (10 nodes)	lab_research (17 nodes)	desk_commercial (18 nodes)	lab_commercial (1 node)	Energy centre
GIA	86,774	176,173	181,908	9,473	N/A
Roof area	26,715	72,648	69,824	5,535	
Technologies		PV, NGB, ST, StoreE, StoreT			CHP, StoreT, GridE, GridNG

Table 3
Dates corresponding to Cambridge term times and vacations (Vac), as used to define typical days for the Cambridge case study. Dates given for 2015 in *month-day* format. 'Vac3' dates wrap from December to January.

	1	2	3
Term	01–13 to 03–13	04–21 to 06–12	10–06 to 12–04
Vac	03–14 to 04–20	06–13 to 10–05	12–05 to 01–12

with a per unit capacity cost on investment, to better represent technology costs;

- Inter-cluster storage, which allows us to track stored energy between all days in our reference year, even though we are only optimising for 12/24 typical days; and
- A rooftop space limit for solar technologies, which limits the combined capacity of all solar technologies being considered.

A summary of the mathematical formulation of all constraints can be found in Appendix A and the full mathematical formulation can be found online.[4] Models were run on a high performance computing cluster, with optimisation undertaken by the Gurobi solver (v7.5.1).

2.4.1. Side note on tractability of optimisation models

Much of the work in this study is driven by a requirement for tractability. In formal terms, tractability might be considered as the necessary condition for the problem to be computed in polynomial time, i.e. within complexity class P [42, p.4]. Thus $NP-complete$ and $NP-Hard$ (where $P \neq NP$) would be considered intractable. This classic definition is somewhat questionable, as it is clearly more practical to solve a problem in the order of $2^{0.1N}$ (exponential) than in the order of N^{20} (polynomial) for a large N [43]. Additionally, a problem that has complexity in the order of N^2 becomes impractical to solve if the number of computations, N, is large (e.g. 10^{10}). Goderbauer et al. [44] recently formalised the complexity of MINLP energy system problems as $NP-Hard$ but did not discuss MILP problems. In fact, model complexity is rarely formalised. Instead, we interchange tractability and practicality. As such, we consider a problem to be tractable if it reaches an optimal (and feasible) solution in a 'reasonable' amount of time. What is reasonable depends on the modeller and the computational power at their disposal. Here, we consider our models to be tractable if they take less than 12 h to solve on a high performance computing cluster. We cannot model the infinite number of future demand profiles, hence the need for scenarios. Nor can we achieve tractability without a small number of scenarios and time steps, leading to scenario reduction and time clustering.

2.5. Evaluating decisions

It is likely that any energy system, no matter how carefully designed, will need to cope with unexpected future demand. In the context of this study, a robust decision is one which ensures that demand will be met in any realised future. By using SO, we aim to improve

decision robustness. However, the requirement for SR, to maintain tractability, already reduces the number of future demand variations for which we optimise. An effective means of assessing the consequence of our decisions, as well as the effectiveness of the SO approach, is to undertake out of sample (OOS) testing [35]. Although used in both SO [45,39] and robust optimisation [46,47], OOS scenario testing is not common in the majority of studies which incorporate uncertainty.

In our OOS tests, investment decisions have already taken place, following either single scenario or multiple scenario optimisation. The investment technology capacities are fixed, before being exposed to new operating conditions to evaluate their performance. The resulting optimisation model is concerned only with minimising operational costs, with the possibility to draw on a 'slack' variable to balance supply and demand. The slack variable imposes such a high cost on its use that it would only be chosen by the model in instances where there is no other way to balance the system. A robust investment portfolio is one which continues to meet demand, or depends relatively little on the slack variable, when faced with new data.

3. The illustrative studies

The two case studies represent very different techno-economic-geographic contexts and hence test different types of demand (cooling dominated versus heating dominated) and associated technologies. Their energy systems are optimised for minimising total costs (investment and operation, normalised to one year) following the same steps. However, for each, we use a different manner of processing historical energy demand data for the scenario generation step. The reason for this is solely due to the difference in nature and volume of data available from the two sites. At the same time, it is true that historical data from buildings is never uniformly available across sites, and future applications of our proposed work may devise their own techniques for processing their demand data as long as they preserve key properties for estimating future possible demand scenarios. We will discuss these in the following section.

3.1. Bangalore, India

A collection of office buildings within Bangalore, India have been selected, defining an illustrative district. Fig. 3a shows each of these buildings, and the nodes used to represent them. Most nodes consist of several buildings, which are connected at the same point on the district cooling network. Building floor area (Table 1) has been inferred from the external footprint and number of floors for each building. As the development is fictitious, no other information is known about these buildings; we use only their relative size and position to test our modelling approach. We propose four energy centres, at different positions on the periphery of the district. Each is sited on a currently undeveloped piece of land, according to satellite data.

The demand data used in this study has been acquired from a single office building in Bangalore. Five-minute sub-metered data is available for this building for a range of end-uses. These end-uses include air conditioning, lighting, and uninterruptible power supply. The data was captured from December 2015 to November 2016 (inclusive) and is

---

[4] https://github.com/calliope-project/calliope.

Table 4
Calculated bandwidths for each KDE subset in the Cambridge case study input dataset. Underlined values refer to those selected for use with a 'top-hat' kernel; all others use a 'gaussian' kernel. Background bars highlight relative magnitude of bandwidth.

	Heat commercial		research		Electricity commercial		research	
	desk	lab	desk	lab	desk	lab	desk	lab
Term1_wd	0.04997	0.05887	0.20264	0.12394	0.08637	0.03682	0.03647	0.0373
Term1_we	0.03101	0.08895	0.08343	0.057	0.01124	0.04261	0.03714	0.03247
Vac1_wd	0.1158	0.08505	0.07766	0.07911	0.0483	0.29988	0.09826	0.03743
Vac1_we	0.13675	0.1211	0.10433	0.05748	0.05955	0.07582	0.06453	0.05468
Term2_wd	0.49099	0.11374	0.07045	0.26323	0.04754	0.381	0.05114	0.14478
Term2_we	0.0457	0.02587	0.04204	0.03974	0.11849	0.01128	0.04833	0.02955
Vac2_wd	0.04018	0.06603	0.08173	0.04574	0.39096	0.02128	0.02703	0.05492
Vac2_we	0.01451	0.04203	0.02651	0.02466	0.01844	0.04253	0.03385	0.035
Term3_wd	0.05666	0.19518	0.04705	0.12393	0.06037	0.07404	0.04446	0.058
Term3_we	0.02685	0.04919	0.05434	0.08039	0.02702	0.03195	0.07398	0.03055
Vac3_wd	0.05348	0.12213	0.07176	0.04954	0.0512	0.07614	0.15696	0.05651
Vac3_we	0.07554	0.0764	0.08497	0.07413	0.02375	0.04223	0.06033	0.03264

shown clustered by typical day in Fig. 4. Air conditioning electrical consumption is converted to cooling demand by using the variable refrigerant flow system coefficient of performance (COP) of 1.6. Current literature would suggest a COP between 3 and 4 [48]. The low COP is that which was recorded for the system in operation within the metered office building. Given that the system has been in place for many years, older literature suggests that it is not an unreasonably low COP. Xia et al. [49] recorded a COP lower bound of 1.9, while Zmeureanu [50] found their rooftop units to have operational COP of 1.68 ±0.19 and 1.86 ±0.37, which compared particularly unfavourably to the rated COP of 2.9. To match the resolution of climate data, the five-minute metered consumption was resampled to hourly data.

There are clear trends visible in Fig. 4 per typical day, with weekends exhibiting a greater variation in demand. As expected, climate clearly affects cooling demand, with high April external temperatures causing high cooling requirements. In May, there are many days with zero demand. These likely correspond to building shutdown days in the Summer holiday period. Some zero demand days may be caused by metering errors, but this top-down method for assessing demand data makes it difficult to assess whether they are erroneous points or truly zero-demand days.

