Highlights

A Study on the Transferability of Computational Models of Building Electricity Load Patterns across Climatic Zones

Rebecca Ward, Cheryl Sze Yin Wong, Adrian Chong, Ruchi Choudhary, Savitha Ramasamy

- Models of plug loads in office buildings are transferable between countries
- Machine learning approaches can generate stochastic models of electricity demand
- The Autoencoder model generates accurate predictions given sufficient data
- The Functional Data Analysis model gives insight into the data structure

A Study on the Transferability of Computational Models of Building Electricity Load Patterns across Climatic Zones

Rebecca Ward[a,b], Cheryl Sze Yin Wong[c], Adrian Chong[d], Ruchi Choudhary[a,b], Savitha Ramasamy[e]

[a] *The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK*
[b] *Energy Efficient Cities initiative, Cambridge University Engineering Department, Trumpington Street, Cambridge CB2 1PZ, UK*
[c] *Institute for Infocomm Research (I^2R), Agency for Science, Technology and Research, 1 Fusionopolis Way, # 21-01 Connexis, Singapore 138632, Singapore*
[d] *Department of Building, School of Design and Environment, National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore*
[e] *Institute for Infocomm Research (I^2R), A*STAR Initiative for Artificial Intelligence and Analytics (AI3) Agency for Science, Technology and Research (A*STAR), Singapore*

Abstract

Significant reduction in energy demand from non-domestic buildings is required if greenhouse emission reduction targets are to be met worldwide. Increasing monitoring of electricity consumption generates a real opportunity for gaining an in-depth understanding of the nature of occupant-related internal loads and the connection between activity and demand. The stochastic nature of the demand is well-known but as yet there is no accepted methodology for generating stochastic loads for building energy simulation. This paper presents evidence that it is feasible to generate stochastic models of activity-related electricity demand based on monitored data. Two machine learning approaches are used to develop stochastic models of plug loads; an autoencoder (AE) and a functional data analysis (FDA) model. Using data from two office buildings located in different countries, the transferability of models is explored by training the models on data from one building and using the trained models to predict demand for the other building. The results show that both models predict plug loads satisfactorily, with a good agreement with the mean demand and quantification of the uncertainty.

Keywords: Electricity demand, Plug loads, Stochastic model, Machine learning, Transferability, Autoencoder (AE), Functional Data Analysis (FDA)

1. Introduction

Across the world, buildings are responsible for almost 40% of energy- and process- related greenhouse gas emissions and this is currently increasing [1]. To reverse the trend, the IEA Global Status Report for Buildings and Construction calls for prioritisation of actions that will result in a reduction of emissions at a rate of 3% per year going forwards. While decarbonisation of the electricity supply will have a significant impact, reduction in demand is also essential if the targets are to be met.

Different countries have made their own commitments to reducing the energy demand attributable to buildings. In the EU, a target of 27% energy improvement by 2030 is in place [2]. In alignment with this, the UK has a challenging net-zero goal by 2050 [3]. Similar targets are in place across Asia; Singapore is committed to reducing the emission index by 36% from 2005 levels and having 80% green buildings by 2030 [4]. These targets will require significant reduction in demand from non-domestic buildings which currently account for 37% of the total electricity consumption in Singapore [5] and 63% in the UK [6].

To reduce demand, it will be necessary to reverse current trends. Electricity demand is dependent on the internal loads in the space, with contributions from the occupants, lighting and plug loads. Occupancy and occupant behaviour are key, as lighting, plug loads, ventilation and internal heat gains are all dependent to some extent on occupancy levels [7, 8]. Specifically, occupant behavior has been shown to have an impact on a building's lighting and plug loads [9], and indeed occupancy has been shown to be able to be simulated from proxy data such as monitored lighting or plug load electricity consumption [10, 11]. Appliance consumption data mining has also been widely used to learn individual occupant behaviour - enabling targeted demand reduction strategies - and to develop occupant group schedule models, facilitating improved HVAC system design [12]. A significant body of work has been carried out by the IEA (Annex 66) in order to understand and find ways of quantifying and simulating occupant behaviour as it impacts on building energy demand, concluding that data collection is of fundamental

importance for modelling [13]. As a consequence, the subsequent IEA Annex 79 features data-driven methods and highlights as a key challenge the need to develop new techniques, to use new data sources and advanced data analytic methodologies from different fields to inform building design and operation [14].

Building energy simulation (BES) has an important part to play in improving design and quantifying retrofit options - and also for fault identification, where comparison of anticipated and actual performance can identify system failings [15, 16]. Yet incorrect estimation of occupant-based internal loads is one of the principal contributing factors for the poor agreement between observed energy demand and the demand simulated using building energy simulation tools - the so-called 'performance gap' [17, 18]. Diversity schedules or profiles from standardised libraries and national standards are most often used to represent occupancy, lighting, and plug loads in BES tools. A comprehensive review of these national standards [19] makes recommendations for updating of these outdated and simplistic schedules, and to explore mechanisms for including local contextual factors that reflect and update societal trends such as homeworking. As an example, Figure 1 shows the schedules assumed for plug loads in office buildings in 5 countries worldwide according to their national standards. The schedules are prescribed, deterministic and make no attempt to quantify the associated stochasticity in the demand or to reflect the individuality of activities within the buildings: the differences arising from offices that have activities that need power hungry devices over others, for example. As a result, they may not be truly representative. In an example using occupancy data from a large multi-tenanted office building, Duarte et al. [20] showed that diversity factors could differ by as much as 46% from those used in standardized schedules. Similarly, Lamano et al. [21] observed plug load intensities for the Nanyang Technological University in Singapore of $11W/m^2$, which are significantly lower than the standard value of $16W/m^2$. There is a real need to learn from observed demand and to develop models for use in design and retrofit that more accurately quantify the uncertainty in the expected demand.

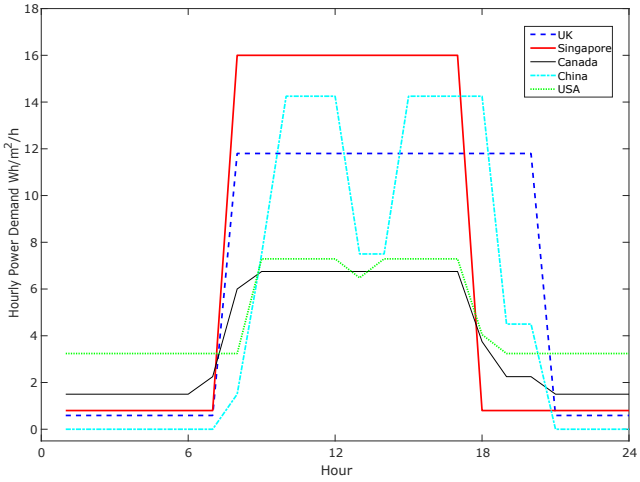


Figure 1: National standard schedules for plug loads to be used in building energy simulation for office buildings

To date there has been no consensus or collective attempt to replace the national standard schedules with stochastic data models that can enable the use of transient and stochastic schedules for energy predictions. Whilst it is true that no two buildings will have *exactly* the same characteristics and occupants [17], one can find standard sets of activities and operations that occur in different types of building (eg. open plan office space, class room, IT office, cafeteria, etc.). In fact, even the national standard schedules are typically based on average estimated consumption pattern per dominant activity in a building. One can take the argument further and even claim that in this global world, the energy demand per activity would be consistent insofar it is not influenced by local climate (i.e. daylighting and HVAC).

As buildings become more information intensive, there is a real opportunity for borrowing strength across different studies and building up stochastic data models of energy demand in buildings that can be representative of underlying activities they represent. Such models could not only be used to generate reference values of occupant related electricity demand, but also help understand the range of variability in the demand. Indeed, this variability is likely the reason behind the difference in amplitude of plug loads in offices seen in Figure 1. An important distinction to be made here is that our proposition is *not* to build a data repository of demand and apply it directly to other buildings. Indeed, it is not appropriate to take data directly

from one building and apply it to another building. Our proposition is to develop appropriate data models of activities and test their applicability to correctly generate occupant-related demand for zones in other buildings that have similar activities. Whereas regular machine learning and data mining techniques study the training data under the major assumption that the test data are collocated or have the same distribution as the training data, in this study we aim to test whether we can train a model on data from a building in one location to generate sample demand for a building in a different location. We are not performing *transfer learning*, in which a model is trained on data then the model is adapted before application to a different scenario; instead we seek to take the trained model and apply it directly to the new scenario - this is termed *model transferability*.

This paper tests two data models that can facilitate transferability of models of occupant-related electricity demand in office buildings. Specifically, we consider the plug loads of two buildings in different parts of the world with similar functions - both in a broad sense are university buildings - with the aim of transferring representations of energy demand for activities from one building to another. We consider the buildings to be similar as they house the same type of activity, a university office. Yet the first building is located in Cambridge, UK and the second in Singapore. Having two buildings in very different locations allows us to test if the nature of standard activities, and demand for energy thereof, across buildings is consistent i.e. whether the assumption of 'similarity' is valid. It is convenient therefore to consider plug loads for three reasons: (a) they are, in most cases, directly controlled by occupants and thus a good indicator of occupant influence on demand, (b) they are entirely activity driven, and (c) they are less influenced by local climate or micro-climate, which would be the case for lighting or HVAC demand.

We develop and test two different modelling approaches; an autoencoder (AE) and a functional data analysis (FDA) model. The AE model is a good choice for correlated input data and creates a compressed representation of the data that retains the correlation structure. The FDA model, by comparison, deconstructs the data into an expanded representation that allows interpretation of the significance of different features of the data. For each approach, a model has been trained using a training dataset from certain zones (hereafter called training zones). The AE and the FDA models are then applied to predict stochastic samples of the demand for the test zones. The model predictions are compared against the test data to ascertain the

extent to which the AE and FDA models are able to replicate important features of the test data. In this way we explore the transferability of the activity models for plug loads across buildings and even across an international context.

In the following section, the literature relating to transferability of models relevant to the context of this study is briefly described. Section 3 outlines the modelling methodologies and the approach used for this study. The results of application of the approach to the real world data are then described and the implications discussed in Sections 4 to 6, with detailed discussion of the data model structures provided in Appendix A. Finally conclusions are drawn and recommendations made for further work.

2. Related Works

The use of machine learning techniques in the building energy simulation field is relatively new. While there have been recent studies exploring transfer learning i.e. the use of models trained on one problem as the starting point of a related problem, such as learning for HVAC control [22] and urban building energy modelling [23], to date we are not aware of any research into the direct transferability of models from one building to another. However, studies have been conducted across several other domains towards exploring transferability of machine learning models across geographical locations, notably species distribution, aspects of climate science and land use including contamination. For example, Duque-Lazo et al. [24] explored the probability of occurrence of species in different geographies by modeling their statistical relationships with environmental variables. The study involved comparison of 10 regression models, of which the Gaussian Linear Model and Generalized Additive Model performed better. However, it was reported that the MaxEnt statistical model provided robust transferability in predicting occurrence of species between Spain and Australia. A study on the temporal transferability of species distribution models in the marine environment [25] showed that machine learning models are more sensitive than regression-based models.

In the climate science domain, although research in assessing wind resources using machine learning approaches is still emerging, geographical transferability of wind estimation has been studied by Veronesi et al. [26] using linear regression and decision tree models. The transferability of local climate zone mapping to regional and global scales through random forest algorithms has been explored by Tong et al. [27]. The study noted that

transferability of climate zones can be successful with a good representative training set. However, a study on transferability of spectroscopic diagnosis models towards estimation of arsenic contamination in soil [28] showed that transfer component analysis provides better estimation over a large area at low cost.

Deep neural networks have also been explored as potential candidates for transferability of representations across geographical locations. Specifically, it has been shown that convolutional neural networks are capable of better transferability for land cover classification [27], which can help to resolve the challenges due to increased complex information and the data disturbances caused by increased spatial resolution and different image acquisition conditions. The latent representation of a Deep Belief network has been exploited towards transferability of ocean characteristics to predict wave height across varied geographical regions by Kumar et al. [29]. The challenges in transferability of models across geography and time have been well summarized by L.Yates et al. [30].

As transfer component analysis and deep neural networks have shown some success in transferability of natural phenomena (climate, land cover, soil contamination etc), in this study we endeavour to explore the application of a functional data analysis model and a deep learning autoencoder model to spatial and temporal transferability of plug load distribution across distinct geographical regions.

3. Modelling methodologies

As aforementioned, we have selected two different data model approaches, namely Autoencoder and Functional Data Analysis. The two approaches differ in the way that the data are used and in their interpretability. The AE model produces a compressed representation of the data in the latent dimension. This is an efficient representation of the data and the model is quick to train. Test data are encoded using the weights learned from the training data, and the encoded representations are used to generate predictions. As an unsupervised machine learning tool, it is necessary to learn from a training dataset and predict from a test dataset as otherwise the model will simply replicate the data used for training; but prediction is even quicker than the training. Provided the features of the test dataset exist in the training data, the AE model gives a faithful representation of the test dataset and predictions agree closely with the test data.

The FDA model, by comparison, deconstructs the data into an expanded representation that aims to uncover the underlying functions from which the data are composed. The initial training of the model is time consuming and computer intensive; alignment of the data and extraction of the phase relationships between data samples is a lengthy process. However, having aligned the data, extraction of the functional principal components and scores is quick. Given the time-consuming nature of the initial data alignment process, it is not ideal to generate a new model for each new building considered. Instead, it is more efficient to map new data to an existing set of PCs, such as those generated from the training data. This is quick to do and facilitates direct comparison between datasets.

The two models are outlined in more detail below.

3.1. Autoencoder (AE)

An autoencoder [31] is a type of artificial neural network for unsupervised representational learning. Figure 2 provides an overview of the autoencoder used in this problem. It generally consists of an encoder and a decoder. The encoder compresses the data into an encoded representation in the latent dimensions and the decoder reconstructs the original data. This network can be trained by minimizing the reconstruction error with a loss function, which measures the differences between the original input and the consequent reconstruction. Therefore, the objective of the autoencoder is to optimize its representation of the input through its low dimensional latent dimension. Let us assume the input is given by $\mathbf{x} \in \Re^m$, where m is the dimension of the input. Let the encoder be given by $f(\mathbf{x})$ and the decoder be given by $g(f(\mathbf{x}))$, the autoencoder then minimizes the loss function (in this case the root mean square error):

$$\mathcal{L}\left(\mathbf{x}, g(f(\mathbf{x}))\right) \tag{1}$$

In the context of this study, the weights in the autoencoder (encoder and decoder) are learnt during the training phase and remain unchanged in the testing phase. In the testing phase, the encoder is first used to encode the testing data for a particular activity. These encoded latent representations are then utilized to build a multivariate Gaussian distribution, from which a total of 1000 samples are drawn. The new samples are then put into the decoder to reconstruct the predicted plug load profiles.

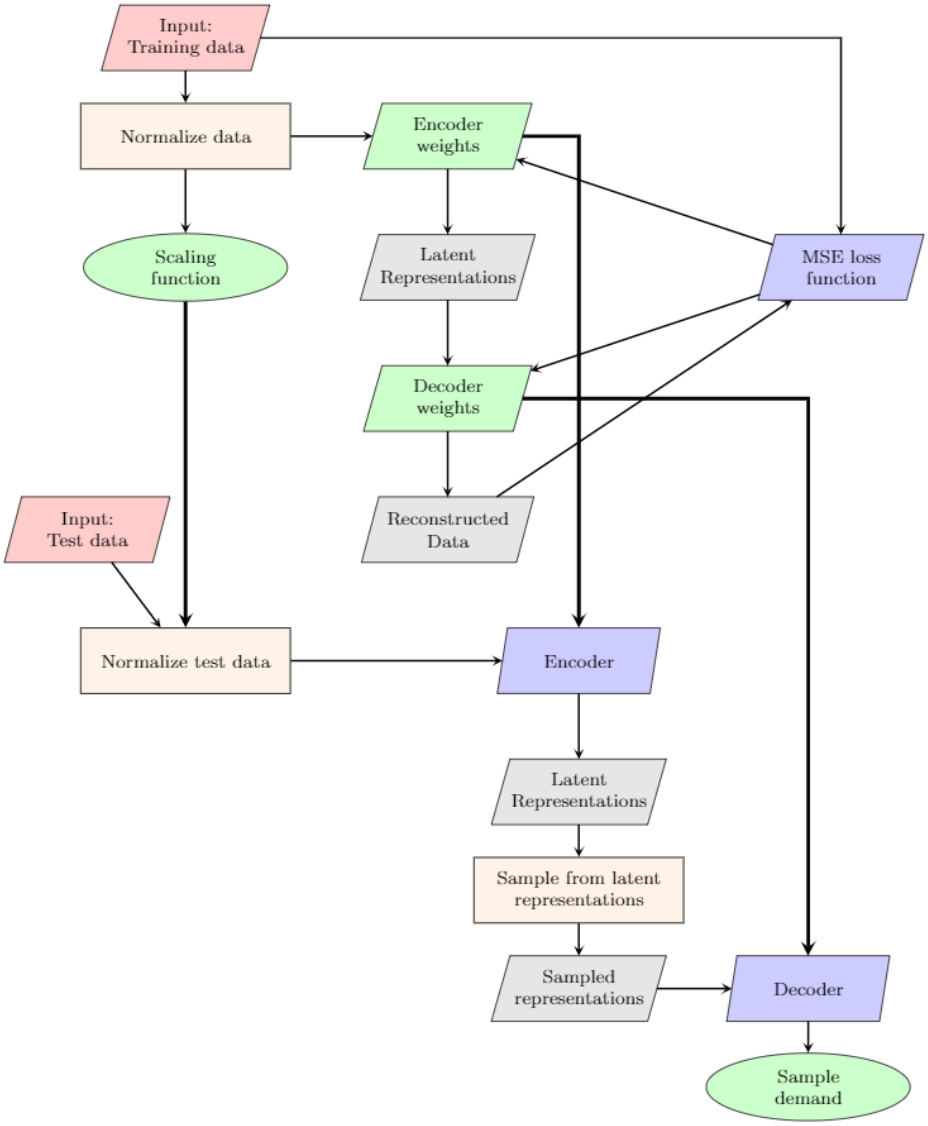


Figure 2: Schematic of the autoencoder process. Thicker lines here indicate that the scaling function, encoder weights and decoder weights that are used/trained in the training phase are brought over for inference in the testing phase

3.2. Functional Data Analysis (FDA)

An alternative methodology that offers insight into patterns in the data is Functional Data Analysis (FDA), a set of techniques for analysing data that represent an underlying continuous function [32]. Hourly data points are assumed to be indicative of an underlying continuous function of time, and for this study each set of 24 hourly values representing one day constitutes a data sample. Given a set of data samples, functional Principal Component Analysis (fPCA) is used to derive a set of *functional* Principal Components (PCs) - each one a function of time - and a set of coefficients or *scores* on the PCs that control how much each PC contributes to each data sample. The PCs are same across all data samples, so it is the PC scores that distinguish one data sample from another. Each data sample, $\phi(t)$ can therefore be expressed simply as a sum of a mean function, $\mu(t)$ and the weighted summation over i functional PCs, $\nu_i(t)$, with scores α_i, i.e.

$$\phi(t) = \mu(t) + \Sigma_i \alpha_i \nu_i(t) \tag{2}$$

It is useful in the analysis of transient energy demand to separate the data into its phase and amplitude, i.e. the timing component and the magnitude of the demand, as both are quantities of interest. In order to do this, the data are aligned to a common mean using a standard dynamic programming algorithm [33], giving a set of warping functions that align each data sample to the mean and a set of aligned amplitude functions. fPCA is then performed on the warping and amplitude functions separately, yielding separate PCs for phase and amplitude that are the same across all data samples. This procedure also yields a unique set of scores on the PCs for each data sample. The fPCA methodology ensures that the order of the PCs corresponds to the amount of variation in the data that is explained by each PC, in the same way as standard PCA for multivariate data.

The FDA model thus comprises the phase and amplitude PCs together with the scores that can be used in conjunction with the PCs to regenerate the original data - the training zones. If we want to generate new sample data for the training zones, it suffices to take a random sample of the scores and use the sampled scores with the PCs to generate sample data. It is necessary to fit a multivariate probability distribution to the scores to ensure correlations are captured - here we use a copula as the scores are not normally distributed. If we wish to generate sample data for new building zones we have two options - in the absence of any data we would have to assume that

the demand of the new zones is sufficiently similar to the training zones that we can use the scores for the training zones directly. However if we have data, as here - even a small quantity - we project the data onto the PCs and extract the scores; we then take a random sample of the new scores and generate new samples with the sampled scores and the PCs. This facilitates direct comparison between the test data and the training data. This process is illustrated schematically in Figure 3. Full details of the model rationale and development are given in Ward et al. [34].

3.3. Approach

Monitored plug-load data, in particular plug loads sub-monitored for individual building zones representing different activities within university buildings are studied in this paper. A subset of the data is used to train the AE and FDA models. Then, taking new data - a test dataset, which in this study is data for a new zone - the models are used in conjunction with the test data to predict energy demand for plug loads of the new zone. The procedure for training the data is slightly different for the two models. In the FDA model, the data are normalised prior to training by subtracting the median base load and dividing by the median load range on a zone-by-zone basis. The training dataset is aligned and PCs extracted following the methodology described in Section 3.2, then the test data are mapped to these PCs and scores extracted. To predict sample plug loads for the new zones, random samples are drawn from a probability distribution fitted to these scores and are combined with the PCs to generate the sample data. In the AE model, the AE builds a latent dimension representation of the training data as described in Section 3.1, together with an encoder and decoder. Various architectures of autoencoder have been tested and the AE with only one layer of 8 hidden nodes, which also represents the latent dimension, was found to be the best in modelling the data. For hyperparameter tuning, it is found that using batch sizes of 16,32,64,128 with learning rates of 0.1, 0.01, 0.001 yielded similar results. The AE model is trained in batches of 32 with the mean square error (MSE) cost function using the Adam optimizer with a learning rate of 0.001. Normalisation of the data occurs within the AE and not prior to the training process. The test data are then encoded using the encoder and the encoded latent representations are used to build a multivariate Gaussian distribution from which sample data are drawn. These samples are then put through the decoder to reconstruct sample load profiles.

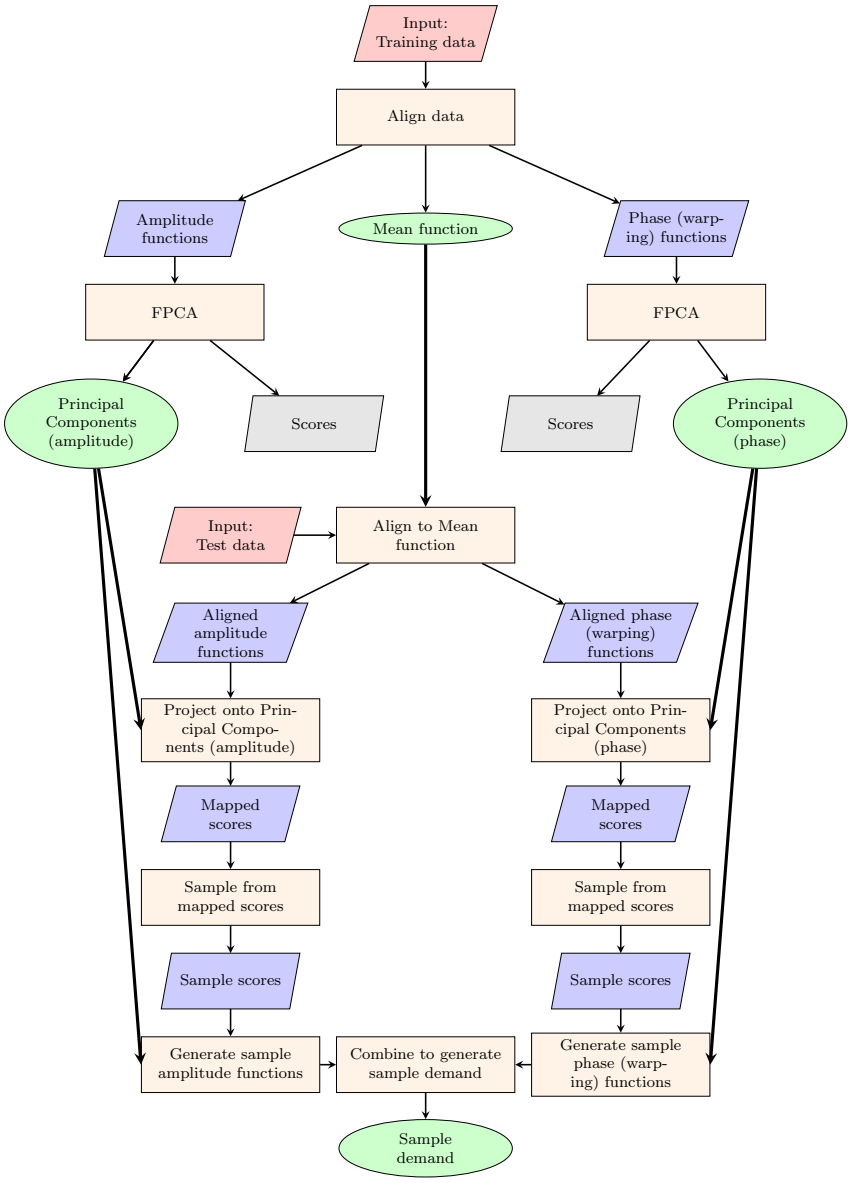


Figure 3: Schematic diagram of FDA process. The thicker lines indicate how the outputs of the training phase are used in conjunction with the test data

The predicted data are then compared against the test dataset using three comparison tools to identify to what extent the predicted data are comparable to the test data. First, a visual comparison is performed in order to identify whether the AE and FDA models are generating good predictions. Second, it is illuminating to compare certain key performance indicators (KPIs) to identify to what extent the models are providing a faithful representation of the data. The KPIs considered here are the mean and standard deviation of the parameters which describe the aspects of the demand that are important for building energy simulation and energy management, i.e.

- base load, i.e. the minimum hourly demand per day, typically the demand at night,
- peak load, i.e. the maximum hourly demand per day, important for sizing of the HVAC system,
- total daily demand, important for cost and total demand estimation, and
- the timing of the daily peak, important for load scheduling and demand response management.

Finally, another measure of how well the generated samples represent the observed data is given by the Kullback-Leibler (K-L) divergence value, calculated using the approach given in Duchi [35] for two multivariate Gaussian distributions. The KL divergence between two multivariate Gaussians can be calculated as follows:

$$KL = \frac{1}{2}[\log \frac{|\Sigma_2|}{|\Sigma_1|} - Tr\{\Sigma_2^{-1}\Sigma_1\} + (\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1) - d] \qquad (3)$$

where Σ_1 and Σ_2 represent the determinants of the diagonal covariances of the two distributions respectively. μ_1 and μ_2 are the mean of the two multivariate Gaussian and d is number of variables in the Gaussian distribution. This is a measure of how well one probability distribution represents another; it is non-negative and values closer to zero imply better agreement. We use the K-L divergence to assess how well the predicted plug loads match the test datasets.

4. Data Description

Monitored plug loads from two university buildings in two distinct locations, namely the University of Cambridge in the United Kingdom (UK) and the National University of Singapore (SG), have been used for this study, as detailed in Table 1.

Table 1: Building Data

	UK	SG
Location	Cambridge	Singapore
Climate	Temperate	Tropical
Number of floors	3	3
Number of sub-metered zones	15	4
Floor Area m^2	10,241	5445
Use	Research office	Research office
Date of construction	2001	2014

The UK and SG data comprise hourly recorded plug loads from 15 and 4 spatial zones, respectively. The terminology used to identify the zones throughout this paper uses the location - UK or SG for UK and Singapore - together with an indication of the activity within the zone. For example, UK:O1 is an office in the UK building. Space-use or activity abbreviations are as follows: O - office, M - meeting room, C - classroom, LT - lecture theatre, IT - IT lab, L - Library and K - canteen kitchen.

4.1. UK Data

The UK building is a 3-storey research office at the University of Cambridge. The building is sub-metered for plug loads and lighting for 15 spatial zones, each of which is designated by activity e.g. office, meeting room, canteen, or user e.g. student office, administrative staff office. As aforementioned, we first use plug loads from 10 of these spatial zones for training the model. These data are illustrated in Figure 4. As there are differences in the plug load patterns between weekdays and weekends, we indicate the mean and 90% confidence limits for the weekday (red) and weekend (blue) for each zone. It can be observed from the figure that the plug loads vary quite diversely despite similarity in use types or activities. The plug loads in most of the zones peaks during the middle section of the day on weekdays. It is advantageous to have a wide range of demand profiles in the training

dataset so that the models can learn possible different variations of demand. It is for this reason that we include both the weekend and weekday data in training dataset 1.