Table 1 shows the technologies that we allow at each node. There is no requirement that a given technology is installed at any particular node, as the investment step of the optimisation will decide this. At a building level, national grid electricity (GridE), a diesel generator (DG), and photovoltaic solar panels (PV) are possible technologies to meet electricity demand. In addition to the district cooling system, individual air conditioning units (AC) can meet cooling demand. In the central energy centres, a large scale electric water chiller (ECh) or a combined cooling, heat, and power plant (CCHP) can be installed. The CCHP is either a diesel or biomass fuelled generator, whose waste exhaust heat is redirected through an absorption chiller to produce cooling. Thermal storage (StoreT) is possible at the energy centre, but due to the relatively low energy density of cold water, we do not consider StoreT at a building level. Technology costs are collated from various sources. Where costs specific to India were not available, values from the (UK specific) SPON'S mechanical and electrical services price book [51] have been used, assuming a currency conversion factor of 90 INR/GBP. More detail on the district and technology definitions is available online,[5] including costs and their sources.
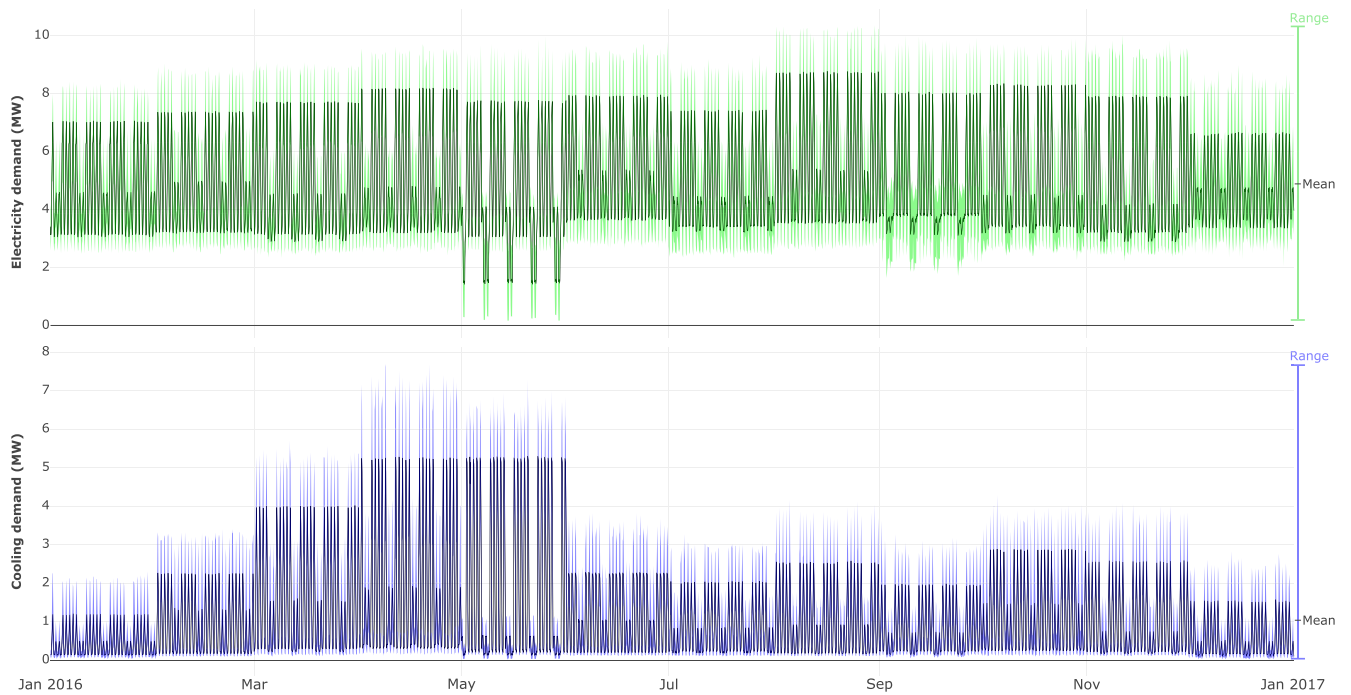
3.2. Cambridge, UK

Unlike the Bangalore 'representative' district, the Cambridge district is based on intended development by the University of Cambridge. The West Cambridge site is a campus of the University, in which there exists already a number of academic, residential, leisure, and commercial buildings. The plan[6] is to construct 383,000 m² of new floorspace, through a combination of greenfield and brownfield development (the latter directly replacing current buildings). According to the masterplan [28], the district will have a 42 GWh annual heating load, 70% met by Combined Heat and Power (CHP), and 88 GWh annual electricity load, 29 GWh of which will be met by the same CHP. To determine this expected load, the buildings on the proposed site have been categorised by archetypes: 'desk-based research', 'medium intensity laboratories', and 'high intensity laboratories'. Our chosen archetypes are further disaggregated in this study, into commercial and research usage. This different categorisation follows consultation with Aecom, the contracted consultants for the energy plan of the West Cambridge site. Thus, four building archetypes are considered: 'Desk-based Commercial', 'Desk-based Research', 'Lab-based Commercial', and 'Lab-based Research'.
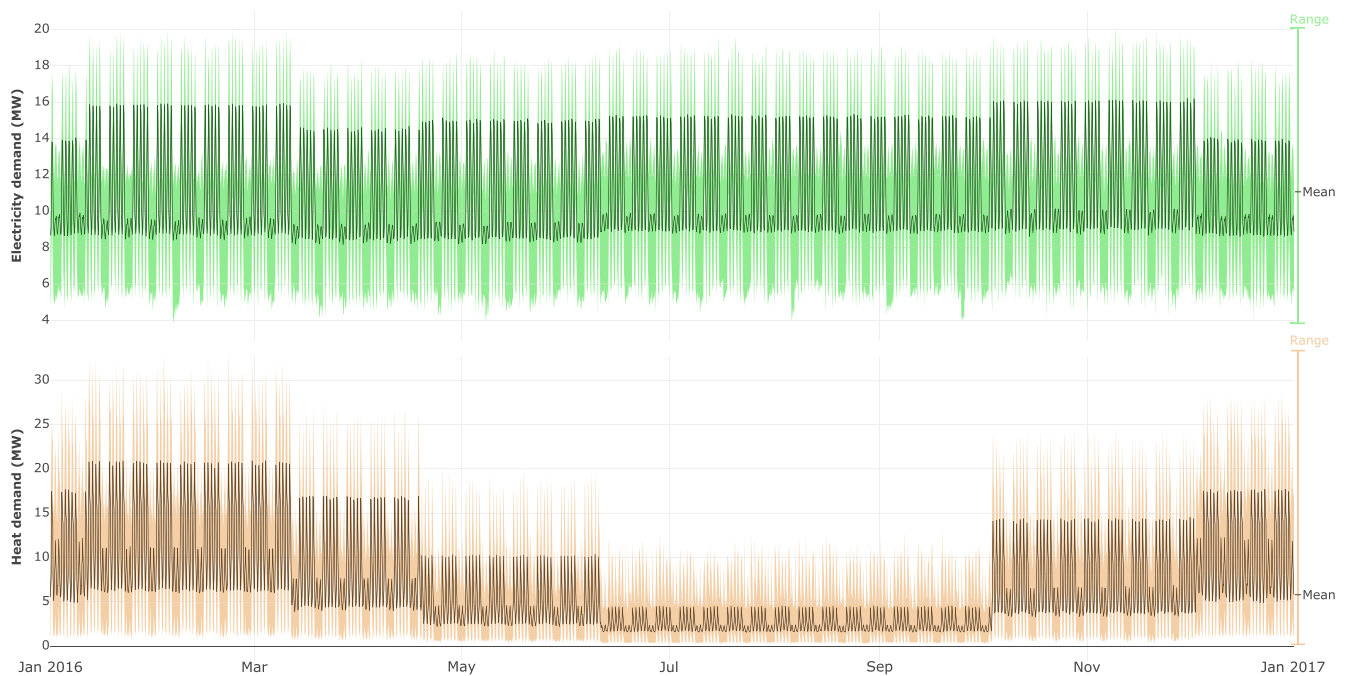
The masterplanners used expert judgement and/or the mean demand to inform the expected demand of proposed buildings within each archetype on the West Cambridge site. We have started from the same position, using existing buildings on the University estate to inform expected demand within the district. Gas and electricity consumption data for 17 buildings have been accessed from across the University. We assume that an 80% efficient boiler is used to meet heat demand from the incoming gas. As with our Bangalore case, we cluster our input into typical days. Many of the buildings are academic, so we have chosen to cluster the data based on weekends/weekdays and the University of Cambridge term dates (Table 3). This gives us 12 typical days. Different profile sub-clusters originating from different buildings within the same archetype are evident in Fig. 5. In some cases, this can lead to an order of magnitude variation in the possible peak on a given day. Electricity demand profile shapes are more pronounced than heating demand profiles, as in the Bangalore data. Although, there is a pronounced morning heating peak in desk-based research/commercial buildings. On weekends, demand is lower but the profile shape is more sporadic, particularly in the 'lab' archetypes. This is likely caused by lab

---

[5] https://github.com/brynpickering/bangalore-calliope.

[6] More detail on the West Cambridge plan can be found at http://www.westcambridge.co.uk/.
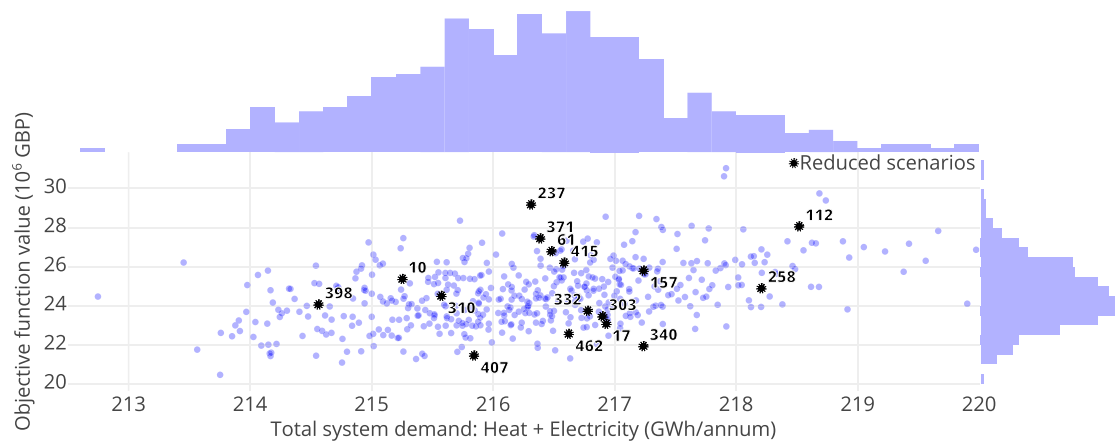
(a) Bangalore, India.

(b) Cambridge, UK.

Fig. 6. Total hourly district demand, as sampled for 500 scenarios for the year 2016. Mean profiles are shown as black lines. To ensure clarity, we have not displayed the profile of all 500 scenarios, instead the minimum to maximum range of all scenarios is given as a shaded region.

occupants choosing to work on weekends and unsupervised, energy intensive lab experiments taking place over weekends. The buildings require less heating in summer ('Vac2'), with hot water requirements being the likely cause of the remaining heat demand on these typical days.