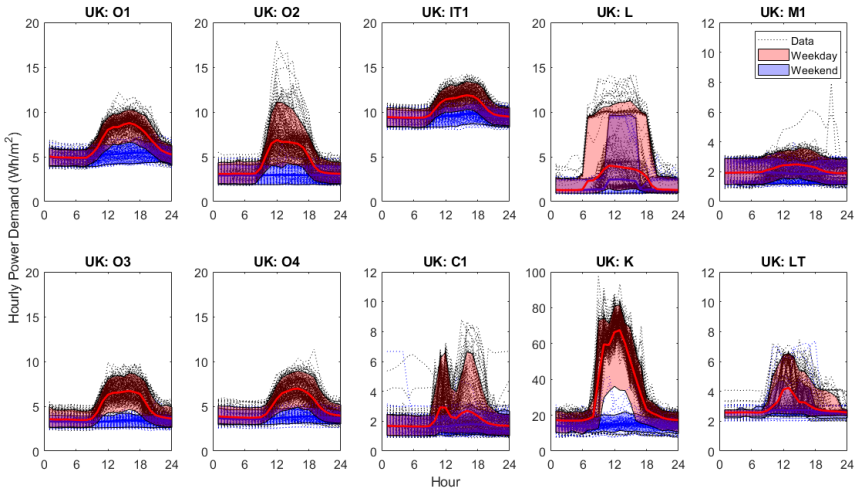


Figure 4: UK training data (dataset 1): daily monitored plug loads for 1 year for 10 different spatial zones showing the 90% confidence limits for weekdays (red) and weekends (blue)

For the validation of the models, we use a smaller subset of the data, comprising term-time weekday data from the 5 remaining spatial zones (dataset 2). A consistent subset of the data is used, rather than all of the data as the consistency makes it more straightforward to assess the quality of the predictions using the K-L divergence as a metric. The hourly plug loads data for these zones are shown in Figure 5. It can be observed from these patterns that they exhibit some similarities to the training dataset. For example, the weekday data for UK:O5 and UK:O1 are quite similar. Similarly, weekday data of UK:IT1 and UK:IT2 are also quite similar, albeit with a slightly higher base load in UK:IT2.

4.2. SG Data

The SG building is a research office located on the campus of the National University of Singapore. The building consists of 3 blocks, each of which is 3 storeys high. There is no heating but the HVAC system operates year round to provide cooling. Sub-metered data for 4 zones have been used for this

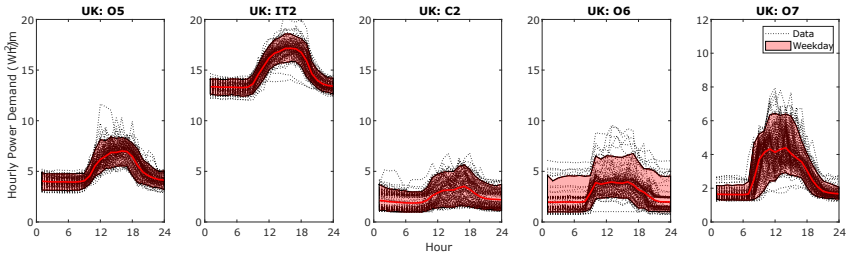


Figure 5: UK validation data (dataset 2): term-time weekday daily monitored plug loads for 5 different spatial zones showing the 90% confidence limits (red)

analysis. Unlike the UK building, the sub-metering is at the distribution board and hence each zone is a mixture of different activities.

Similar to the UK validation dataset, the SG test dataset comprises term-time weekday data as illustrated in Figure 6. The data are fairly consistent, particularly zone SG: O2 which exhibits only a small degree of daily variation from the mean.

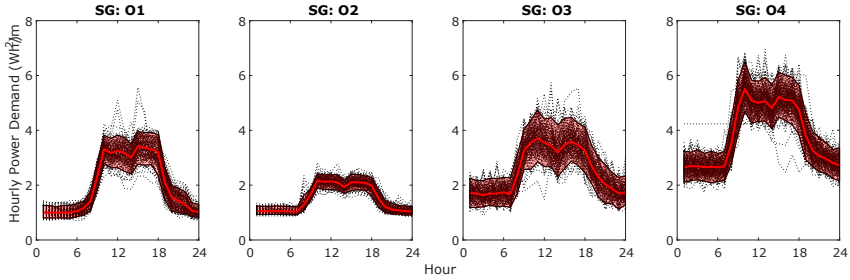


Figure 6: SG test data (dataset 3): term-time weekday daily monitored plug loads for 4 different spatial zones showing the 90% confidence limits (red)

4.3. Model Setup

As the UK data is the larger dataset, comprising sub-metered data from 15 different zones, we first train the models using data from 10 of the UK building zones (dataset 1). Three cases are then examined:

1. we perform a validation study in which the applicability of the models to new data from the same building is assessed using a subset of data from the 5 remaining spatial zones of the UK building (dataset 2). This provides a benchmark against which to compare subsequent analyses.

2. the models are then used to predict plug loads for a test dataset comprising the 4 SG zones (dataset 3).
3. finally we reverse the analysis and use dataset 3 as training data to train new models thus exploring the ability of the new models to predict the plug loads of UK validation zones (dataset 2).

A visual comparison of the data presented in Figures 4 - 6 suggests there are similarities, but it is not clear which zones are the most similar. The strength of the AE and FDA models is that they both provide a means to explore the similarities in more detail. This is discussed further in the following sections and in detail in Appendix A, but as an example consider the projection of the latent dimensions from the AE model to a 2D t-Distributed Stochastic Neighbor Embedding (t-SNE) plot [36] shown in Figure 7. In the figure, the red points represent the training data from all zones. The data points in various zones are clustered together, indicating that the plug load profiles in the given zones possess similar characteristics. For the UK validation zones, the UK:IT2 data (turquoise - on the far right of the plot) do not overlap with the training data at all. This suggests that the plug loads in this zone are quite dissimilar to the training data (dataset 1). For the SG test data, all 4 zones do not really overlap with the training data, but are quite centrally located in the plot, and zone SG:O3 overlaps with zone UK:O7. The implications of this comparison are discussed further in Section 6 and Appendix A.

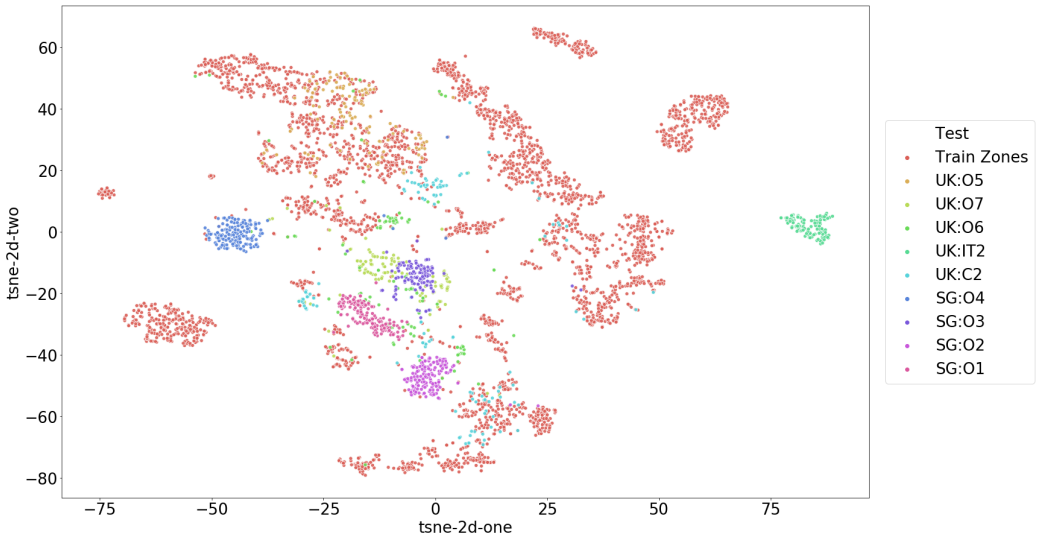


Figure 7: t-SNE: Training Data (red) compared against the UK Validation Zones and the SG Test Zones

5. Modelling Results

As discussed above, the prime output of the models is predicted plug load profiles for each zone of the building. These are presented in this section, first for the validation study, then for the test case. The K-L divergence and the KPIs described in Section 3.3 are also calculated and discussed for both the validation study and the test case.

5.1. Profiles

Considering first the validation study, Figure 8 shows the the validation data in red (dataset 2), the predicted plug loads generated using the AE model in blue and the predicted plug loads generated using the FDA model in green. In these plots, the solid line is the sample mean and the shaded area is the 90% confidence limit.

It is clear that the predicted plug loads from the AE model fit dataset 2 well in terms of the mean profile shape, with the exception of UK:O7. In this zone, the predicted mean profile shape seems to increase early, around 5am and plateaus for a couple of hours before increasing again, but in the actual data the profile increases steadily starting from around 7am. If we examine the training data (Figure 4), it is clear that the only zone which

exhibits a similar start time to that observed for zone UK:O7 is zone UK:L, which has the early start followed by a plateau in the load before rising again. This suggests that the AE model is associating the early start time with the following plateau and hence the predictions follow this pattern. The mean demand of the FDA predicted samples is also in good agreement with the test data, but the rise at the start of the day is smoother in the predicted load profiles than the test data. For example, in zone UK:O6 at around 8am a sharp rise can be seen in the red solid line which is somewhat more shallow in the green line. This is due to the hourly discretisation of the FDA model; a more refined model would be able to match sudden changes in demand better, but with a corresponding increase in computational time.

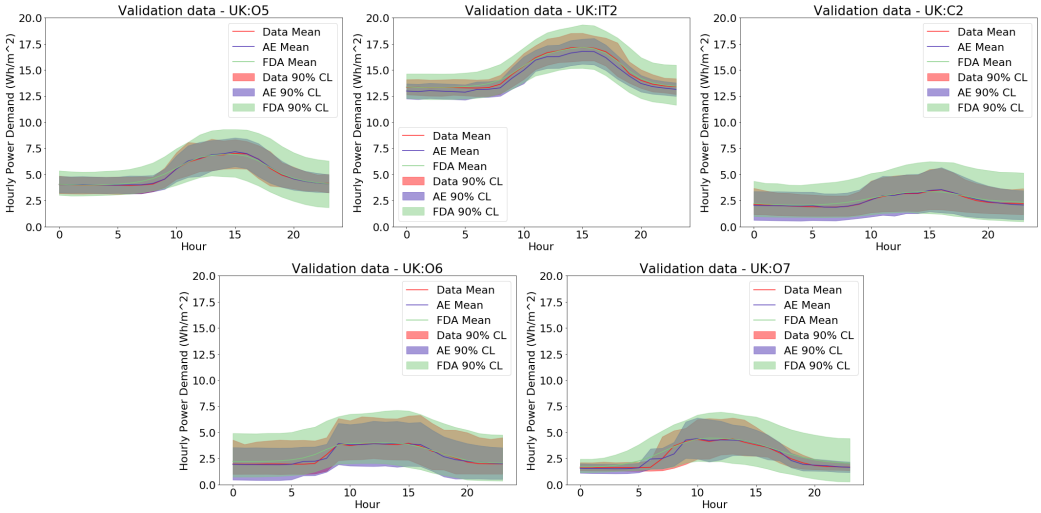


Figure 8: UK validation dataset: comparison of samples generated using the autoencoder (blue) and FDA (green) models against the validation dataset 2 (red)

The variability of the AE predictions as indicated by the 90% confidence limits - the blue shaded areas in Figure 8 - for zones UK:O5 and UK:C2 match the data very well, with hardly any of the red shaded area visible. For the remaining three zones there are areas where the AE model variability is lower than the data, particularly for zone UK:O6 for which the red shaded area is visible at the top of the range across all hours. This is due to the distribution of the test data for zone UK:O6 being more asymmetric about the mean than the majority of the data on which the AE model is trained.

By comparison, and as is typical, the 90% confidence limits of the FDA

predictions (shown in green) cover a wider range than dataset 2. It is particularly noticeable for zone UK:O7, where the predictions show a wide variance at the end of the day, in strong contrast to the test data that are tightly defined at the start and end of the day. The FDA model depends on a weighted summation of the PCs, and features such as the drop at the end of the day arise due to complex correlations of the PC weightings for opposing PC contributions. These correlations are taken into account in sampling from the PC scores, but as one might expect there are samples for which opposing contributions do not fully cancel. Furthermore, in fitting a probability distribution to empirical data there are inevitably sparsely populated areas within the probability distribution. Sampling from these areas of the distribution gives rise to combinations of PC not seen in the data and a wider distribution of predictions.

The predictions by both models visually indicate that both machine learning approaches work for the validation zones in the same building. Do they work for prediction of plug loads in the Singapore building?

Figure 9 shows the plug load profiles (90% confidence limit) of the SG test dataset (dataset 3) in red and the predicted data generated using the AE model in blue and the FDA model in green. In general, it can be observed that the samples predicted from the AE model do still fit the original data, however, the agreement is not as good as for the validation dataset. Similar to zone UK:O7 the predicted profiles for all four SG zones do not capture the gradual increase at the start of the day but rise earlier than the data and plateau before rising again. Again, this is due to the lack of profiles that have an early start time in training dataset 1. The variability of the AE predictions is a similar order to the data with the exception of zone SG:O3 for which the AE model predicts a much lower variability in base load. Considering the FDA predictions shown in green in Figure 9, the sharp rise at the start of day is smoother in the predictions than in the test data. When comparing Figures 8 and 9 it appears that the start of day, i.e. the time at which demand starts to rise, is earlier in the SG building than the UK building. This is particularly noticeable for zones SG:O3 and SG:O4 where the start of the working day appears to occur just after 6am on average. Despite being based on data for the UK building, the FDA model can mimic this, albeit somewhat smoothed. The variability of the FDA model predictions is again greater than the data, with a wide variance at the end of the day, particularly for zone SG:O4.