Buildings within our chosen archetypes are scattered across the

development site. Fig. 3b shows the proposed CHP would be based in an energy centre at the western edge of the district. The heating network follows the road network in connecting to the buildings. We also include a gas network cost along the same network, to account for the installation of gas pipework. To make the most of a possible energy centre, we also include the possibility of a large-scale ground source

(a) Cambridge, UK.
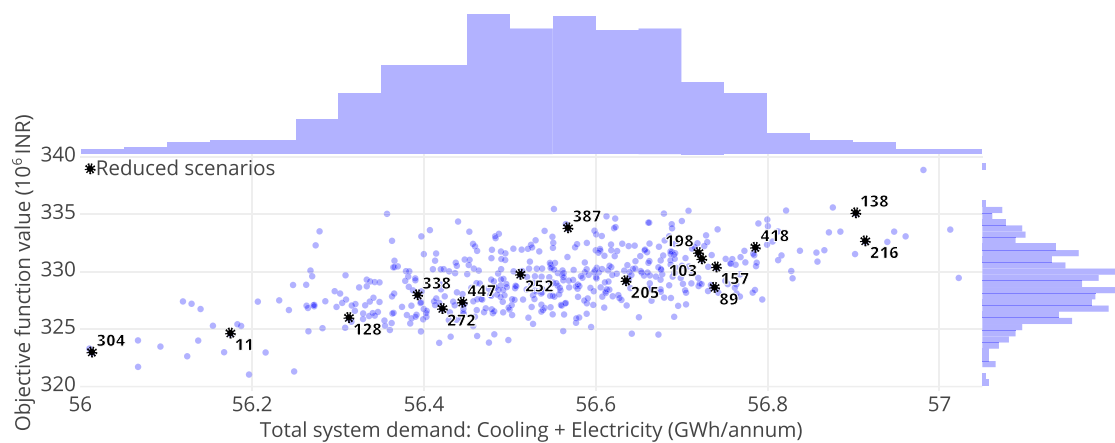
(b) Bangalore, India.

Fig. 7. Total district demand compared to independently optimal objective function value for 500 demand scenarios. Scenarios chosen to represent the full set, following scenario reduction, are highlighted and numbered. The distribution of demand and objective function value is given outside the scatter plot.

heat pump (GSHP) and thermal energy storage (StoreT) to enter into the district system. Nodes in the district correspond to buildings of different archetypes, with roof area available for solar technologies and building-level technologies made available, if a district system is not favourable (see Table 2). Building-level heat demand can be met by natural gas boilers (NGB) or solar thermal panels (ST), and can be stored using building-level StoreT. Photovoltaic solar panels (PV), grid electricity (GridE), and CHP output can meet electricity demand. Electricity can also be stored in batteries (StoreE) at a building level. Technology costs are primarily taken from the SPON'S mechanical and electrical services price book [51]. More detail on the district and technology definitions is available online, including costs and their sources.[7]

4. Results & discussion

4.1. Scenario generation

Different methods were used to generate scenarios for the Bangalore and Cambridge case study districts. The Bangalore samples were generated prior to this study, using fPCA [32]. In this study, the samples

were applied to case study buildings by scaling them to building floor area and randomly assigning them to days corresponding to our pre-defined typical days.

For Cambridge, the acquired data included several buildings describing each archetype. Consequently, KDE was applied to multi-building sets. As we use KDE for each energy type (2), archetype (4), and typical day (12), we require our limited observations to generate 96 probability density functions. To make the most of our observations, 5-fold cross-validation was used to calculate the best-fit bandwidth in the range (0, 1] with either the 'gaussian' or 'top hat' kernel, using the Python package Scikit-learn [52] for KDE and Hyperopt [53] for optimising the error. Table 4 shows the resulting bandwidths and kernels following 5-fold cross-validation. Most probability density functions (PDFs) are described using a gaussian kernel and the bandwidth varies from 0.01 to 0.49. A larger bandwidth refers to more variable data, requiring a sufficiently smoothed PDF to describe the full dataset.

Fig. 6 gives the range of all the 500 scenarios of district-wide energy demand per utility for both case studies. Mean demand profiles are shown as lines and minimum to maximum range of any scenario is given as a shaded region (also marked on the right side of the y-axis). Each scenario is an aggregation of the demand for each building and every day in the year. The demand in Bangalore is described with 24 distinct typical days (weekend and weekday per month) and the demand in Cambridge is described with 4 archetype buildings, each with

---

[7] https://github.com/brynpickering/cambridge-calliope.

Table 5

Description of the 16 reduced scenarios chosen to represent the full set of 500, given to three significant figures. Scenarios are ordered by objective function value (left to right ascending), as this is the value on which SR takes place. Highest values in each category are highlighted in red colour.

Scenario	407	340	462	17	303	332	398	310	258	10	157	415	61	371	112	237
Demand (x10^{8}kWh)																
Heat	6.58	6.56	6.52	6.58	6.47	6.59	6.46	6.45	6.68	6.60	6.63	6.66	6.48	6.62	6.70	6.63
Electricity	15.0	15.2	15.1	15.1	15.2	15.1	15.0	15.1	15.1	14.9	15.1	15.0	15.2	15.0	15.1	15.0
Sum	21.6	21.7	21.7	21.7	21.7	21.7	21.5	21.6	21.8	21.5	21.7	21.7	21.6	21.6	21.9	21.6
Objective function value (x10^{7}GBP)	2.15	2.19	2.26	2.31	2.35	2.37	2.41	2.45	2.49	2.54	2.58	2.62	2.68	2.74	2.81	2.92
Weight (%)	3.00	2.40	6.20	8.40	7.80	6.40	9.60	9.40	11.2	8.20	7.40	8.20	4.40	4.00	2.40	1.00

(a) Cambridge, UK.

Scenario	304	11	128	272	447	338	89	205	252	157	103	198	418	216	387	138
Demand (x10^{7}kWh)																
Cooling	1.05	1.06	1.05	1.05	1.07	1.07	1.07	1.05	1.06	1.07	1.07	1.07	1.07	1.06	1.07	1.08
Electricity	4.56	4.56	4.58	4.59	4.57	4.57	4.61	4.61	4.59	4.61	4.60	4.60	4.61	4.64	4.59	4.61
Sum	5.60	5.62	5.63	5.64	5.64	5.64	5.67	5.66	5.65	5.67	5.67	5.67	5.68	5.69	5.66	5.69
Objective function value (x10^{8}INR)	3.23	3.25	3.26	3.27	3.27	3.28	3.29	3.29	3.30	3.30	3.31	3.32	3.32	3.33	3.34	3.35
Weight (%)	1.80	3.60	6.60	7.20	9.00	7.80	9.80	7.80	8.00	10.2	5.80	5.20	4.80	5.20	4.80	2.40

(b) Bangalore, India.

12 typical days (academic term times and vacation). These typical days are evident in the samples, particularly between weekdays and weekends. Electricity demand shows less variation in throughout the year than thermal demand. Indeed, the variation in thermal demand is not sufficiently well described by the choice of typical days. A greater number of typical days would better represent the data, leading to smoother transitions in both case studies.

There is a high possible deviation from the mean in any of the 500 scenarios, particularly for electricity and heat demand in Cambridge (Fig. 6b). The demand in any time interval could range from 0.5 to 1.5 times the mean. Within a set of days associated with the same typical day, there is much less variation in the peak demand. This variation remains below 1 MW in the mean curve and 5 MW in the min/max range. As the profiles in Fig. 6 aggregate all buildings in the district, the deviation of the profile mean and min/max is lower than that which is visible on a building-level.

The total annual demand of the Cambridge district is approximately 1.61 times the amount predicted in the energy masterplan. The total annual demand is 142 GWh electricity and 68.2 GWh heat compared to 88 GWh and 42 GWh respectively predicted by the masterplan. The masterplan is based on a single archetypal annual demand mapped to all buildings of that archetype. As such, it is possibly less accurate than our estimate, suggesting that a greater demand can be expected on the site. We cannot validate either our result or that of the masterplan until all buildings are commissioned in approximately ten years' time. In Bangalore, we cannot compare the demand to a district prediction, as no masterplan exists for this illustrative study.

As aforementioned, we model only typical days, not the full year given in Fig. 6. Thus, only one day can be used to represent each typical day in the optimisation. For each typical day (e.g. January weekday), the mean profile of all days represented by it (e.g. January weekdays) is used as the model input demand. The impact of a typical day on the objective function is scaled to the number of days in the year it represents. The impact of just using the typical days, instead of the full timeseries, can be calculated ex-ante. In Cambridge, the mean annual demand for electricity and heat demand shown in Fig. 6b varies by +5.6% & −3.5%, respectively, compared to the description of the annual demand used in the optimisation. This change is more pronounced in Bangalore, where electricity and cooling vary by +7% & +20%, respectively. Time dimension reduction is standard practice in MILP modelling, as previously discussed in Section 2.1. Optimisation results will differ due to this reduction [24], but we have implemented intra-cluster storage management [54,55] to help improve the accuracy of the results.