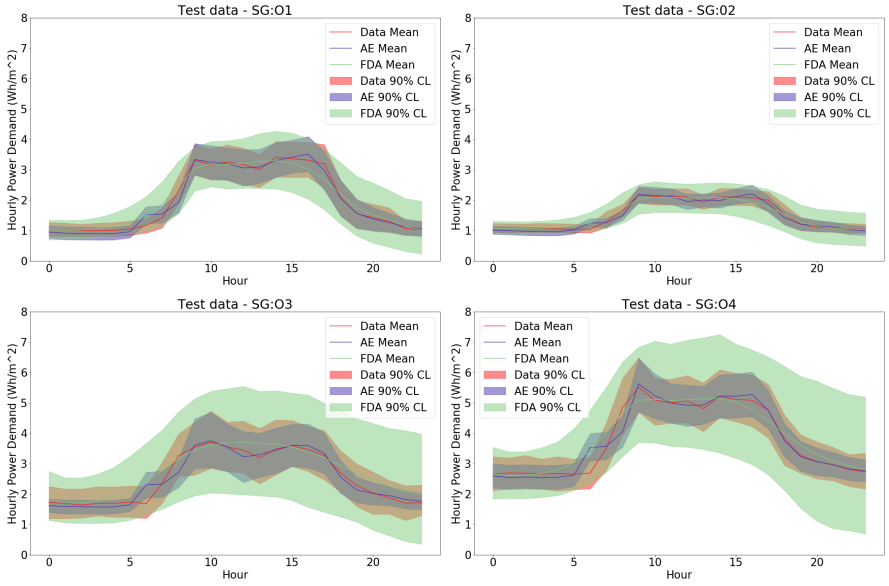


Figure 9: SG test dataset: comparison of samples predicted using the autoencoder (blue) and FDA (green) models against test dataset 3 (red)

The difference between the observed and predicted datasets can be quantified statistically using the K-L divergence. Table 2 shows the K-L divergence values for the samples predicted by each of the models compared against the data for the validation/test zones.

For the UK validation data, the lowest values of KL divergence for the AE model are zone UK:C2, closely followed by zone UK:O5. As observable in the profiles in Figure 8, the mean and area of overlap between the data and the model predictions are more closely matched for these two zones than zones UK:O7 and UK:IT2 which have higher K-L divergence values. Similarly, in Figure 9, comparing zones SG:O1 and SG:O2, which have lower K-L divergence values against zones SG:O3 and SG:O4 which have higher K-L divergence values, the lower values correspond to a visually better match between the data and predictions. This gives confidence that the K-L divergence values do reflect the similarities and differences between the data and predicted plug load profiles.

For the UK validation data, the FDA model also generates the lowest K-L value for zone UK:C2 but with a value of 1.02, higher than the AE value of 0.07, reflecting the higher variance of the FDA model results compared against the AE model. However, for the FDA model the K-L divergence

Table 2: K-L divergence (the maximum values are highlighted in bold font, whereas the minimum values are in italics)

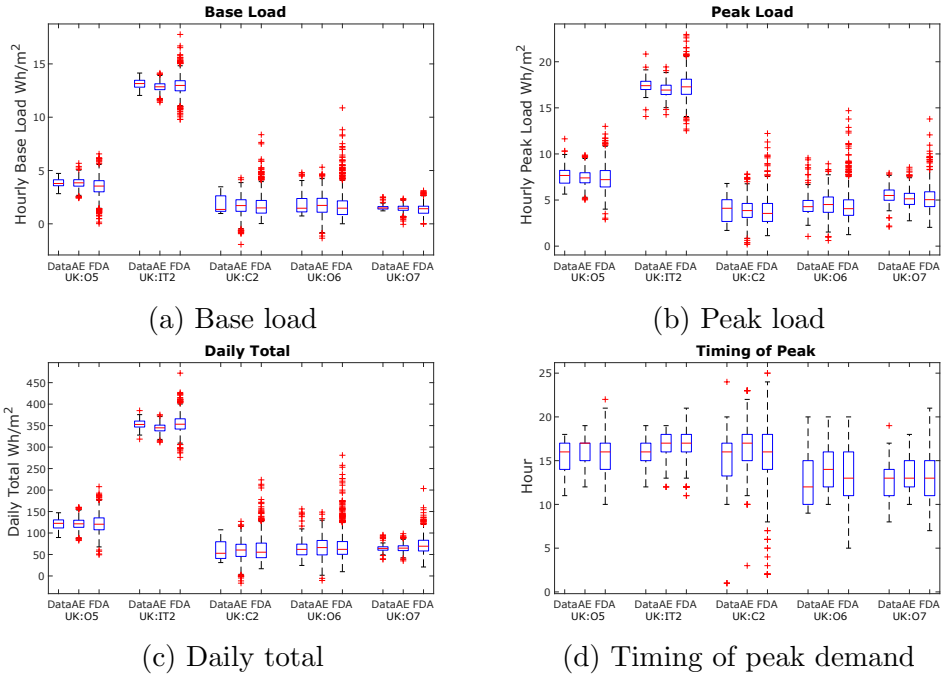
Zone	Autoencoder	FDA
UK:O5	0.09	1.82
UK:IT2	**3.04**	1.74
UK:C2	*0.07*	*1.02*
UK:O6	0.25	1.50
UK:O7	1.47	**3.42**
SG:O1	1.33	*2.84*
SG:O2	*1.17*	2.86
SG:O3	**3.97**	3.23
SG:O4	3.26	**3.90**

suggests a better fit for zone UK:IT2 than zone UK:O7 (1.74 for zone UK:IT2 compared against 3.42 for zone UK:O7) in direct contrast to the results for the AE model. Considering Figure 8, it is clear that the FDA model predictions for zone UK:IT2 are closer to the original data than for zone UK:O7, whereas the opposite is true for the AE model. This is due to differences in normalisation of the data; whereas for the FDA model the data are normalised relative to the zone median base load and load range prior to model development, in the AE model the normalisation occurs within the AE training process. This means that the AE model is less able to predict for zones that lie outside the training data, giving a higher K-L value for zone UK:IT2 which has a base load higher than any of the training data.

The predicted data show good agreement with the validation dataset for both models, and this is reflected in the low values of the K-L divergence. For the SG test data the agreement is slightly less good, with correspondingly higher K-L divergence values. In particular the FDA model does not represent the dip in demand around hour 13 which is clearly observable in the data for all four zones. But the highest K-L divergence value is observed for the AE model for zone SG:O3 which has a value of 3.97. This is due to the inability of the AE model to accurately capture the early morning demand - again, owing to the absence of these features in the training data.

5.2. KPIs

As described in Section 3.3, while a visual comparison of the predicted demand profiles and the monitored data can give good insight into the applicability of the models, it is the KPIs that give quantifiable measures that



(a) Base load (b) Peak load

(c) Daily total (d) Timing of peak demand

Figure 10: KPIs for the UK validation dataset compared against samples generated using the autoencoder (AE) and functional data analysis (FDA) models

are most relevant to the building energy simulator.

Tables 3(a) and (b) and Figure 10 show a comparison of KPIs derived from the UK validation dataset 2 and from the predictions generated from each model. In Table 3(a), the mean values of each of the KPIs are given for both models, together with the percentage difference between the mean of the models and the data. This is equivalent to the Normalised Mean Bias Error (NMBE), which is one of the metrics recommended in ASHRAE Guideline 14 for comparison of simulated and monitored building energy demand. In this instance, we are not making predictions in the same sense, but instead are generating demand that is typical of the building zone in question. Hence the acceptability criteria of ASHRAE Guideline 14 are not strictly applicable here [37]. Nonetheless, we can use the guideline as an indication - if we consider a 10% difference between the mean of the models and the data to be an acceptable limit, both models give good results; the only KPI for which this threshold is exceeded is the base load for zone UK:O7 predicted by the FDA model which has a % difference of -15.48%. This is due to the wide

Table 3: KPIs for UK validation data (dataset 2): a) mean, and b) standard deviation (the maximum absolute difference is highlighted in bold font, whereas the minimum absolute difference is in italics)

a) Mean

			Autoencoder		FDA	
	Zone	Data	Model	% Difference	Model	% Difference
Base Load	UK:O5	3.80	3.83	*0.71*	3.51	-7.76
	UK:IT2	13.14	12.85	-2.23	12.97	-1.29
	UK:C2	1.79	1.69	**-5.88**	1.75	-2.70
	UK:O6	1.78	1.73	-2.75	1.81	*1.27*
	UK:O7	1.54	1.46	-5.28	1.30	**-15.48**
Peak Load	UK:O5	7.61	7.41	-2.62	7.35	-3.31
	UK:IT2	17.45	16.95	-2.88	17.33	*-0.69*
	UK:C2	3.93	3.84	-2.33	3.79	-3.61
	UK:O6	4.53	4.52	*-0.16*	4.45	-1.67
	UK:O7	5.51	5.17	**-6.31**	4.98	**-9.65**
Time of Peak Demand	UK:O5	15.24	15.34	0.69	14.94	-1.94
	UK:IT2	15.92	15.99	*0.43*	15.98	*0.38*
	UK:C2	15.24	15.64	2.57	14.55	-4.56
	UK:O6	12.76	13.33	**4.43**	12.42	-2.66
	UK:O7	12.64	12.27	-2.89	11.83	**-6.39**
Total Daily Demand	UK:O5	120.51	121.36	0.71	122.34	1.52
	UK:IT2	352.99	344.57	**-2.39**	354.47	*0.42*
	UK:C2	59.15	59.36	0.35	62.60	5.83
	UK:O6	66.16	66.22	*0.09*	71.82	**8.54**
	UK:O7	64.84	64.67	-0.25	67.67	4.37

b) Standard Deviation

			Autoencoder		FDA	
	Zone	Data	Model	% Difference	Model	% Difference
Base Load	UK:O5	0.43	0.47	9.20	0.93	115.76
	UK:IT2	0.44	0.42	*-5.20*	0.87	93.99
	UK:C2	0.73	0.84	15.40	1.08	*48.40*
	UK:O6	0.87	0.95	8.80	1.33	51.63
	UK:O7	0.24	0.29	**20.66**	0.60	**147.87**
Peak Load	UK:O5	1.06	0.81	**-23.40**	1.36	27.82
	UK:IT2	0.82	0.71	-12.81	1.34	**62.66**
	UK:C2	1.31	1.16	-10.87	1.54	17.55
	UK:O6	1.44	1.21	-16.28	1.67	15.23
	UK:O7	0.95	0.90	*-5.94*	1.10	*14.73*
Time of Peak Demand	UK:O5	1.70	1.58	-7.14	2.00	17.52
	UK:IT2	1.40	1.33	*-4.71*	1.50	7.07
	UK:C2	3.23	2.32	**-28.15**	3.73	**15.34**
	UK:O6	3.12	2.57	-17.65	2.73	-12.44
	UK:O7	2.28	1.94	-14.54	2.43	*6.81*
Total Daily Demand	UK:O5	12.25	12.58	2.72	21.38	73.88
	UK:IT2	10.56	9.69	**-8.21**	20.22	**90.73**
	UK:C2	20.32	21.27	4.66	28.77	*41.03*
	UK:O6	23.49	23.75	*1.12*	33.97	44.08
	UK:O7	9.04	8.94	*-1.12*	16.78	84.84

variation exhibited by the FDA model predictions at the end of the day for this zone, as also observed in Figure 10(a). The good match in the prediction of the peak load is particularly encouraging as it is an important feature for the design of HVAC systems as well for the management of demand-side response. In addition, the timing of the peak demand is significant as it can have cost implications arising from real-time energy pricing. The mean timing for both models is in good agreement with the data, with a maximum difference of 4.43 % for the AE model (zone UK:O6) and 6.39% for the FDA model (zone UK:O7). To put this in context, the decrease of 6.39% is a change from 12.38pm to 11.50am to, i.e. a difference of just over 45 minutes. Timing of the peak is generally very difficult to simulate precisely. Indeed, as the data illustrate (Figure 4), daytime demand is highly variable and peaks may not be consistent or identifiable.

The FDA model in general has a tendency to give a higher variability in the KPIs, as detailed in Table 3(b), which shows the standard deviation of each KPI, together with the % difference between the standard deviation of the simulations and the data. This can also be observed in Figure 10. In terms of the magnitude of the demand, i.e. KPIs Base Load, Peak load and Total Daily Demand, the FDA model gives a higher standard deviation for the KPI than the AE model. The increase in standard deviation is particularly high for the base load of zones UK:O5 and UK:O7, a consequence of the variability observable in Figure 8 at the end of the day for these zones. As explained previously, the tendency for the FDA model to overestimate the variability is due to samples being drawn from a probability distribution fitted to the PC scores which may encompass space for which there are no data. This leads to the relatively high number of outlying values as observed in Figure 10; as an example, considering the peak load for zone UK:O6 shown in Figure 10(b) it seems clear that the higher standard deviation calculated for the FDA approach is due to the higher number of outlying points. The AE and FDA models exhibit opposite tendencies in the prediction of the variability of the peak load - the FDA model overestimates while the AE underestimates the standard deviation of the peak load for all zones. While we expect the FDA model to overestimate the variability, the way in which the AE model learns from the training data means that the samples predicted by the AE model necessarily lie within the model's experience i.e. it is less likely to predict outlying values than the FDA model, and hence is more likely to underestimate the variability.

Tables 4(a) and (b) show the corresponding results for the SG test zones,

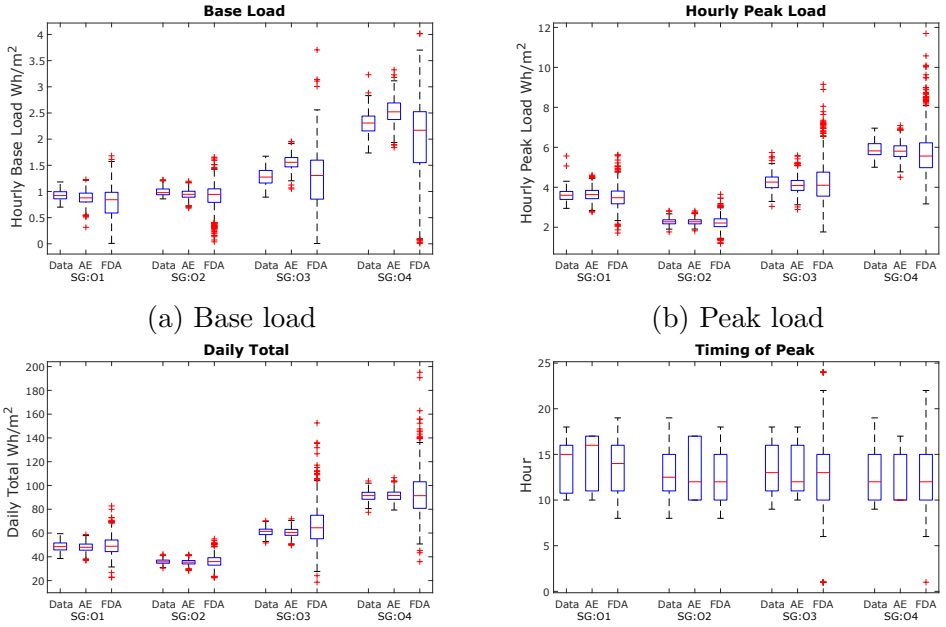
Table 4: KPIs for SG test zones (dataset 3): a) mean, and b) standard deviation (the maximum absolute difference is highlighted in bold font, whereas the minimum absolute difference is in italics)

a) Mean

			Autoencoder		FDA	
	Zone	Data	Model	% Difference	Model	% Difference
Base Load	SG:O1	0.93	0.88	*-4.65*	0.79	**-14.98**
	SG:O2	1.00	0.95	-5.05	0.91	-8.83
	SG:O3	1.28	1.56	**21.59**	1.22	*-4.52*
	SG:O4	2.31	2.53	9.70	2.01	-12.63
Peak Load	SG:O1	3.61	3.64	0.90	3.47	-3.96
	SG:O2	2.27	2.27	*0.34*	2.22	*-1.98*
	SG:O3	4.28	4.10	**-4.11**	4.10	-4.27
	SG:O4	5.89	5.82	-1.27	5.59	**-5.12**
Time of Peak	SG:O1	14.05	14.40	2.51	13.25	**-5.72**
Demand	SG:O2	12.98	13.63	5.08	12.42	-4.26
	SG:O3	13.25	13.06	*-1.42*	12.61	-4.80
	SG:O4	12.76	12.01	**-5.87**	12.29	*-3.71*
Total Daily	SG:O1	48.61	48.07	-1.09	48.89	0.59
Demand	SG:O2	36.05	35.42	**-1.75**	36.12	*0.20*
	SG:O3	61.00	60.55	-0.73	63.63	**4.30**
	SG:O4	91.34	91.34	*-0.01*	91.61	0.29

b) Standard Deviation

			Autoencoder		FDA	
	Zone	Data	Sample	% Difference	Sample	% Difference
Base Load	SG:O1	0.10	0.12	**19.37**	0.31	196.44
	SG:O2	0.08	0.08	*1.17*	0.22	*175.72*
	SG:O3	0.15	0.13	-11.58	0.57	**278.52**
	SG:O4	0.22	0.23	6.12	0.76	253.00
Peak Load	SG:O1	0.34	0.30	-10.93	0.48	*43.39*
	SG:O2	0.16	0.15	-7.11	0.30	85.54
	SG:O3	0.45	0.39	**-14.59**	0.91	99.82
	SG:O4	0.41	0.39	*-3.24*	0.95	**133.25**
Time of Peak	SG:O1	2.84	3.03	6.82	2.69	-5.14
	SG:O2	2.53	3.15	**24.61**	2.42	-4.23
	SG:O3	2.82	2.82	*0.02*	3.21	**13.89**
	SG:O4	2.79	2.74	-1.76	2.76	*-1.18*
Total Daily	SG:O1	3.85	3.70	-3.92	7.30	*89.16*
	SG:O2	2.19	2.09	**-4.53**	4.87	121.51
	SG:O3	3.69	3.76	*1.77*	14.65	**295.51**
	SG:O4	4.42	4.50	1.85	16.21	265.87



(a) Base load

(b) Peak load

(c) Daily total

(d) Timing of peak demand

Figure 11: KPIs for the SG test dataset compared against samples generated using the autoencoder (AE) and functional data analysis (FDA) models

with the results illustrated in Figure 11. Again using a difference of 10% as an acceptable limit, there are instances where the error in the mean results for the base load fall above this value. Specifically, the AE model predictions generate a mean base load higher than the data by 21.59% for zone SG:O3 and the FDA models generates mean base loads lower than the data by 14.98% and 12.63% respectively for zones SG:O1 and SG:O4. As described above for the validation dataset, the mean base loads generated by the FDA model are affected by the propensity for the model to generate a wider range of demand at the end of the day than observed, reducing the mean base load. For the AE model, it is clear from Figure 9 that the predicted data for zone SG:O3 are much less variable at the start and end of day i.e. in the base load, than the data. This is corroborated in Table 4(b) which indicates a reduction in standard deviation of the base load when compared against the data for this zone.

The differences in base load prediction are not reflected in the mean total daily demand which shows good agreement for both models (<5% difference), and good agreement is also observed for both models for the timing of the peak demand which as discussed previously, is difficult to predict.

Similar to the results for the UK validation dataset, the standard deviation of the model predictions of the SG test data KPIs given in Table 4(b) is much greater for the FDA model than the AE model. This is supported by the visual comparison of Figure 9, in which the wider variability of the FDA model is clearly visible, and Figure 11 which clearly indicates the increased variability of the FDA sample demand. Looking at the SG test data, the standard deviation values are lower than for the UK validation data implying more consistency in the monitored data. Whereas the percentage difference in the results for the AE model predictions of standard deviation are a similar order of magnitude for both datasets, for the FDA model, the standard deviation for all of the magnitude-related KPIs is greater for the SG test data predictions than for the UK validation data predictions (Table 3(b)). This is again due to the sampling of scores from probability distributions fitted to the empirical data - because the data are more consistent, the effect of outlying scores is more significant and sampling from sparsely populated areas of the fitted probability distribution leads to a higher variability in predicted data.

The fundamental differences between the AE and FDA techniques are further investigated by examining the latent dimensions and the functional principal components and scores in more detail. These are discussed in Sec-

tion 6, but for brevity the details are provided in Appendix A.

5.3. Reverse direction study

The results discussed in the previous sections indicate that there is sufficient similarity between the plug loads for the two buildings studied that it is acceptable to use the training dataset from the UK building to develop a model for demand for the Singapore building. If this is the case, it should be equally feasible to build a model based on the Singapore data and apply the model to generate sample demand for the UK building. To test this hypothesis, the test dataset from the Singapore building (dataset 3) has been used to build new AE and FDA models and these models have been used to predict plug load data for the validation zones from the UK building (dataset 2).

Given the purely data-driven characteristics of the autoencoder, a drop in performance is expected when it is trained with a lesser quantity of more uniform data but tested on a diverse range of data. As seen in the spread of data in the t-SNE plots in Figure 7, it can be observed that the range of the SG data is narrower than the range of zones in dataset 2, spread over [-50, 10] on the first axis and [-60, 10] on the second axis. Therefore, we would expect a drop in performance particularly in regions that fall outside of this area - in particular zones UK:O5 and UK:IT2. Figure 12 provides the predicted plug load profiles generated from the sampled data in the latent dimensions. It can be observed that the autoencoder does a poor job reconstructing the profiles of zones UK:O5 and UK:IT2, UK:IT2 in particular. This is because the UK:IT2 data is very different from the data that the autoencoder is trained on. The autoencoder would be able to provide a better generated profile given a more diverse range of training data.

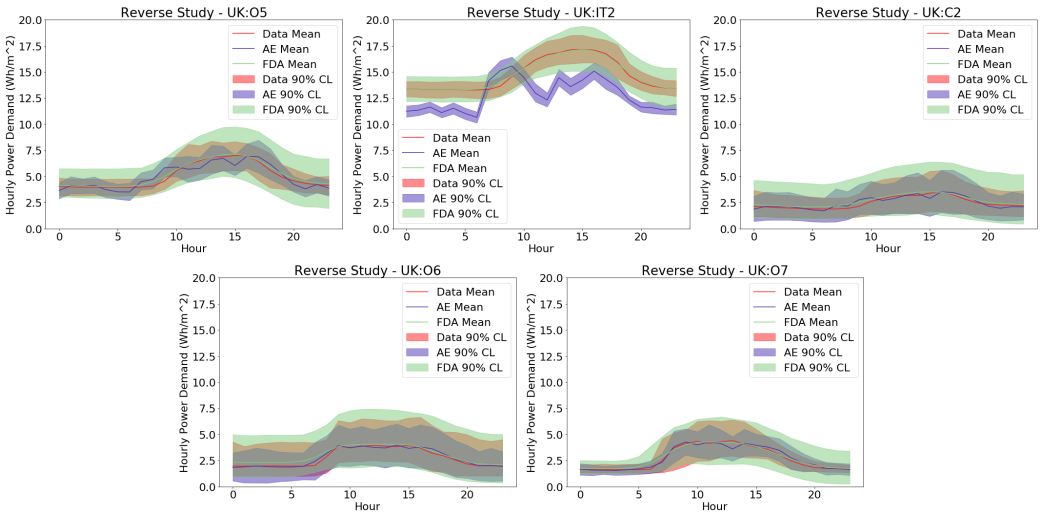


Figure 12: Profile for Reverse Study: Model generated using SG data as training dataset for UK test zones (dataset 2)

A new FDA model has also been generated using the SG data as the training dataset, and a new set of phase and amplitude PCs have been derived. The UK test data are then projected onto the new PCs to extract the scores, and these scores have been sampled from in order to predict new plug loads as before. The predicted plug load profiles are given in Figure 12 in green. Comparing this figure against Figure 8, it is clear that visually there is very little difference in the range of demand generated; this is despite the much smaller quantity of data used as a training dataset.

Table 5 shows the K-L divergence values calculated for the reverse approach. The K-L divergence values again reflect how similar is the simulated profile when compared to the actual profile as seen in Figure 12. For the AE model, the K-L divergence value is exceptionally high for UK:IT2, reflecting the dissimilarity in actual and simulated profiles. The K-L divergence is also high for the UK:O5 zone as it falls outside of the range of the training data used here. For the other zones (UK:C2, UK:O6, UK:O7), the K-L divergence values seem reasonable. For the FDA model values for zones UK:O5, UK:C2 and UK:O6 are slightly higher than before, but for zones UK:IT2 and UK:O7 the values are lower.

Table 5: Model generated using SG data: K-L divergence (the maximum values are highlighted in bold font, whereas the minimum values are in italics)

Zone	Autoencoder	FDA
UK:O5	5.47	2.37
UK:IT2	**175.28**	1.47
UK:C2	0.92	*1.14*
UK:O6	*0.43*	1.59
UK:O7	1.64	**2.50**

6. Discussion

This study enables us to explore the ease and suitability of transferring machine learning models of plug load data from one zone to another zone comprising similar activities, even across different buildings in different climatic zones.

Both the AE and FDA models used here are able to reconstruct the plug load profiles of the SG buildings relatively well (measured by the key performance indicators) when trained on data from the UK building zones. The models both succeed particularly well in predicting the mean magnitude and timing of the peak demand and the mean total daily demand for all four SG zones. The base load is seemingly harder to predict for both models; for the AE model, zone SG:O3 predictions suggest a base load 22% higher than observed, whereas for the FDA model zones SG:O1 and SG:O4 predictions are 13 and 15% lower than observed respectively.

The ability to predict depends to some extent on the similarity between datasets. As explained in detail in Appendix A, the latent dimensions of the AE model and the fPCA scores from the FDA model give separate indications of the similarity of the plug loads across the different zones. Projection of the AE model latent representation onto a 2D t-SNE plot helps to visualise how the different zones relate to each other (Figure 7 and Figures A.14-A.15). In the plots, plug loads from the same zone generally cluster together indicating that the plug load profiles in the zone are similar throughout the dataset. While there are some overlaps - notably zones SG:O3 and UK:O7 - on the whole the clusters remain distinct from each other implying differences in the plug loads between zones. The differences are related more to base load and load range than the shape however. In the FDA model the data are normalised prior to training by subtracting the base load and dividing

by the load range on a zone-by-zone basis. A comparison of the resulting datasets is thus a comparison of the *relative* changes in hourly demand, i.e. the shape of the demand, rather than the absolute magnitude. A detailed exploration of the scores given in Appendix A reveals the similarity between scores for zones with similar activity, for example between zones SG:O1 and SG:O2 and zones UK:O1-O4. This suggests that the shape of the demand is similar for these zones.

This difference in the way the data are processed has implications for the prediction capability of the models. Given similarity in plug loads data from the two buildings despite their different locations, it should be possible to train a model using a training dataset from either building for use in prediction. The study shows success for training the models on data from the UK building, but the reverse process (Section 5.3) highlights the difference between the two approaches. Whereas the FDA approach is able to build a reasonable model using the SG data, the range of data in the SG dataset is insufficient for the AE model to be able to predict zone UK:IT2 accurately, primarily because the base load for this zone is so far outside the range of the training data. The FDA model benefits here from the normalisation of the data as described above which facilitates comparison of shape rather than magnitude, resulting in a more accurate prediction for zone UK:IT2.

The AE model relates the test data to the predicted data via the encoded latent dimensions of the model. Data that are similar in the original dimension tend to be similar in the latent dimension, however as explored in Appendix A, it is difficult to assign physical meaning to the latent dimensions. Each latent dimension contributes to the plug load profile over the entire day, so it is not possible to specifically relate a latent dimension to a distinct feature such as a higher peak load or an earlier start time. As an example, consider the results for the SG test data described in Section 5. In the latent representation, an early start time is clearly associated with a plateau in demand around 7am. This is because there are insufficient samples in the training data with an early start without this plateau, but it means that the predictions for the SG data cannot accurately match the early start and subsequent rise. By comparison, the functional PCs generated by the FDA approach relate directly to features of the data and have a direct interpretation. For example the phase PCs relate directly to the start time and length of the working day. This enables more flexibility to fit features that are not observed in the training data. Interpretation is not always straightforward, however, as different combinations of PCs can give similar results,

exacerbated to some extent by the separation of the plug loads into phase and amplitude which gives even more combination options.

As is clear in the study, while both models give predictions which are in good agreement with the test data, there are differences in the nature of the predictions. The most obvious difference is in the variability of the results. Predicted data are generated from both models by sampling from a probability distribution fitted to the trained model. In the case of the AE model, the latent representations are used to build a multivariate Gaussian distribution from which the samples are drawn, generating accurate predictions closely aligned with the test data with few outlying values. In the FDA approach, the score distributions are fitted to a copula which preserves the non-Gaussian shape of the score distributions and retains correlations between the scores. Sampling from the copula gives scores which are used in conjunction with the PCs to generate the predicted data. This approach gives predicted data with a mean demand in good agreement with the test data. The spread of the results is substantially higher and outlying results are not uncommon, however, owing to samples being drawn from regions of the copula that are sparsely populated. As described above, this is primarily due to different score combinations potentially yielding similar plug load profiles; this is discussed in more detail in conjunction with Figure A.25.