4.2. Scenario reduction

When discussing the methodology in Section 2.2, we introduced two approaches to scenario reduction. Figs. 7a and b allows us to compare two metrics describing our 500 scenarios: annual total system demand and independently optimal objective function value. The better the correlation between the two metrics, the more likely we could use the simpler of the two (annual total system demand) for scenario reduction. Annual demand varies in Cambridge from 212 GWh/a to 220 GWh/a, compared to an objective function value variation of 20.2 million GBP to 31 million GBP (Fig. 7a). There is a trend for higher system cost with an increase in total system demand. However, the highest objective function value is associated with a system demand of ~218 GWh/a and the lowest objective function value with a system demand of ~213.8 GWh/a. There is also a distribution skew in Cambridge in objective function value, towards lower cost, that is not exhibited in the total system demand. These factors all reinforce the use of the additional step in scenario reduction, where an independently optimal objective function value was sought for each of the 500 scenarios. In Bangalore, the use of the additional step is less clear. There is a similarly small percentage variation in total system demand, from 56 GWh/a to just over 57 GWh/a (Fig. 7b). But the variation in objective function value, 321 million INR to 339 million INR, is relatively well correlated to the demand. Still, the lowest and highest demand scenarios are not the lowest and highest cost scenarios.
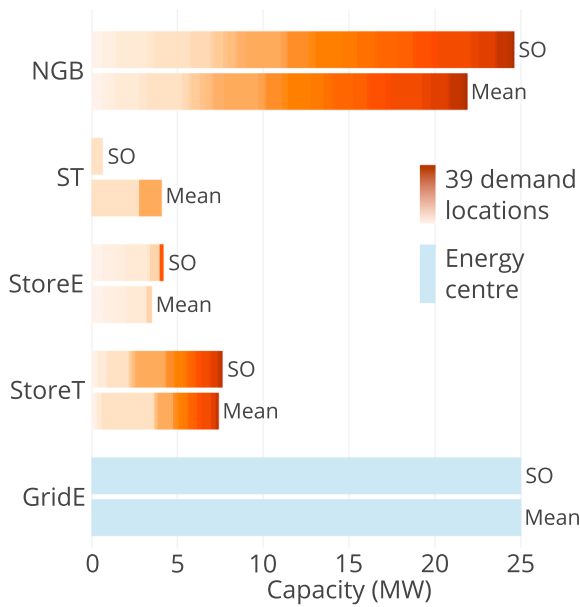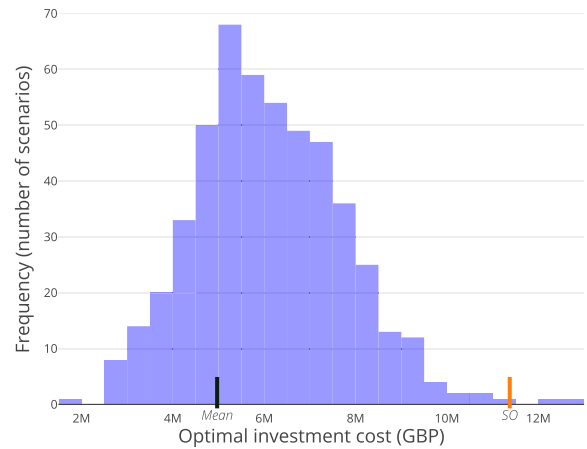
(a) Cambridge, UK.
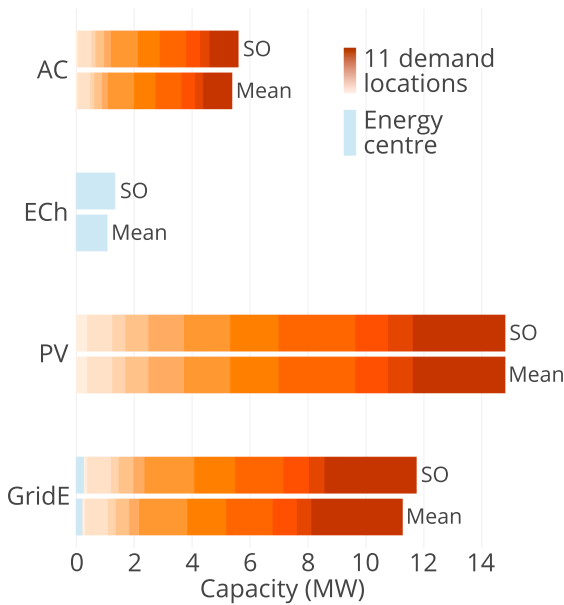
(b) Bangalore, India.

Fig. 8. Installed capacity of technologies to achieve the optimal objective function value in both mean (single scenario) and SO cases. Building-level technologies have been aggregated over all demand nodes. The contribution from each demand node to the total technology capacity has been differentiated with a colour gradient. Although, in some cases (e.g. 'ST' in (a)), not all nodes have installed capacity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The reduced scenarios chosen to represent the full set are also highlighted in Fig. 7. They are more bunched towards the centre of total system demand in Cambridge, compared to the greater spread exhibited in Bangalore. As a result, the scenario with the highest objective function value in Cambridge (237) is the 12th highest total system demand scenario. In Bangalore, the corresponding scenario (138) is the second highest total system demand scenario. Table 5 inspects the reduced scenarios in detail. We can see that maximum thermal and
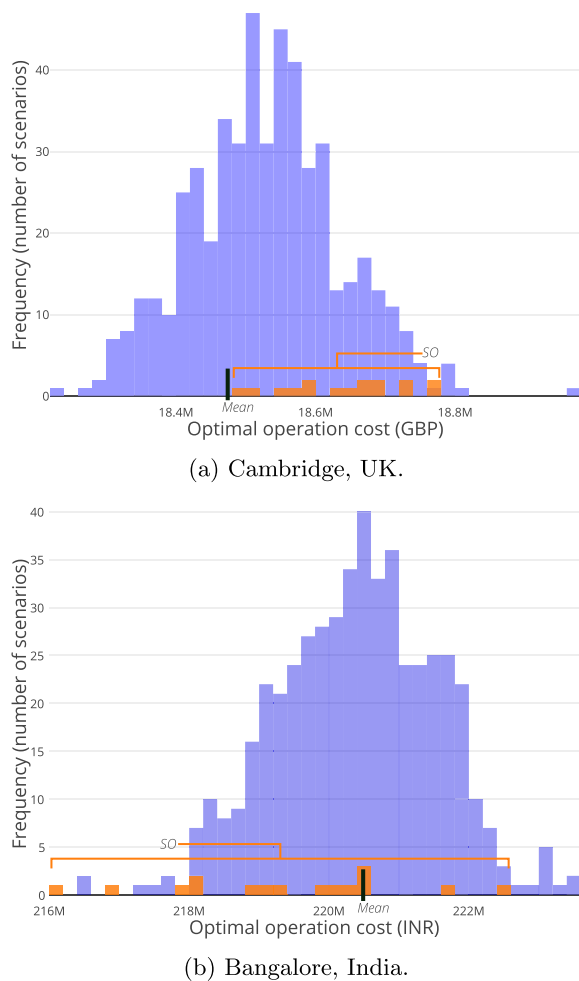

(a) Cambridge, UK.
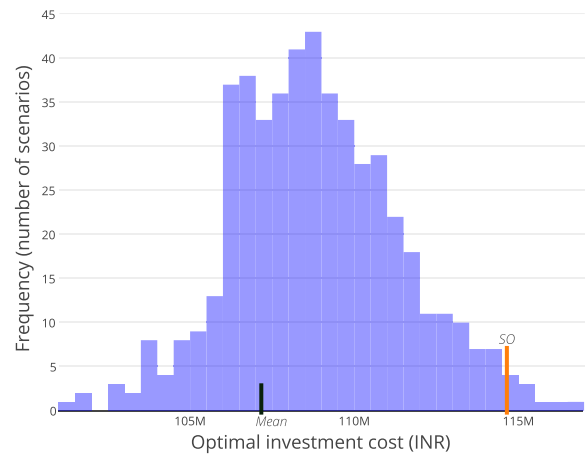
(b) Bangalore, India.

Fig. 9. Contribution of investment cost to the objective function value for all 500 independent scenarios, compared to mean and SO contributions. The 500 scenario costs form the histogram, while the singular results from the mean scenario and from SO have been pinpointed on the x axis.

electricity demand is not given by the same scenario in either district. The proportion of scenarios a single reduced scenario represents ranges from 1% to 11.2%. We use this to weight each scenario in the objective function. The lower weight reduced scenarios are, as expected, those at the extremes of objective function value. Thus, costs incurred in an extreme scenario will have approximately 10 times less influence on the objective function value than in a moderate scenario (see Eq. (1)). In the optimisation, this translates to a risk-based compromise, where we are willing to accept a relatively high operational cost in unlikely futures (our extreme scenarios) as they have an order of magnitude lower impact on objective function value.

4.3. Scenario optimisation

Optimisation of the 16 reduced scenarios by SO took 4.12 h and 8.51 h to reach a solution for the Cambridge and Bangalore cases, respectively. The results follow similar trends in both case studies. Neither considers an energy centre and district network to be particularly important (Fig. 8). This is caused by the prohibitive cost of laying thermal network pipework, a cost not often included in district network studies. Albeit small (<2 MW), a centralised electric chiller with a higher efficiency is considered in an energy centre in the Bangalore

(a) Cambridge, UK.

(b) Bangalore, India.

Fig. 10. Contribution of operation cost to the objective function value for all 500 independent scenarios, compared to mean and SO (16 scenarios) contributions. The 500 scenario costs form one histogram and the cost of the 16 SO future realisations form another. The singular result from the mean scenario has been pinpointed on the x axis.

case. The optimisation deems its small size relative to the total cooling demand to still be worth the land purchase and pipework cost. Practically, it could probably fit within one of the existing structures, which would make it more financially viable. Renewable energy technologies are not particularly important in the Cambridge case. Rather than increasing thermal storage to handle uncertainty in the use of heat from solar thermal panels, it is more cost optimal to add several MW of boiler capacity. This is less of a concern in Bangalore, where electricity from solar photovoltaics is sufficiently cost beneficial to be installed across the maximum available roof space in all optimisation runs. Other than the decreased dependence on solar thermal energy in Cambridge, there is little difference in installed energy capacity between the mean and SO runs.

Investment costs are high as a result of SO: +6.4 million GBP in Cambridge, +7.5 million INR in Bangalore, compared to the mean investment (Fig. 9). Compared to the 500 independent scenario runs, the investment cost is in the top 5% for SO in both case studies. This is a high investment penalty, for little additional capacity investment. It is unlikely that system designers would be willing to accept this investment increase unless it led to a clear improvement in system robustness. No matter the scenario, operating costs are higher than investment costs. The operating costs for each of the 16 scenarios are based on the

same technology investment portfolio, but with a different set of demand profiles being realised. In the Cambridge case, the operating costs, no matter which scenario is realised, will be higher than the costs incurred according to the result of the mean scenario (Fig. 10a). This does not make the mean model better, as the cost it portrays relies on the demand being exactly the mean of the uncertain profiles. It cannot necessarily meet demand in the future scenarios used in SO, let alone doing so at a lower cost. The spread is greater in the Bangalore case, where most scenarios have a lower operating cost than the mean case, including an operating cost lower than any of the 500 independent scenarios (Fig. 10b). This possible lower operating cost is balanced by a higher investment cost, hence why such a low operating cost was not seen when running the 500 scenarios through independent optimisations.
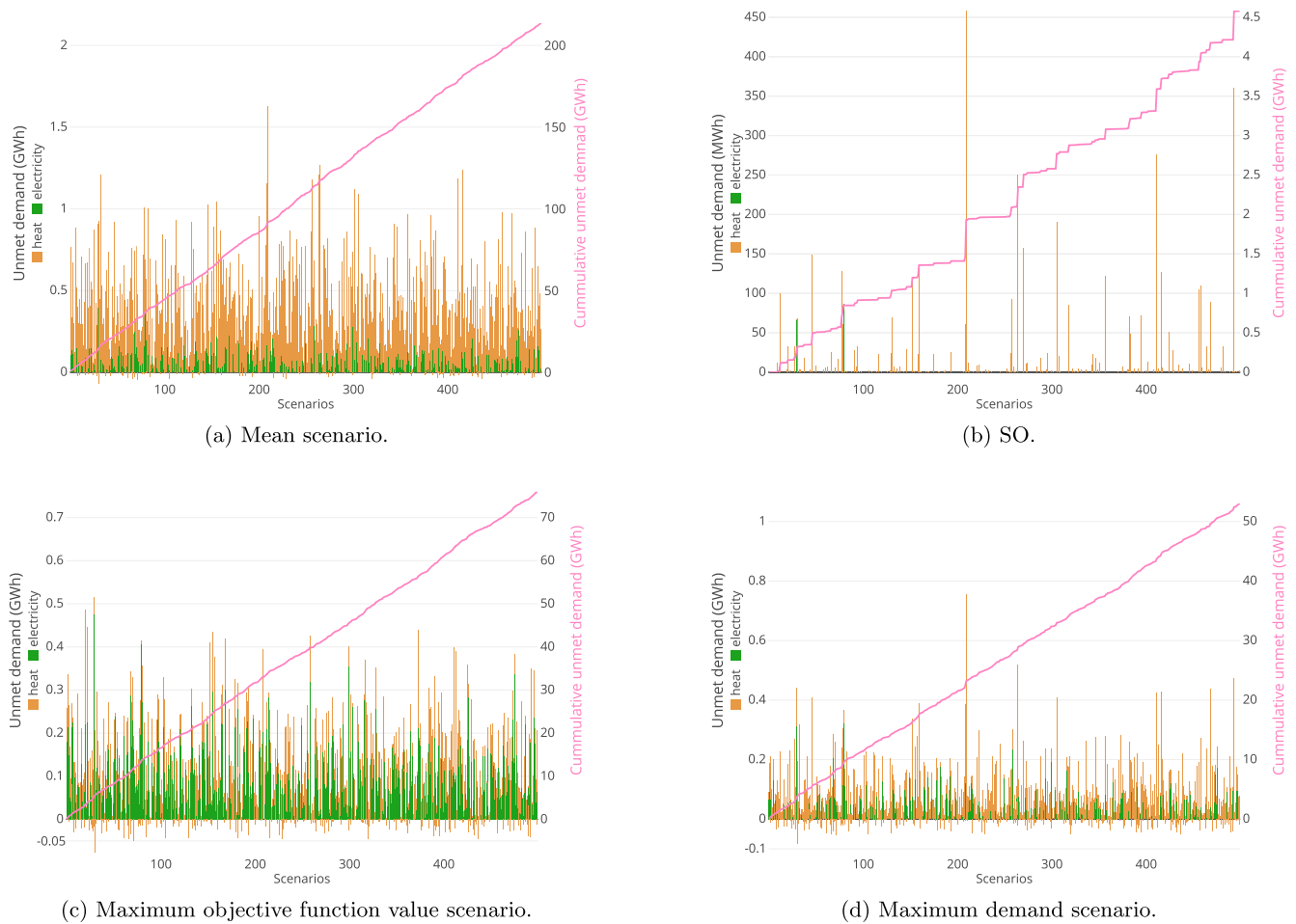
4.4. Out of sample tests

When conducting SO, demand must be met in all scenarios. This is also the case in the independently run scenario models. We can test the robustness of the optimal technology portfolios in select models by re-running our 500 scenarios with fixed technology capacities and capturing the resulting supply/demand imbalance. The optimal technology portfolio of four models are considered in this section: mean, SO, highest objective function, and highest total system demand. In each test, we run the optimisation with fixed capacities and a 'slack' decision variable in the system balance constraint. In any time step where the supply cannot match demand due to lack of capacity (or over-capacity, in the case of unwanted energy produced by e.g. solar thermal panels), the slack variable will record the imbalance. The greater the total imbalance across the 500 scenarios, the less robust the technology portfolio.

SO leads to a more robust technology portfolio, in both Cambridge (Fig. 11) and Bangalore (Fig. 12). By considering our 16 scenarios in SO, we have ensured that we cover a much greater range of futures in our system design, without increasing the system cost above the 'worst case'. We can gain a 50× and 25× reduction in system imbalance in the Cambridge and Bangalore districts, respectively if we use SO instead of the mean scenario. Comparing to our extreme independent scenarios, SO retains a more robust solution. If we spent more on technology investment by designing to the result of the highest objective value scenario it would be less robust than SO. A similar result would be obtained by designing to the result of the highest total system demand scenario. In fact, it is only the SO investment portfolio that leads to multiple scenarios having no imbalance. Granted, using 16 scenarios has not ensured full balancing when testing against all 500 scenarios, as seen in Figs. 11b and 12b. This is an expected result of our scenario reduction process, much in the same way that there are likely to be more extreme scenarios than those encompassed by our initial 500 scenarios. Currently, these simplifications are a necessity for tractability.

Although more robust, the increased monetary cost for SO may not be worth the risk reduction. The total worst case imbalance in any scenario is never greater than 1.6 GWh in Cambridge, 60MWh in Bangalore. This compares to annual system demands of >200 GWh in Cambridge and ~56 GWh in Bangalore, 2–3 orders of magnitude greater than the imbalance. We drastically reduce the imbalance in the SO case, but it is of small consequence.

4.5. The need for a new objective

Conventional building-level systems (Boiler & AC) are flexible at meeting varying demand and are still relatively cheap. Hence, they are the choice for meeting the demand in a district, whether or not we are aiming for robustness to future demand scenarios. The most efficient

(a) Mean scenario.

(b) SO.

(c) Maximum objective function value scenario.

(d) Maximum demand scenario.

Fig. 11. Cambridge case study system unmet cooling and electricity demand when running out of sample (OOS) optimisation tests. the 500 OOS scenarios are those generated in SG. Technology capacities have been fixed in the tests to the result of optimising the a. mean scenario, b. 16 reduced scenarios in SO, c. scenario with highest objective function value (when optimised independently), and d. scenario with the highest total annual demand. Cumulative imbalance is the rolling sum of the unmet demand from individual tests.

systems, such as the CHP and CCHP cogeneration technologies, have not been chosen. This is particularly problematic when making decisions under uncertainty, where existing technologies such as building-level air conditioning and boilers are better placed to handle varying demand profiles. A new objective is required if environmental agendas are to be prioritised in the design of energy infrastructure. This may be the inclusion of the environmental impact, be it only operational or for full technology lifecycles. In including environmental concerns in the objective function, we can also extend our work to include risk aversion by considering unmet demand as being met by a highly polluting backup system. The existing optimal solutions also depend heavily on having a national grid connection and are thus not robust to intermittency nor the requirement for energy autarky.
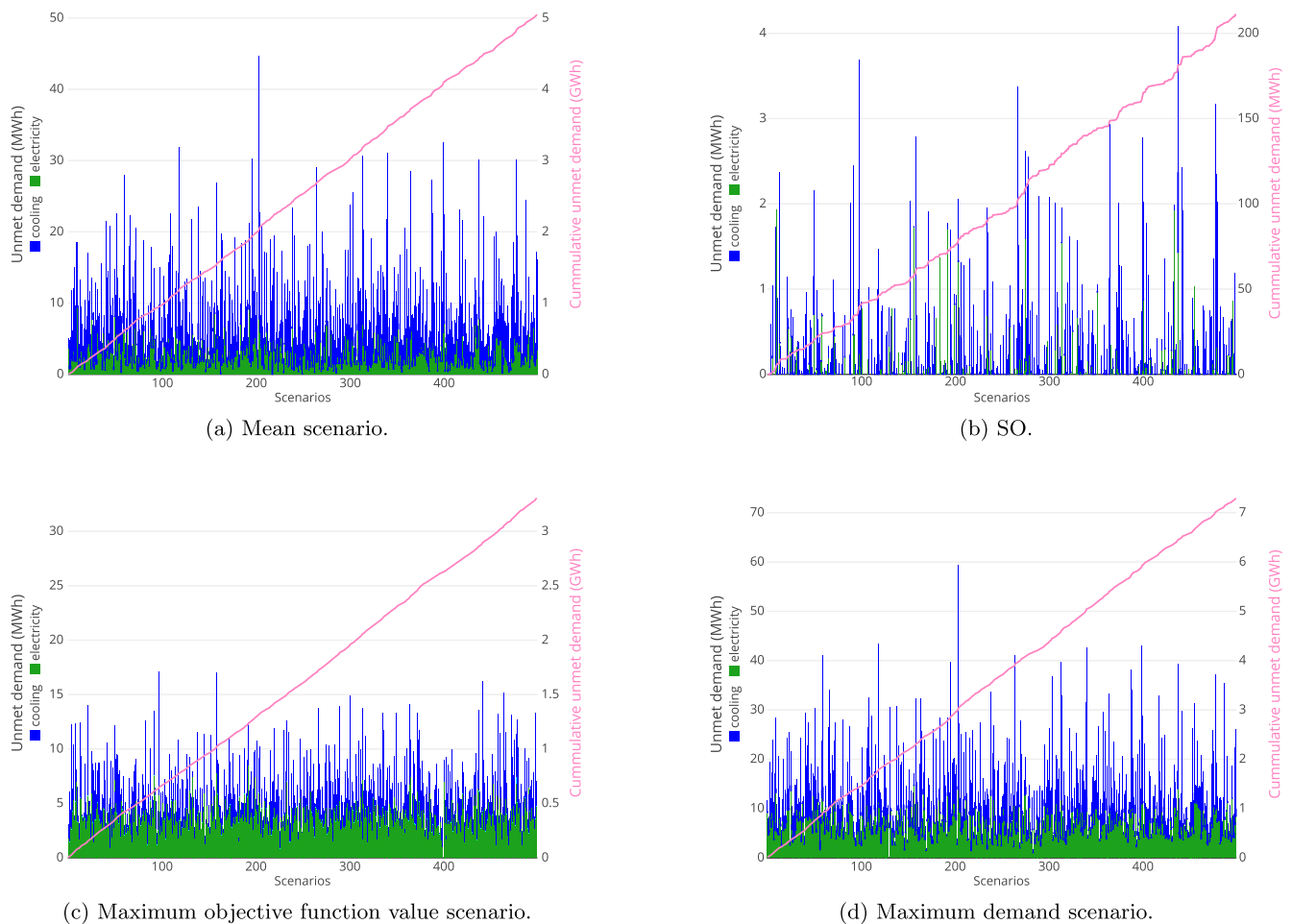
5. Conclusion