The choice of model for use in building energy simulation is likely dependent on the nature of the simulation being performed and also the level of confidence in the data. For example, in a study of retrofit measures for an existing building it may be essential to have predicted loads as close as possible to those observed - there is a high level of confidence in the demand, and the AE model may be a natural choice. By comparison, where there is more uncertainty - at the design stage for example - the FDA model gives predictions which typically encompass the actual demand and may be a better choice. Given the sample demand profiles it is straightforward to generate a stochastic time history of plug loads for input into a BES. Having demonstrated here the similarity between plug loads in offices, despite their different locations, it is not unreasonable to foresee the development of a set of stochastic schedules based on data for inclusion in national standards and indeed standardisation across the globe. This may be true for other activities - further work is required to assess the transferability of models in this way for different types of common activity and end-use.

7. Conclusions

In summary, we have explored the transferability of models for plug loads in buildings across climatic zones by using data from two case study buildings from Singapore and the UK. Both buildings have similar use type, being research offices located on a university campus, and both are sub-monitored for plug loads.

Two modelling approaches have been explored; an autoencoder and a functional data analysis model. Both models were able to predict plug loads satisfactorily for the SG building when trained on data from the UK building. The study also facilitated an in-depth comparison of the strengths and weaknesses of the two models used in this context. The stand-out differences between the two approaches are as follows:

- the AE model is fast to train and generates predicted plug loads quickly. Initial training of the FDA model is computationally intensive and can take several hours depending on the quantity of data. Once trained, however, mapping new data to the PCs and generating predicted data is quick,
- the latent dimensions of the AE model cannot easily be interpreted in terms of features in the data. By comparison the functional PCs generated for the FDA model relate directly to features such as base load, load range and timing,
- the plug loads predicted by the AE model are very close to the test data in terms of mean and standard deviation provided features of the test data exist in the training data. The FDA model predictions match the mean test data well and typically predict a wider variability than observed in the test data,
- the FDA model does not rely on features of the test data being explicitly present in the training data as the features of the data are constructed from a summation of component functions - this means that fewer data samples are required in order to generate a model adequate for prediction purposes

We have demonstrated that given a sufficient breadth of training data, either model can be used to predict plug loads for these two buildings. Hence

the assumption that the buildings are similar is valid in this respect. The choice of model is likely dependent on the purpose of the simulation; we have considered two possible approaches here, both of which predict plug loads in good agreement with the test data. The feasibility of using plug loads from one location to build a model for prediction of plug loads in other locations opens up the possibility that deterministic schedules defined by location could be replaced by a stochastic modelling approach. This would improve efficiency in generation of models for building energy simulation and ensure greater comparability between simulations. Specifically, a stochastic approach could help to understand the range in variability of the demand and its relation to the activity performed in the building zone in question.

This study has considered plug loads as they are indicative of occupancy, are activity-driven and are not highly influenced by local climate. The results suggest, however, that the machine-learning approach may offer useful insights for transferability of models for other loads such as HVAC and lighting loads. Future studies will consider the wider applicability of the approach, taking into consideration the dependency of such loads on external parameters such as local climate.

Funding

This work was supported by the Lloyd's Register Foundation and The Alan Turing Institute Data-Centric Engineering Programme, together with the EPSRC Strategic Priorities Fund - AI for Science, Engineering, Health and Government, under grant EP/T001569/1.

References

[1] IEA. 2019 Global Status Report for Buildings and Construction. Technical report, United Nations Environment Programme, 2019.

[2] Elena Cuerda, Olivia Guerra-Santin, Juan José Sendra, and Fco. Javier Neila. Understanding the performance gap in energy retrofitting: Measured input data for adjusting building simulation models. *Energy and Buildings*, 209, 2020.

[3] UK House of Commons. Energy efficiency: building towards net zero. Technical Report HC 1730, House of Commons Business, Energy and Industrial Strategy Committee, 2019.

[4] BCA. Super low energy building technology roadmap. *Building and Construction Authority, Singapore*, 2018. URL `https://www.bca.gov.sg/greenmark/others/SLE_Tech_Roadmap.pdf`.

[5] Sushanth Babu, Adrian Lamano, and Priya Pawar. Sustainability assessment of a laboratory building: case study of highest rated laboratory building in Singapore using Green Mark rating system. *Energy Procedia*, 122:751 – 756, 2017.

[6] Office for National Statistics. Energy consumption in the UK (ECUK) 1970 to 2018. Technical report, 2019.

[7] Simone D'Oca and Tianzhen Hong. Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, 88:395–408, 2015.

[8] Da Yan, William O'Brien, Tianzhen Hong, Xiaohang Feng, H Burak Gunay, Farhang Tahmasebi, and Ardeshir Mahdavi. Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107:264–278, 2015. doi: https://doi.org/10.1016/j.enbuild.2015.08.032.

[9] Arsalan Heydarian, Claire McIlvennie, Laura Arpan, Siavash Yousefi, Marc Syndicus, Marcel Schweiker, Farrokh Jazizadeh, Romina Rissetto, Anna Laura Pisello, Cristina Piselli, et al. What drives our behaviors in buildings? A review on occupant interactions with building systems from the lens of behavioral theories. *Building and Environment*, page 106928, 2020. doi: https://doi.org/10.1016/j.buildenv.2020.106928.

[10] Ki-Uhn Ahn, Deuk-Woo Kim, Cheol-Soo Park, and Pieter de Wilde. Predictability of occupant presence and performance gap in building energy simulation. *Applied Energy*, 208:1639 – 1652, 2017.

[11] Yang-Seon Kim, Heidarinejad Mohammed, Matthew Dalhausen, and Jelena Srebric. Building energy model calibration with schedules derived from electricity use data. *Applied Energy*, 190:997 – 1007, 2017.

[12] Jie Zhao, Bertrand Lasternas, Khee Poh Lam, Ray Yun, and Vivien Loftness. Occupant behaviour and schedule modeling for building energy simulation through office aapliance power consumption data mining. *Energy and Buildings*, 82:341 – 355, 2014.

[13] Da Yan, Tianzhen Hong, Bing Dong, Ardeshir Mahdavi, Simona D'Oca, Isabella Gaetani, and Xiaohang Feng. IEA EBC Annex 66: Definition and simulation of occupant behaviour in buildings. *Energy and Buildings*, 156:258 – 270, 2017.

[14] William O'Brien, Andreas Wagner, Marcel Schweiker, Ardeshir Mahdavi, Julia Day, Mikkel Baun Kjaergaard, Salvatore Carlucci, Bing Dong, Farhang Tahmasebi, Da Yan, Tianzhen Hong, H. Burak Gunay, Clayton Miller, and Christiane Berger. Introducing IEA EBC Annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment*, 178, 2020.

[15] Laura E. Fedoruk, Raymond J. Cole, John B. Robinson, and Alberto Cayuela. Learning from failure: understanding the anticipated-achieved building energy performance gap. *Building Research and Information*, 43:6:750 – 763, 2015.

[16] Fu Xiao and Cheng Fan. Data mining in building automation system for improving building operational performance. *Energy and Buildings*, 75:109 – 118, 2014.

[17] Patrick X.W. Zou, Xiaoxiao Xu, Jay Sanjayan, and Jiayuan Wang. Review of 10 years research on building energy performance gap: life-cycle and stakeholder perspectives. *Energy and Buildings*, 178:165 – 181, 2018.

[18] Pieter de Wilde. The gap between predicted and measured performance of buildings: a framework for investigation. *Automation in Construction*, 41:40 – 49, 2014.

[19] William O'Brien, Farhang Tahmasebi, Rune Korsholm Andersen, Elie Azar, Verena Barthelmes, Zsofia Deme Belafi, Christiane Berger, Dong Chen, Marilena De Simone, Simone d'Oca, Tianzhen Hong, Quan Jin, Dolaana Khovalyg, Roberto Lamberts, Vojislav Novakovic, June Young Park, Manfred Plagmann, Vina Subashini Rajus, Marika Vellei, Silke Verbruggen, Andreas Wagner, Eric Willems, Da Yan, and Jin Zhou. An international review of occupant-related aspects of building energy codes and standards. *Building and Environment*, 179, 2020.

[20] Carlos Duarte, Kevin Van Den Wymelenberg, and Craig Rieger. Revealing occupancy patterns in an office building through the use of occupancy sensor data. *Energy and Buildings*, 67:587–595, 2013.

[21] Adrian Sansoldi Lamano, Wu Xiangyu, Zhou Jian, and Bharath Seshadri. Office plug load metering study on NTU campus. Technical report, Nanyang Technological University, 2015.

[22] Shichao Xu, Yixuan Wang, Yanzhi Wang, Zheng O'Neill, and Qi Zhu. One for many: Transfer learning for building HVAC control, 2020.

[23] Alex Nutkiewicz and Rishee K. Jain. Exploring the integration of simulation and deep learning models for urban building energy modelling and retrofit analysis. In *BS 2019, 16th IBPSA International Conference, 2-4th September, Rome*, 2019.

[24] J. Duque-Lazo, H. Van Gils, T. A. Groen, and R. M. Navarro-Cerrilo. Transferability of species distribution models: The case of phytophthora cinnamomi in Southwest Spain and Southwest Australia. *Ecological Modelling*, 320:62–70, 2016.

[25] Camino Fernández de la Hoz, Elvira Ramos, Araceli Puente, and Jose Antonio Juanes. Temporal transferability of marine distribution models: The role of algorithm selection. *Ecological Indicators*, 106 (105499), 2019.

[26] Fabio Veronesi, Athina Korfiati, René Buffat, and Martin Raubal. Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modelling. In *Societal Geo-innovation*, pages 297–310, 2017.

[27] Xin-Yi Tong, Gui-Song Xia, Qikai Lu Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237(111322), 2020.

[28] Chao Tao, Yajin Wang, Wenbo Cui, Bin Zou, Zhengrong Zou, and Yulong Tu. A transferable spectroscopic diagnosis model for predicting arsenic contamination in soil. *Science of The Total Environment*, 669: 964–972, 2019.

[29] N Krishna Kumar, R Savitha, and Abdullah Al Mamun. Ocean wave characteristics prediction and its load estimation on marine structures: A transfer learning approach. *Marine Structures*, 61:202–219, 2018.

[30] Katherine L.Yates, Phil J.Bouchet, M. Julian Caley, Kerrie Mengersen, Christophe F.Randin, Stephen Parne, Alan H. Fielding, Andrew J.Bamford, Stephen Ban, A. MárciaBarbosa, Carsten F.Dormann, Jane Elith, Clare B.Embling, Gary N.Ervin, Rebecca Fisher, Susan Gould, Roland F.Graf, Edward J. Gregr, Patrick N.Halpin, Risto K.Heikkinen, Stefan Heinänen, Alice R.Jones, Periyadan K.Krishnakumar, Valentina Lauria, HectorLozano-Montes, Laura Mannocci, Camille Mellin, Mohsen B.Mesgaran, Elena Moreno-Amat, Sophie Mormede, Emilie Novaczek, Steffen Oppel, Guillermo Ortuño Crespo, A. Townsend Peterson, Giovanni Rapacciuolo, Jason J.Roberts, Rebecca E.Ross, Kylie L.Scales, David Schoeman, Paul Snelgrove, Göran Sundblad, Wilfried Thuiller, Leigh G.Torres, Heroen Verbruggen, Lifei Wang, Seth Wenger, Mark J.Whittingham, Yuri Zharikov, Damaris Zurell, and Ana M.M.Sequeira. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33:790–802, 2018.

[31] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[32] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, second edition, 2005.

[33] J. Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Comput Stat Data An*, 61:50–66, 2013. ISSN 01679473. doi: 10.1016/j.csda.2012.12.001. URL `http://linkinghub.elsevier.com/retrieve/pii/S0167947312004227`.

[34] R.M. Ward, R. Choudhary, Y. Heo, and J.A.D. Aston. A data-centric bottom up model for generation of stochastic internal load profiles based on space-use type. *Journal of Building Performance Simulation*, 12: 5: 620 – 636, 2019.

[35] John Duchi. Derivations for linear algebra and optimization, 2014. URL `https://web.stanford.edu/ jduchi/projects/general_notes.pdf`.

[36] Laurens van der Maaten and Geoffrey Hinton. Visualising data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[37] Adrian Chong, Weili Xu, Song Chao, and Ngoc-Tri Ngo. Continuous-time Bayesian calibration of energy models using BIM and energy data. *Energy and Buildings*, 194:177–190, 2019.

Appendix A. Interpretation of simulation results

This appendix contains further details of the two models and explores how the models may be interpreted in the light of the simulation results.

Appendix A.1. Autoencoder

The AE model is further examined to understand the underlying reasons for its propensity to correctly predict demand for some test zones better than others. First, the reconstructed profile and the latent dimensions of zones UK:O1 (training data), UK:O5 (test data), UK:IT1 (training data) and UK:IT2 (test data) is illustrated on the left and right of Figure A.13 respectively. It can be observed that similar profiles of zones UK:O5 (red) and UK:O1 (green) (on the left) are also very similar in the latent dimensions. On the other hand, it can also be seen that the dissimilar profile of UK:IT2 (in blue) zone has a much different latent representation as compared to the UK:O5 (red) and UK:O1 (green) zones. Therefore, it can be said that the data that are similar in the original dimensions would tend to be similar in the latent dimensions.

The AE compresses significant characteristics into a small dimension representation. The previous paragraph illustrated that similarities in energy profiles would still be present in the latent dimensions of the autoencoder.

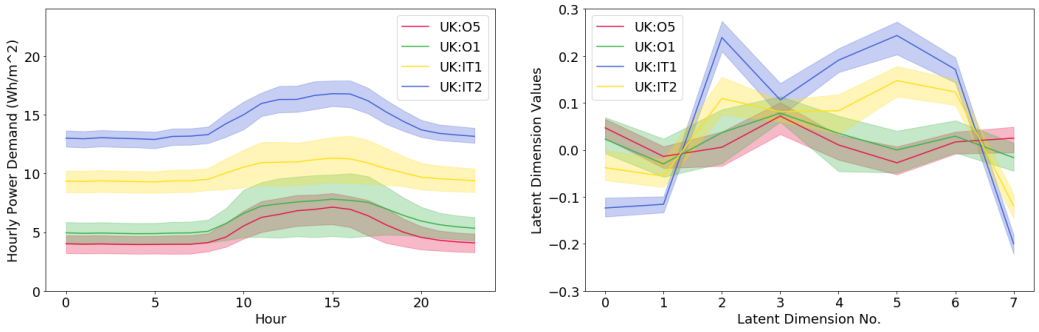


Figure A.13: Reconstructed profile and latent dimensions

Figure 7 shows the similarities and differences between the training and testing datasets. The red points represent the training data from all zones. The data points in various zones are clustered together, indicating the profiles in the given zones possess similar characteristics. For the UK test zones, the UK:O7 and UK:IT2 do not overlap with the training data at all. This in turn suggests that the plug loads in these zones are quite dissimilar to the zones used in the training set. Given that the autoencoder is purely data-driven, this could suggest that the lack of similar data has contributed to the higher K-L divergence values (UK:07 - 1.47, UK:IT2 - 3.04) given in Table 2 and the sampled profiles with greater variation from the actual profiles (shown in Figure 8). For the SG zones, all 4 zones do not really overlap with the training data. This is the reason for the higher K-L divergence values (Table 2) and the greater variation in profiles (shown in Figure 9).