Decision making under uncertainty can be improved by the combination of new sampling and optimisation techniques. We have shown in this study that by following a 3-step method, we can design district energy systems which are more robust than traditional, single scenario models to the risk of not meeting a range of future demand. These three steps are scenario generation, scenario reduction, and scenario optimisation. In scenario generation, we use multivariate nonparametric

sampling to produce 500 future demand scenarios from historic building-level stochastic consumption data. Using scenario reduction, we are able to select 16 scenarios for scenario optimisation, without misrepresenting the probability distribution of our 500 initial scenarios. With 16 representative scenarios, we can run tractable scenario optimisation models.

We applied these three steps to illustrative case studies in Cambridge, UK, and Bangalore, India. By using out of sample scenarios, unmet demand has been quantified for scenario optimisation and single scenario technology investment portfolios. Robustness to unmet demand increased by $50\times$ in the Cambridge case, and $25\times$ in the Bangalore case when using scenario optimisation derived over mean scenario derived technology capacities. However, this robustness comes at a high cost: $+6.4$ million GBP and $+7.5$ million INR investment in the Cambridge and Bangalore cases, respectively. These results have been applied directly to a masterplan-level site (Cambridge), showing that such a method can be readily utilised by practitioners.

However, given increasing calls to reduce the environmental impact of infrastructure, our cost optimal solutions are not helping. Contrary to existing literature, district energy systems are not considered cost optimal in our study. Incumbent energy technologies are preferred, such as building-level boilers and AC. There is some reliance on rooftop renewable solar technologies, for heat and electricity. However, increased

(a) Mean scenario.

(b) SO.

(c) Maximum objective function value scenario.

(d) Maximum demand scenario.

Fig. 12. Bangalore case study system unmet cooling and electricity demand when running out of sample (OOS) optimisation tests. the 500 OOS scenarios are those generated in SG. Technology capacities have been fixed in the tests to the result of optimising the a. mean scenario, b. 16 reduced scenarios in SO, c. scenario with highest objective function value (when optimised independently), and d. scenario with the highest total annual demand. Cumulative imbalance is the rolling sum of the unmet demand from individual tests.

robustness leads to reduced dependence on solar thermal technologies in our Cambridge case study. The lack of energy centre in Cambridge and minimal centralised chiller production in Bangalore suggests that a new objective is required within scenario optimisation. Such an objective would need to prioritise emission reduction or, at least, penalise the use of the most polluting technologies. Future work will concentrate on reformulating the objective function to this end. Additionally, we will test our scenario optimisation derived technology capacities for robustness against national grid intermittency, a particular source of uncertainty in our Bangalore case study.

Acknowledgements

This research was supported by the Engineering and Physical

Appendix A. Model mathematical formulation

The mixed integer linear programming (MILP) framework used to model our two districts is typical of district energy system modelling. Sets, decision variables and constraints are summarised in this appendix. Decision variables are those variables which are unknown at the start of the optimisation. Their values are selected by the optimisation algorithm so that they satisfy the realisation of the objective function value. Constraints are placed upon the decision variables, to ensure they cannot take values which are either physically or mathematically impossible. For a more detailed understanding of the model implementation, readers are referred to the Calliope software repository.[8]

---

[8] https://github.com/calliope-project/calliope.

A.1. Sets and decision variables

Sets

$n \in N$	Set of geographic nodes in the system
$(n, n_r) \in L$	Node pairs, defining distribution links
$x \in X$	Set of available technologies
$t \in T$	Set of operational timesteps
$c \in C$	Set of energy carriers

Subsets

store	Storage technologies
conv	Conversion technologies
supply	Supply technologies
dem	Demand technologies
dist	Distribution technologies
ex	Technologies which can export a carrier
prod	Technologies which can produce a carrier
con	Technologies which can consume a carrier
area	Technologies which use physical area

Decision Variables

$\mathbf{P}^{\mathbf{cap}}_{n,x} \in [0, inf)$	Energy capacity (kW)
$\mathbf{S}^{\mathbf{cap}}_{n,x} \in [0, inf)$	Storage capacity (kWh)
$\mathbf{R}^{\mathbf{area}}_{n,x} \in [0, inf)$	Resource capture area (m²)
$\mathbf{Y}_{n,x} \in \{0, 1\}$	Purchase switch
$\mathbf{P}^{+}_{n,x,t,c} \in [0, inf)$	Energy carrier production (kW)
$\mathbf{P}^{-}_{n,x,t,c} \in (-inf, 0]$	Energy carrier consumption (kW)
$\mathbf{P}^{\mathbf{ex}}_{n,x,t,c} \in [0, inf)$	Energy carrier export (kW)
$\mathbf{S}_{n,x,t} \in [0, inf)$	Stored energy (kWh)

A.2. Constraints

Constraints are applied across the full sets of each decision variable and parameter indices, unless explicitly noted in the constraint.

A.2.1. Technology capacity

Our capacity decision variables must be kept below a given parameter, denoted with the subscript *max*.

Storage capacity:

$$\mathbf{S}^{\mathbf{cap}}_{n,x} \leqslant S^{cap}_{max_{n,x}} \quad \forall\ (n, x) \in (N, X)_{store}$$

Energy capacity:

$$\mathbf{P}^{\mathbf{cap}}_{n,x} \begin{cases} \leqslant P^{cap}_{max_{n,x}} \times \mathbf{Y}_{n,x}, & \text{if} cost_{purch} \geqslant 0 \\ \leqslant P^{cap}_{max_{n,x}}, & \text{otherwise} \end{cases}$$

Resource capture area:

$$\mathbf{R}^{\mathbf{area}}_{n,x} \leqslant R_{area,max_{n,x}} \quad \forall\ (n, x) \in (N, X)_{area}$$

At each node, there is a limit on the available area:

$$\sum_{x \in X} \mathbf{R}^{\mathbf{area}}_{n,x} \leqslant available\_area_n$$

Area use is linked to energy capacity of a technology, where η_{area} is the area required per unit capacity:

$$\mathbf{R}^{\mathbf{area}}_{n,x} = \mathbf{P}^{\mathbf{cap}}_{n,x} \times \eta_{area_{n,x}} \quad \forall\ (n, x) \in (N, X)_{area}$$

Storage capacity is linked to energy capacity by the charge rate:

$$\mathbf{P}^{\mathbf{cap}}_{n,x} \leqslant \mathbf{S}^{\mathbf{cap}}_{n,x} \times charge\_rate_{n,x} \quad \forall\ (n, x) \in (N, X)_{store}$$

Technologies which define a distribution link have the same capacity:

$$\mathbf{P}^{\mathbf{cap}}_{n,x} = \mathbf{P}_{n_r}{}^{\mathbf{cap}}_{,x} \quad \forall\, x \in X_{dist}$$

A.2.2. Dispatch limits
Carrier flow is limited by the technology capacity.
Carrier production:

$$\mathbf{P}^+_{n,x,t,c} \leqslant \mathbf{P}^{\mathbf{cap}}_{n,x} \quad \forall\, (n, x, c) \in (N, X, C)_{prod}$$

Carrier consumption:

$$-1 \times \mathbf{P}^{\mathbf{cap}}_{n,x} \leqslant \mathbf{P}^-_{n,x,t,c} \quad \forall\, (n, x, c) \in (N, X, C)_{con}$$

Stored energy:

$$\mathbf{S}_{n,x,t} \leqslant \mathbf{S}^{\mathbf{cap}}_{n,x} \quad \forall\, (n, x) \in (N, X)_{store}$$

A.2.3. Energy balance
At each timestep, carrier flows must balance. This includes a system-wide energy carrier balance, as well as technology specific balances.
All demand must be met at each node, with no slack for excess supply:

$$\sum_x \mathbf{P}^+_{n,x,t,c} + \mathbf{P}^-_{n,x,t,c} + \mathbf{P}^{\mathbf{ex}}_{n,x,t,c} = 0$$

Demand technologies, where R is the (-ve) timeseries demand:

$$\mathbf{P}^-_{n,x,t,c} \times \eta_{n,x,t} = R_{n,x,t} \quad \forall\, (n, x, c) \in (N, X, C)_{dem}$$