Figure A.14 illustrates the projection of the latent dimensions on a t-SNE plot over all the zones. The latent dimensions of test zone UK:O5 overlaps with the plug loads of the mixed zones, UK:O3 and UK:O4 (training zones). Referring back to the data illustrated in Figure 4 and Figure 5, we can see the similarities in range and shape of the UK:O5 test zone and the UK:O3 and UK:O4 training zones. Furthermore, the plug loads in UK:C2 classroom zone overlaps the plug loads of the UK:C1 Classroom and part of the UK:L zone. Similar plug loads can be seen in these profiles when we refer to Figure 4 and 5. Moreover, it is worth noting that the UK:L zone with large variation of plug load profiles (refer to Figure 4) is also represented with small clusters all over the t-SNE plots.

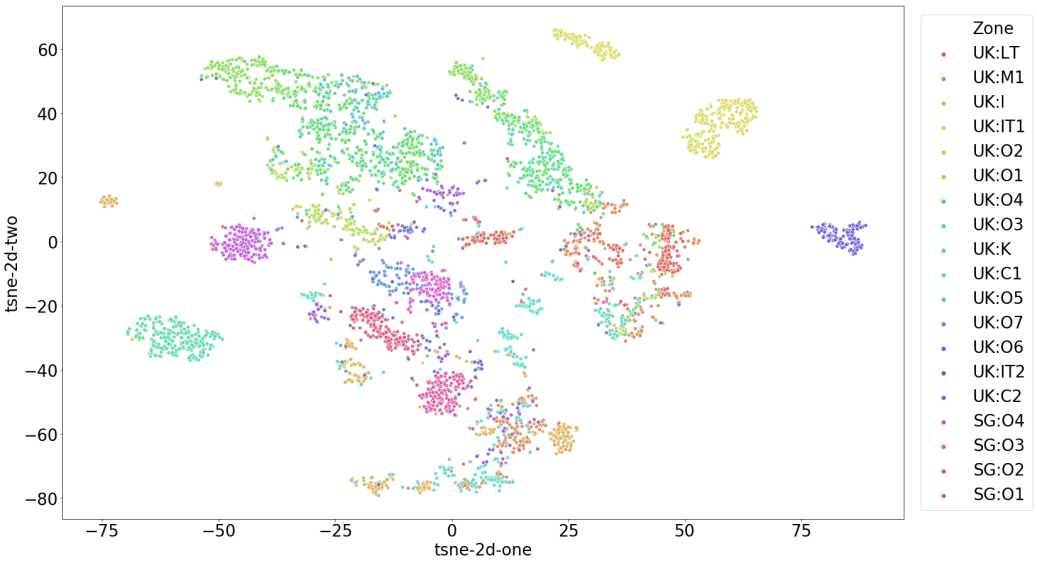


Figure A.14: TSNE: Latent dimensions for all the zones

Last but not least, the projection of the latent dimensions on a t-SNE plot, differentiated by weekday and weekend is illustrated in Figure A.15. It is interesting to see the differences in the weekday and weekend plug loads as weekend plug loads are typically flatter. In the t-SNE plot, with the exception of the UK:L plug loads, the weekend data are clustered together, indicating that that the plug loads over all but one zone are similar over the weekends.

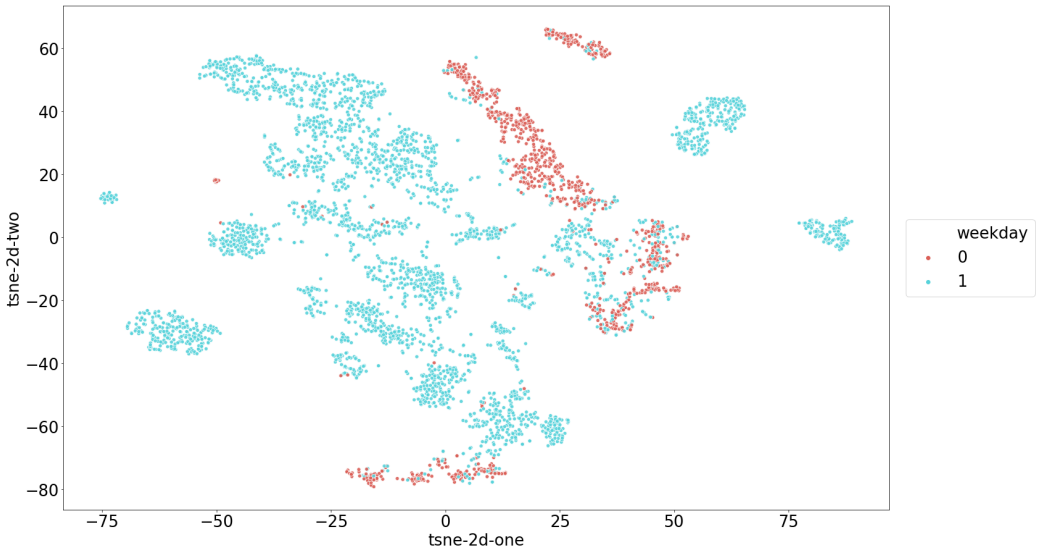


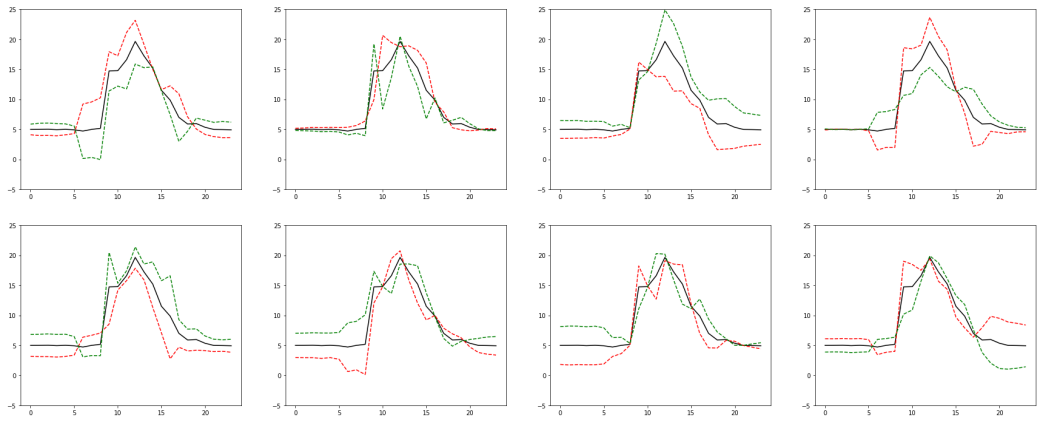
Figure A.15: TSNE: Latent dimensions for all the zones by weekday vs weekends

Next, a sensitivity analysis on the various latent dimensions was performed to explore the possibility of identifying the contribution of each latent dimension i.e. if any particular latent dimension specifically contributes to the base or peak load, or affects the rise and fall of the profile. If this is possible, it would enable the regeneration of profiles by varying the latent dimension values.

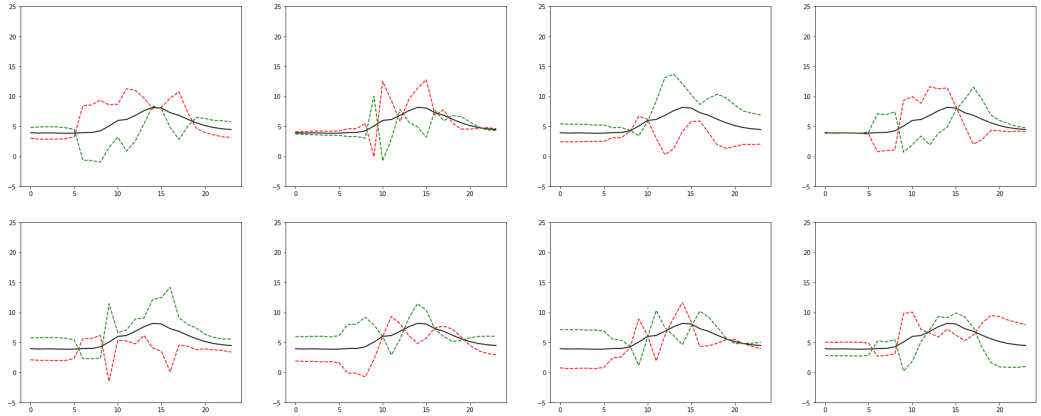
For the sensitivity analysis, two random samples are picked from the UK:K and UK:O1. These zones are chosen because their profiles are dissimilar to each other. In each of the 8 latent dimensions, one latent dimension i is varied by a value of $+/-0.2$ at once, while the other latent dimensions are kept constant. The resulting variations of profiles is illustrated in Figure A.16. In each subplot, the chosen sample energy profile is given in black, profile generated when increasing or decreasing latent dimension i is given in green and red respectively. The top row shows the changes in latent dimensions 1 to 4 from left to right and the second row shows the changes in latent dimensions 5 to 8.

Figure A.16 importantly shows that each latent dimension contributes to the plug load profile over the whole day. For example, it can be observed in both samples that varying the value of latent dimension 3 creates a greater variation in amplitude on the profile towards the end of the day, the changes made would also affect the shape of the overall profile and the amplitude in

the earlier part of the day. Hence, it is not possible to specifically relate a latent dimension to a distinct feature such as a higher peak load or an earlier rise of profile.



UK:K sample



UK:O1 sample

Figure A.16: Sensitivity of latent dimensions - Top row (left to right): Varying latent dimensions 1 to 4, Bottom row (left to right): Varying latent dimensions 5 to 8, + in green and - in red

Appendix A.2. Functional Data Analysis

As outlined in Section 3.2, the FDA model consists of a set of functional principal components for phase and amplitude which are combined with PC scores to generate each data sample. As aforementioned, this study uses normalised UK training data as a basis for extraction of the PCs. Figures

A.17 and A.18 illustrate the first 6 PCs for phase and amplitude respectively. Shown in the figures are the mean function (in black) together with the impact of adding each PC separately with a score of $\alpha = +1$ (dashed red line) or $\alpha = -1$ (dotted blue line). The first phase PC, X1, represents a shift in time of the data, with a positive score (dashed red line) giving an earlier start/finish time and a negative score a later start/finish time. PC X2 has the impact of moving just the start time of the day, so a 'stretch' in time; and PC X3 stretches the time of the peak, with a negative score equating to a much narrower peak. Included in each plot is the proportion of the variance of the data that is attributable to that PC. So the first phase PC accounts for 26% of the variance in the data and the first 6 PCs account for 79%.

The first amplitude PC, Y1 in Figure A.18, relates almost directly to the load range, with a positive score of $\alpha = +1$ giving a high load range (dashed red line) and a negative score giving a low range (dotted blue line). The second amplitude PC governs the base load, but it must be remembered that these are for the normalised data, so the comparison is with the mean of the entire normalised dataset. The higher PCs equate to more localised effects and are responsible for less of the variance in the data. PC Y3 for example has an impact on the timing of the peak demand with a positive score indicating a later peak. Conversely, a positive score for PC Y5 contributes to an earlier peak in the demand. The amplitude PCs account for similar proportions of the variance in the data, with 26% attributable to PC Y1 and 74% to the first 6 Y PCs.

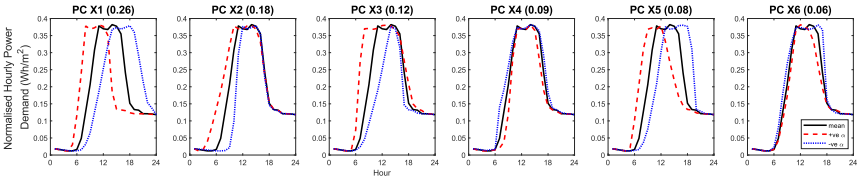


Figure A.17: First 6 Principal Components - Phase

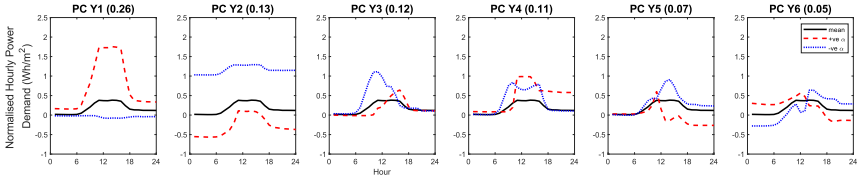


Figure A.18: First 6 Principal Components - Amplitude

Using this approach each data sample reduces to a set of PC scores. Similar data samples typically have similar scores which means it is possible to identify similarities and differences between demand for different zones by comparing the sets of scores.

Figure A.19 shows a t-SNE plot for 4 different zones from the UK training dataset, in which the 48 scores per data sample are compressed to 2D. Clusters of data are observable for the different zones, for example the kitchen zone UK: K is shown in blue and the Lecture Theatre, zone UK: LT in green. These are reasonably distinct - there are some scattered points that don't fall into the clusters, but the majority of points are closer to others from the same zone than to points from the different zone. However, zones UK: O1 and UK: O3 overlap to a much greater extent. These are both office zones, and this plot illustrates the similarity in normalised demand for the zones with similar space use-type.