Distribution technologies, where η is the link efficiency:

$$-1 * \mathbf{P}^-_{n,x,t,c} \times \eta_{(n,n_r),x,t} = \mathbf{P}^+_{n_r,x,t,c} \quad \forall\, (x, c) \in (X, C)_{dist}$$

Conversion technologies, where η is the technology conversion efficiency:

$$-1 * \mathbf{P}^-_{n,x,t,c_{con}} \times \eta_{n,x,t} = \mathbf{P}^+_{n,x,t,c_{prod}} \quad \forall\, (n, x) \in (N, X)_{conv}$$

Multiple-carrier conversion technologies (e.g. CHP, CCHP), where p_2 / p_3 are the additional output carriers of the particular technology and *ratio* is the carrier conversion ratio (e.g. the heat to power ratio):

$$\mathbf{P}_{n,x,t,c^+_{prod}} = \frac{\mathbf{P}^+_{n,x,t,c^+_p}}{ratio_{n,x,c_p}}$$

$$\forall\, p \in \{(p_2), (p_3)\}, \quad \forall\, (n, x) \in (N, X)_{conv}$$

Supply technologies, where *force_R* is a boolean parameter dictating whether an available resource has to be consumed when available (e.g. un-curtailed PV) or not (e.g. diesel fuel), and η is the technology conversion efficiency:

$$\frac{\mathbf{P}^+_{n,x,t,c}}{\eta_{n,x,t}} \begin{cases} = R_{n,x,t}, & \text{if} force\_R_{n,x} \text{ is True} \\ \leqslant R_{n,x,t}, & \text{if} force\_R_{n,x} \text{ is False} \end{cases}$$

$$\forall\, (n, x, c) \in (N, X, C)_{supply}$$

Supply technologies with resource given as per unit area, where *force_R* is a boolean dictating whether an available resource has to be consumed when available (e.g. un-curtailed PV) or not (e.g. diesel fuel), and η is the technology conversion efficiency:

$$\frac{\mathbf{P}^+_{n,x,t,c}}{\eta_{n,x,t}} \begin{cases} = R_{n,x,t} \times \mathbf{R}^{\mathbf{area}}_{n,x}, & \text{if} force\_R_{n,x} \text{ is True} \\ \leqslant R_{n,x,t} \times \mathbf{R}^{\mathbf{area}}_{n,x}, & \text{if} force\_R_{n,x} \text{ is False} \end{cases}$$

$$\forall\, (n, x, c) \in (N, X, C)_{supply}$$

Storage technologies, where S_{loss} is the storage standing loss rate, *res* is the timestep resolution and η is the charge/discharge efficiency:

$$\mathbf{S}_{n,x,t} = \mathbf{S}_{n,x,t-1} \times (1 - S_{loss_{n,x,t}})^{res_t}$$

$$- \mathbf{P}^-_{n,x,t,c} \times \eta_{n,x,t} - \frac{\mathbf{P}^+_{n,x,t,c}}{\eta_{n,x,t}} \quad \forall\, (n, x) \in (N, X)_{store}$$

A.2.4. Cost
The cost decision variables given in Eq. [(1)](#) are auxiliary variables, calculated from the summation of decision variables given above.
Investment cost, where *cost* refers to the cost associated with each capacity and *AF* is the annualisation factor, $\frac{rate^{int}}{1 - (1 + rate^{int})^{L_{n,x}}}$, to scale the investment cost of each technology to one year based on its lifetime L and interest rate $rate^{int}$ (10% in this study):

$$\mathbf{cost}_{\mathbf{invest}} = AF \times \sum_{n,x} (\mathbf{S}^{\mathbf{cap}}_{n,x} \times cost_{S^{cap}_{n,x}} +$$

$$\mathbf{P}^{\mathbf{cap}}_{n,x} \times cost_{P^{cap}_{n,x}} + \mathbf{R}^{\mathbf{area}}_{n,x} \times cost_{R^{area}_{n,x}} + \mathbf{Y}_{n,x} \times cost_{Y_{n,x}})$$

Operational cost, where *cost* refers to the cost associated with carrier production, consumption or export. Note that $cost_{ex}$ is negative, to signify a revenue from export:

$$\textbf{cost}_{\textbf{operate}} = \sum_{n,x,c,t} (\textbf{P}^+_{n,x,c,t} \times cost_{prod_{n,x,t}} +$$

$$\textbf{P}^{\textbf{ex}}_{n,x,c,t} \times cost_{ex_{n,x,t}} - \textbf{P}^-_{n,x,c,t} \times cost_{con_{n,x,t}})$$

Appendix B. Technology characteristics

See Table B.7.

Table B.7
Technology characteristics used in Cambridge/Bangalore case study models. Where there is information overlap between Bangalore and Cambridge, (B) and (C) has been appended, respectively. HTP = heat to power ratio, CTP = cooling to power ratio.

	Efficiency or COP	Capacity	Cost		Other
			Capacity	*Operation*	
NGB	0.82	2 MWp	2,024 GBP + 35.3 GBP/kWp	0.025 GBP/kWh	
PV	0.85[a]	N/A	3.4×10^4INR + 3.6×10^4INR/kWp (B)	0	7 m^2/kWp
			1,500 GBP + 1,000 GBP/kWp (C)		
ST	1[a]	N/A	1,200 GBP + 700 GBP/kWp	0	1 m^2/kWp
CHP	0.405	25 MWp	4.65×10^4 GBP + 703 GBP/kWp	0.029 GBP/kWh[b]	HTP: 0.83
B-CCHP	0.2	100 MWp	4×10^6INR + 5.6×10^4INR/kWh[c]	2,260INR/kWh	CTP: 2.1
D-CCHP	0.45	100 MWp	4×10^5INR + 2.2×10^4INR/kWh[c]	230INR/kWh	CTP: 0.7
DG	0.45	100 kWp	1.44×10^4INR/kWh	230INR/kWh	
AC	3	1 MWp	2.3×10^6INR + 9,510INR/kWh	8INR/kWh	
ECh	5	100 MWp	1.02×10^4INR/kWh	8.25INR/kWh[c]	
GSHP	4.07	22 MWp	2,520 GBP + 4,221 GBP/kWp	0.095 GBP/kWh	
StoreE	1	5 MWh	1,670 GBP + 350 GBP/kWp	0	charge rate: 0.7
StoreT	1 (B)	100 MWh (B)	3×10^3INR/kWp (B)	0	charge rate: 0.5 (B)
	0.9 (C)	10 MWh (C)	527 GBP + 65.7 GBP/kWp (C)		charge rate: 0.3 (C)
					heat loss: 0.01 (C)
Distribution					
Electricity	1	20 MWp (B)	0	0	
		25 MWp (C)			
Thermal	1 (B)	20 MWp (B)	731INR/kWp/m (B)	0	
	2.5×10^{-3} loss/m (C)	25 MWp (C)	294 GBP + 281 GBP/kWp/m (C)		
Gas	1	25 MWp	1.7×10^4 GBP + 10 GBP/kWp/m	0	

[a] Solar energy conversion efficiency has already been accounted for in the timeseries resource data for PV and ST. For PV, 0.85 refers to inverter efficiency.
[b] CHP can export to the grid, using a time variable rate equal to 80% of the wholesale electricity prices during 2015.
[c] There is an additional cost of 5×10^7INR for construction of an energy centre, which is incurred when at least one energy centre technology is chosen.

References

[1] Morvaj B, Evins R, Carmeliet J. Optimising urban energy systems: simultaneous system sizing, operation and district heating network layout. Energy 2016;116, Part 1:619–36. https://doi.org/10.1016/j.energy.2016.09.139. ISSN 0360-5442.
[2] Li L, Mu H, Li N, Li M. Economic and environmental optimization for distributed energy resource systems coupled with district energy networks. Energy 2016;109:947–60. https://doi.org/10.1016/j.energy.2016.05.026. ISSN 0360-5442.
[3] Omu A, Choudhary R, Boies A. distributed energy resource system optimisation using mixed integer linear programming. Energy Policy 2013;61:249–66. https://doi.org/10.1016/j.enpol.2013.05.009. ISSN 03014215.
[4] Jennings M, Fisk D, Shah N. Modelling and optimization of retrofitting residential energy systems at the urban scale. Energy 2014;64:220–33. https://doi.org/10.1016/j.energy.2013.10.076.
[5] Mehleri ED, Sarimveis H, Markatos NC, Papageorgiou LG. A mathematical programming approach for optimal design of distributed energy systems at the neighbourhood level. Energy 2012;44(1):96–104. https://doi.org/10.1016/j.energy.2012.02.009. ISSN 0360-5442.
[6] Buoro D, Pinamonti P, Reini M. Optimization of a distributed cogeneration system with solar district heating. Appl. Energy 2014;124:298–308. https://doi.org/10.1016/j.apenergy.2014.02.062. ISSN 0306-2619.
[7] Haikarainen C, Pettersson F, Saxén H. A model for structural and operational optimization of distributed energy systems. Appl Therm Eng 2014;70(1):211–8. https://doi.org/10.1016/j.applthermaleng.2014.04.049. ISSN 1359-4311.
[8] Fonseca JA, Schlueter A. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. Appl Energy 2015;142:247–65. https://doi.org/10.1016/j.apenergy.2014.12.068. ISSN 0306-2619.
[9] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew Sustain Energy Rev 2009;13(8):1819–35. https://doi.org/10.1016/j.rser.2008.09.033. ISSN 1364-0321.
[10] Girardin L, Marechal F, Dubuis M, Calame-Darbellay N, Favrat D. EnerGis: A geographical information based system for the evaluation of integrated energy conversion systems in urban areas. Energy 2010;35(2):830–40. https://doi.org/10.1016/j.energy.2009.08.018. ISSN 03605442.
[11] The Carbon Trust. Closing the gap: lessons learned on realising the potential of low carbon building design. Tech. Rep., The Carbon Trust; 2011.
[12] Pazouki S, Haghifam M-R, Planning Optimal. Optimal planning and scheduling of energy hub in presence of wind, storage and demand response under uncertainty. Int J Electrical Power Energy Syst 2016;80:219–39. https://doi.org/10.1016/j.ijepes.2016.01.044. ISSN 0142-0615.
[13] Ahn Y, Han J. Economic optimization of integrated network for utility supply and carbon dioxide mitigation with multi-site and multi-period demand uncertainties. Appl Energy 2018;220:723–34. https://doi.org/10.1016/j.apenergy.2018.02.129. ISSN 0306-2619.
[14] Yokoyama R, Ito K. Optimal design of energy supply systems based on relative robustness criterion. Energy Convers Manage 2002;43(4):499–514. https://doi.org/10.1016/S0196-8904(01)00027-9. ISSN 0196-8904.
[15] Mavromatidis G, Orehounig K, Carmeliet J. Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. Appl Energy 2018;214:219–38. https://doi.org/10.1016/j.apenergy.2018.01.062. ISSN 0306-2619.
[16] Rezvan AT, Gharneh NS, Gharehpetian GB. Optimization of distributed generation capacities in buildings under uncertainty in load demand. Energy Build 2013;57:58–64. https://doi.org/10.1016/j.enbuild.2012.10.031. ISSN 0378-7788.

[17] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16(6):3586–92. https://doi.org/10.1016/j.rser.2012.02.049. ISSN 1364-0321.

[18] Reinhart CF, Cerezo Davila C. Urban building energy modeling – A review of a nascent field. Build Environ 2016;97:196–202. https://doi.org/10.1016/j.buildenv.2015.12.001. ISSN 0360-1323.

[19] Braun MR, Altan H, Beck SBM. Using regression analysis to predict the future energy consumption of a supermarket in the UK. Appl Energy 2014;130:305–13. https://doi.org/10.1016/j.apenergy.2014.05.062. ISSN 0306-2619.

[20] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl Energy 2014;123:168–78. https://doi.org/10.1016/j.apenergy.2014.02.057. ISSN 0306-2619.

[21] Gamou S, Yokoyama R, Ito K. Optimal unit sizing of cogeneration systems in consideration of uncertain energy demands as continuous random variables. Energy Convers Manage 2002;43(9):1349–61. https://doi.org/10.1016/S0196-8904(02)00020-1. ISSN 0196-8904.

[22] Yang Y, Zhang S, Xiao Y. Optimal design of distributed energy resource systems based on two-stage stochastic programming. Appl Therm Eng 2017;110:1358–70. https://doi.org/10.1016/j.applthermaleng.2016.09.049. ISSN 1359-4311.

[23] Zhou Z, Zhang J, Liu P, Li Z, Georgiadis MC, Pistikopoulos EN. A two-stage stochastic programming model for the optimal design of distributed energy systems. Appl Energy 2013;103:135–44. https://doi.org/10.1016/j.apenergy.2012.09.019. ISSN 0306-2619.

[24] 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. ISSN 0306-2619.

[25] Pickering B, Choudhary R. Applying piecewise linear characteristic curves in district energy optimisation. In: Proceedings of the 30th international conference on efficiency, cost, optimisation, simulation and environmental impact of energy systems, San Diego, USA, July 2017, p. 1080–92.

[26] Sun Y, Closing the building energy performance gap by improving our predictions. Doctor of Philosophy, Georgia Institute of Technology; 2014.

[27] Ward R, Choudhary R, Heo Y, Rysanek A. Exploring the impact of different parameterisations of occupant-related internal loads in building energy simulation. Energy Build 2016;123:92–105. https://doi.org/10.1016/j.enbuild.2016.04.050. ISSN 0378-7788.

[28] Cambridge City Council. Cambridge west site, planning application 16/1134/OUT, Cambridge City Council; 2016.

[29] Ramsay J, Silverman BW. Functional data analysis. Springer series in statistics. 2 ed.New York: Springer-Verlag; 2005. ISBN 978-0-387-40080-8.

[30] Ward R, Choudhary R, Heo Y, Guillas S. Data driven bottom-up approach for modelling internal loads in building energy simulation using functional principal components. In: Proceedings of the 3rd IBPSA-England Conference BSO 2016, Great North Museum, Newcastle, 2016, p. 1115–22.

[31] Lu D, Bao Z, Li Z. Load sampling for Scuc based on principal component analysis and kernel density estimation. In: 2016 IEEE Power and energy society general meeting (PESGM), 2016, p. 1–5. https://doi.org/10.1109/PESGM.2016.7741946.

[32] Ward R, Choudhary R, Aston J. A Stochastic data-centric model for quantification of end-use energy demand in buildings. In: Proceedings of BSO, 2018, p. 307–14, URL http://www.ibpsa.org/proceedings/BSO2018/3C-2.pdf.

[33] Turlach BA. Bandwidth selection in kernel density estimation: a review. Tech. Rep. Technical Report 9317, C.O.R.E. and Institut de Statistique; Universite Catholique de Louvain, URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.6770.

[34] Refaeilzadeh P, Tang L, Liu H. Cross-validation. Encyclopedia of database systems Boston, MA: Springer; 2009. p. 532–8. https://doi.org/10.1007/978-0-387-39940-9_565.

[35] Conejo AJ, Carrión M, Morales JM. Decision making under uncertainty in electricity markets vol. 1. Springer; 2010.

[36] Römisch W. Scenario reduction techniques in stochastic programming. Watanabe O, Zeugmann T, editors. Stochastic algorithms: foundations and applications, vol. 5792. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 1–14. https://doi.org/10.1007/978-3-642-04944-6_1. ISBN 978-3-642-04943-9 978-3-642-04944-6.

[37] Mavromatidis G, Orehounig K, Carmeliet J. Design of distributed energy systems under uncertainty: a two-stage stochastic programming approach. Appl Energy 2018;222:932–50. https://doi.org/10.1016/j.apenergy.2018.04.019. ISSN 0306-2619.

[38] Good N, Mancarella P. Flexibility in multi-energy communities with electrical and thermal storage: a stochastic, robust approach for multi-service demand response. IEEE Trans Smart Grid 2017;PP(99):1. https://doi.org/10.1109/TSG.2017.2745559. ISSN 1949-3053.

[39] Bruninx K, Delarue E. Scenario reduction techniques and solution stability for stochastic unit commitment problems. In: 2016 IEEE international energy conference (ENERGYCON), 2016, p. 1–7. https://doi.org/10.1109/ENERGYCON.2016.7514074.

[40] Maurovich-Horvat L, Rocha P, Siddiqui AS. Optimal operation of combined heat and power under uncertainty and risk aversion. Energy Build 2016;110:415–25. https://doi.org/10.1016/j.enbuild.2015.11.009. ISSN 0378-7788.

[41] Pfenninger S, Pickering B. Calliope: a multi-scale energy systems modelling framework. J Open Source Software 2018;3(29):825. https://doi.org/10.21105/joss.00825.

[42] Meurant G. Algorithms and complexity. Elsevier; 2014. ISBN 978-0-08-093391-7.

[43] Rotman B. Will the digital computer transform classical mathematics?, Philosophical transactions of the royal society of London A: mathematical. Phys Eng Sci 2003;361(1809):1675–90. https://doi.org/10.1098/rsta.2003.1230. ISSN 1364-503X), 1471-2962.

[44] Goderbauer S, Comis M, Willamowski FJL. The synthesis problem of decentralized energy systems is strongly NP-hard, repORt (preprint) 2018 (043); 2018, 16.

[45] Rockafellar RT, Uryasev S. Conditional value-at-risk for general loss distributions. J Bank Finance 2002;26(7):1443–71. https://doi.org/10.1016/S0378-4266(02)00271-6. ISSN 0378-4266.

[46] Zugno M, Morales J, Madsen H. Commitment and dispatch of heat and power units via affinely adjustable robust optimization. Comput Operat Res 2016;75:191–201. https://doi.org/10.1016/j.cor.2016.06.002.

[47] Zhou Y, Wei Z, Sun G, Cheung K, Zang H, Chen S. A robust optimization approach for integrated community energy system in energy and ancillary service markets. Energy 2018;148:1–15. https://doi.org/10.1016/j.energy.2018.01.078.

[48] Lee JH, Kim H, Song Y-h. A study on verification of changes in performance of a water-cooled VRF system with control change based on measuring data. Energy Build 2018;158:712–20. https://doi.org/10.1016/j.enbuild.2017.10.014. ISSN 0378-7788.

[49] Xia J, Winandy E, Georges B, Lebrun J. Testing methodology for VRF systems. In: Proceedings of IRACC 2002, vol. Paper 542, Purdue University, 9, 2002.

[50] Zmeureanu R. Prediction of the cop of existing rooftop units using artificial neural networks and minimum number of sensors. Energy 2002;27(9):889–904. https://doi.org/10.1016/S0360-5442(02)00027-0. ISSN 0360-5442.

[51] AECOM (Ed.), Spon's mechanical and electrical services price book 2016, CRC Press, 47 har/psc edition edn., 2015, ISBN 978-1-4987-3506-3.

[52] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30. ISSN 1533-7928.

[53] Bergstra J, Yamins D, Cox DD. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th international conference on machine learning, vol. 28, Atlanta, Georgia, USA: JMLR: W&CP; 9, 2013.

[54] Gabrielli P, Gazzani M, Martelli E, Mazzotti M. Optimal design of multi-energy systems with seasonal storage. Appl Energy (ISSN 0306-2619), https://doi.org/10.1016/j.apenergy.2017.07.142.

[55] Kotzur L, Markewitz P, Robinius M, Stolten D. Time series aggregation for energy system design: modeling seasonal storage. Appl Energy 2018;213:123–35. https://doi.org/10.1016/j.apenergy.2018.01.023. ISSN 0306-2619.