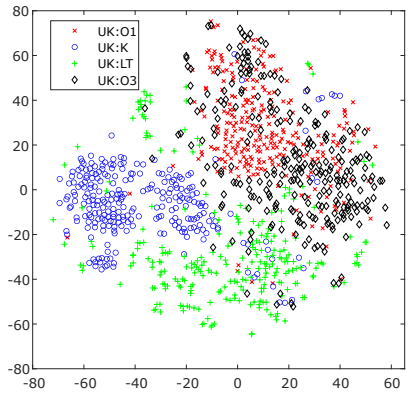


Figure A.19: t-SNE plot for zones UK:O1, UK:K, UK:LT and UK:O3

The scores can used to explore similarities between types of use, be that days of the week, or types of space-use. Consider Figure A.20, in which the scores for the first 10 PCs are plotted for the four office zones UK: O1 - O4. These are quite similar to each other; for the phase PCs, shown in the top plot, the median score for the first PC is negative, around -0.4 - -0.9. Using this information in conjunction with the PC plots in Figure A.17 suggests that the office zones typically have a working day which starts and ends later than the mean of the training data. The scores for the amplitude PCs indicated in the lower figure exhibit a similar level of consistency, with the

first PC again being dominant with a positive score indicating a high load range. Compare these against two different zones in Figure A.21. Zone UK: K is the canteen kitchen, while zone UK: LT is a lecture theatre. These exhibit very different patterns of scores, different from each other and different from the office zones. Indeed the scores for UK: LT, are almost opposite those of the offices for the first 2 phase PCs, with a positive value for PC X1 and a negative value for PC X2, indicating a day shifted earlier than the mean, but with a later start. The negative score for the amplitude PC Y3 reinforces this trend, with an early peak.

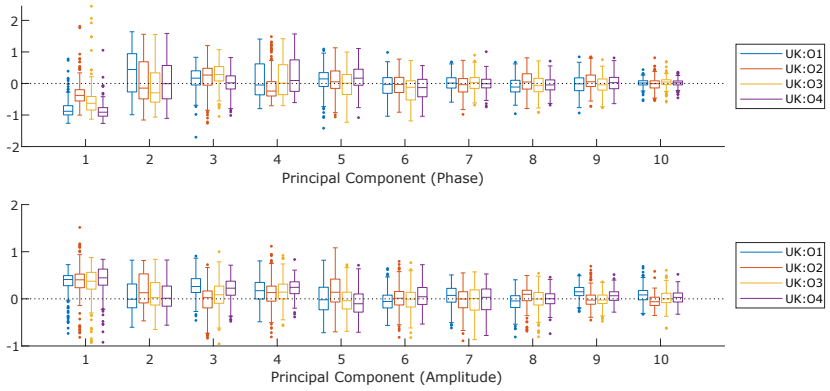


Figure A.20: Comparison of PC scores for similar space use types, UK: O1 - UK: O4

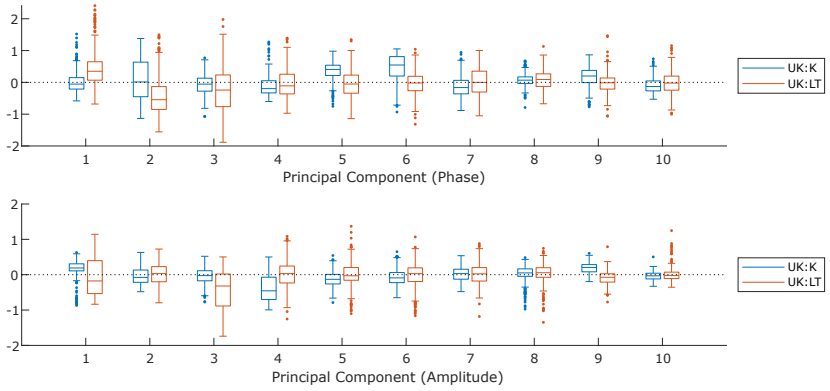


Figure A.21: Comparison of PC scores for different space use-types, UK: K and UK: LT

The KPIs and the K-L divergence discussed in section 5.2 illustrated the level of agreement between the samples derived from the model and the test data. Comparing the PC scores derived for the test data against the training data and also against each other can help understand if similarities in the scores can be related to similarities in use-type.

Figure A.22 illustrates the scores for the first 10 PCs for the UK validation dataset. The top row shows boxplots of the scores for the Phase PCs and the bottom row gives the scores for the amplitude PCs. The first three zones are office zones, and the remaining two are the IT lab, UK:IT2 and a classroom, UK:C2. The median scores for each zone, taken in consideration with the PC plots of Figures A.17 and A.18 give an indication of the impact of each PC on the shape of the demand, while the interquartile range gives an idea of the degree of variability in the data. Considering the phase PCs first, the median scores for PC X1 are very similar for zones UK:O6 and UK:O7. By comparison, zones UK:O5 and UK:IT2 have much more negative scores for this PC, implying a shift in working day to a later start/finish time for these zones. Figure 5 confirms this interpretation; the median start time for zones UK:05 and UK:IT2 is around 10am, whereas it is earlier, around 8am for zone UK: O6 and UK: O7. The range for PC X2 is similar for all zones, suggesting all have a similar degree of overall variability around the timing of the start of day, with zone UK:O6 having the most negative median score and smallest interquartile range, implying a tendency to have a shorter day. Zone UK:O6 also has the most positive median score for PC X3, stretching the extent of the peak demand; again the broad peak is clearly visible in Figure 5. The breadth of the peak is also indicated by the scores for the amplitude PCs, shown in the lower plot. In this case it is a negative score for PC Y4 which has this effect, as seen for zone UK:O6. The first amplitude PC relates to the load range of the normalised data. Zones UK:O5, UK:IT2 and UK:C2 have the most positive median scores for this PC, suggesting a higher load range than the mean across the normalised training dataset. The second amplitude PC relates to the base load, with a negative score implying a higher base load than the mean of the training dataset. A negative median score is seen for zone UK:C2. Given that the data are normalised by subtracting the median demand and dividing by the load range, a score different from zero suggests a degree of skew in the data i.e where the median base load is different from the mean. This is indeed the case for zone UK:C2 for which the median base load is 1.7 and the mean is $1.3 Wh/m^2$.

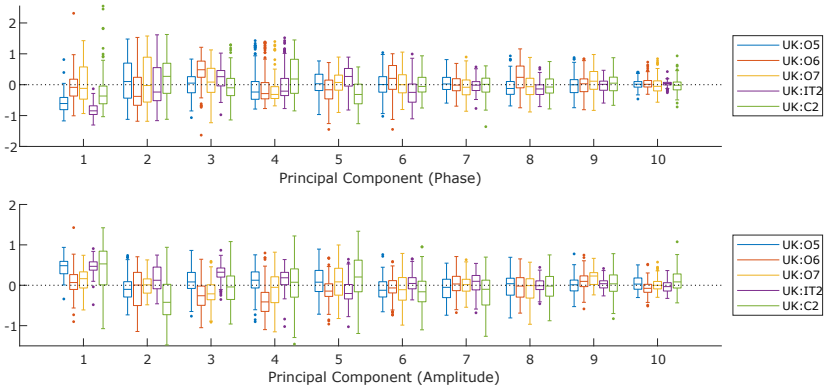


Figure A.22: PC scores for UK validation data

Figure A.23 shows corresponding results for the SG test data. Again, the plots show the scores for the first 10 PCs for phase (top) and amplitude (bottom). Here there are some very clear differences between zones. Typically, zones SG:O1 and SG:O2 are quite similar, but different from zones SG:O3 and SG:O4. For example in the phase PCs, zones SG:O1 and SG:O2 have negative median scores for PC X1 and positive median scores for PC X3 with small interquartile ranges. As we have seen, the negative score for PC X1 reults in a shift in time of the day to a later start and end, but here the positive score for PC X3 counterbalances this with a stretch at the start of the day; the result is a longer working day. By comparison zone SG:O3 has a much wider interquartile range for PCs X1 and X3, giving a greater degree of variability in start and end time, as observed in the data (Figure 6). Zone SG:O4 also shows a wider interquartile range for the phase PCs than the first two zones, reflecting the greater variability in demand.

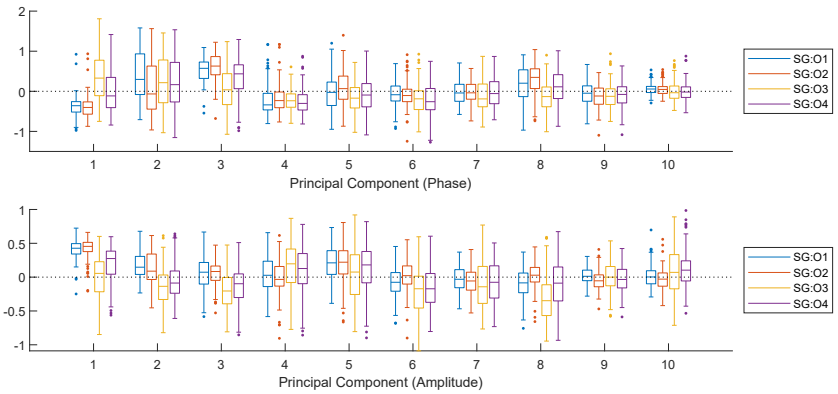


Figure A.23: PC scores for SG test data

For a more detailed example, consider Figure A.24. This shows the normalised data for the test zone UK:C2 (blue) compared against training data zone UK:C1 (red), with the mean demand shown as a line and the 90% confidence limit shown as the shaded area. These zones are both classrooms, located next to each other in the building, and it might be expected that the demand profiles could be similar; certainly in a building energy simulation they would typically be assigned the same demand schedule. The figure shows a comparison of the normalised data on the left, and on the right there are plots illustrating the comparison between the scores for the first 10 phase and amplitude PCs. There are clear differences in the data - zone UK:C2 shows a much greater variability and the median peak is late in the working day, around 6pm. By comparison zone UK:C1, included in the training dataset, has much less variability in the data, a lower base load and a noticeable drop in demand in the early afternoon. Considering the scores on the right of the figure, the top plot shows the scores for the phase PCs and immediately differences between the two zones are obvious. Whereas the training zone UK:C1 has a large negative median score for PC X3, the validation zone UK:C2 has a median score of around zero; conversely zone UK:C1 has a positive median score for PC X6 whereas the value for UK:C2 is again close to zero. These scores for UK:C1 contribute significantly to the dual peak observed for this zone: a negative score for PC X3 gives a late peak whereas a positive score for PC 6 gives an early peak. In contrast, the maximum PC contributions to the zone UK:C2 data are for PCs X1, X2 and X5; a positive score for PC X2 stretches the day, and on its own would

bring the start of day earlier. However, the negative scores for PC X1 and PC X5 shift the whole day later and give rise to the later day start and late peak. There are also clear differences in the scores for the amplitude PCs. The test zone UK:C2 has a high positive median score for PC Y1 - whereas zone UK:C1 has a negative median score. This gives rise to the higher load range for zone UK:C2 compared against zone UK:C1 observed in the data. Similarly, the negative score for PC Y2 for the test data gives rise to a higher median base load as observed. It is interesting to consider how these zones are used. Zone UK:C1 is a larger space used for teaching classes, whereas zone UK:C2 is used more as a study space; students typically attend classes in UK:C1 and move into UK:C2 for individual study once the classes are finished. This is also reflected in the demand (Figures 4 and 5).

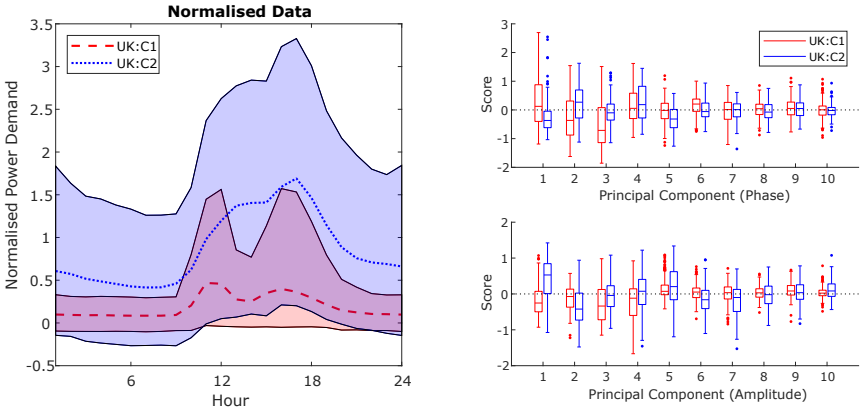


Figure A.24: Relation of PC scores to normalised data - comparison of zone UK:C1 against zone UK:C2

Figure A.25 shows a comparison of the data and PC scores for test zones SG:O1 and SG:O3. As mentioned above, zones SG:O1, SG:O2 and SG:O4 show quite similar scores, suggesting similar demand profiles, but the scores for SG:O3 are a bit different. Looking first at the amplitude scores in the bottom right plot of Figure A.25, SG:O3 has a near zero score for PC Y1 and a negative score for PC Y2, giving rise to a lower load range and a higher base load than zone SG:O1; clearly observable in the data shown on the right side of the figure. There is also a strongly negative score for PC Y3, giving an early peak, and for PC Y8, which tends to give rise to higher demand later in the day and is reinforced by the negative scores for PCs Y6 and Y7. By comparison, the median scores for SG:O1 are positive for the

first 3 Y PCs, giving rise to a higher load range, lower base load and later peaks. Considering the scores for the phase PCs in the top right plot, again clear differences between the two zones are visible. The negative median score for PC X1 for SG:O1 and positive for SG:O3 should give rise to a later start time for SG:O1 and earlier for SG:O3. However, the positive median score for SG:O1 for PC X3 brings the start time earlier again, resulting in an overall stretching of the day relative to the mean. This is also achieved by zone SG:O3 but in a different way: the negative median score for PC X5 has the opposite effect to a positive score for PC X1, again resulting in an overall stretching of the day. The net results are similar for both zones, albeit achieved with a different combination of PC scores. This supports the assertion that it is beneficial to include all PCs in the analysis rather than using the fPCA to reduce the dimensionality.

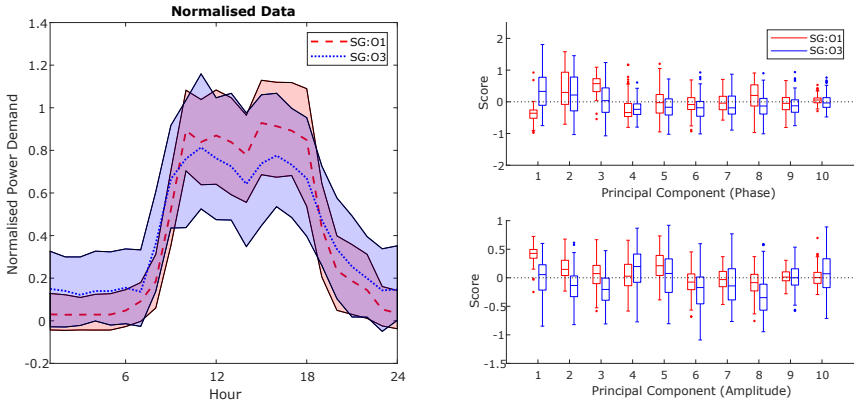


Figure A.25: Relation of PC scores to normalised data - comparison of SG zones SG:O1 and SG:O3