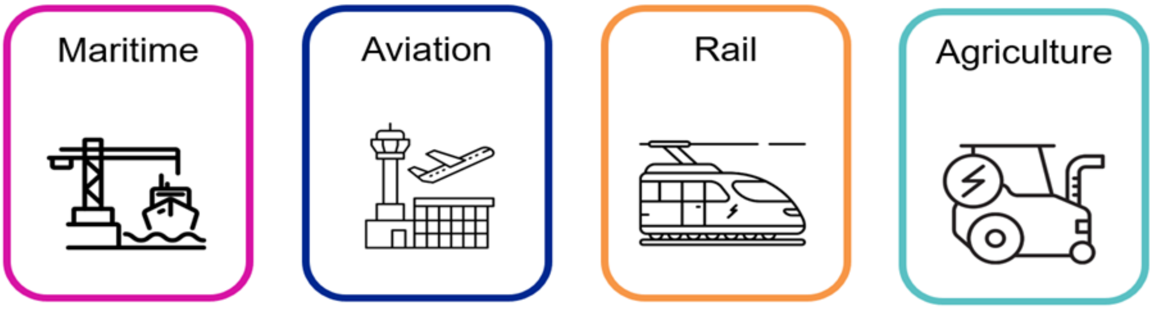
Electricity
Distribution

DFES 2024 Enhancements

New Technologies

With each annual DFES cycle we incorporate and project new technologies in our analysis. In DFES 2024 we have explored how the electrification of aviation, maritime, rail and agricultural machinery will impact operation of our distribution system. These are sectors with significant uncertainty on the pathway to net zero, so early insight is key to ensure that we can support these customers on their decarbonisation journey.



Further detail on these new technologies (and all technologies) can be found in the Technology Summary Reports for each NGED licence area on our DSO website:

https://dso.nationalgrid.co.uk/planning-our-future-network/forecasting-for-future-need/dfes-volumes

Whilst these technologies will not be added by local authorities to LAEP+, their incorporation in DFES will increase the accuracy of load forecasts. Accurate forecasts enable users to understand future load of the network and better plan where to add new LCT or domestic development connections, supporting the PRIDE project.

LAEP Integration

Local Authority Energy Plan ambitions and targets were included for the first time in DFES 2024. Where available, LAEPs superseded data obtained by the Energy Strategy Survey. The Energy Strategy Survey asks questions such as ‘How are you working with local authorities in your area on low-carbon transport ambitions?’. With the integration of LAEPs, specific targets such as 100 EV Charge Points by 2030, enabled simple comparison to volume projections in the DFES.

Incorporation of LAEP data in DFES 2024 lays down the groundwork for how ‘LAEP-stye Plans’ will inform changes to the DFES as part of the PRIDE project. The format we receive LAEP data helps us understand the format NG DSO would need to extract data from LAEP+.

Approved low carbon technology and new development projects in LAEPs can be incorporated directly into volume forecasts in a similar fashion to our connections pipeline, assuming they are not already accounted for in our connections pipeline. Broader decarbonisation ambitions in LAEPs can influence volume projections by warranting a percentage ‘uplift’ if they are credibly above the DFES scenarios envelope.

Enhanced Domestic Energy Profiles

Cluster analysis using machine learning was newly implemented to assign 57 highly specific energy profiles to new domestic developments. Similar energy profiles were grouped together to create clusters and their proportion of technologies were analysed. Using planning knowledge on which technologies will be connected to new substations, the profile of these future substations can be predicted using the profile of the clusters. Detailed methodology found in appendix A.

Local Authority Workbook

A Local Authority Workbook has been created to support local authorities accessing and querying their DFES data. It contains energy and volume projection graphs across all technologies and subtechnologies in addition to underlying raw data tables. It also provides information on the DFES process and how Local Authority data is incorporated to inform projections. The workbook was produced as an .xlsx file in direct response to feedback in last year's stakeholder engagement webinars. This will support PRIDE by allowing LAEP+ users to query the data underlying the DFES layer in addition to visualising it spatially.

Local Authority Workbook: https://connecteddata.nationalgrid.co.uk/dataset/dfes/resource/43ed83f6-234a-47aa-951c-efc16d40ae03

Appendix A - Primary Substation Clustering

Methodology

In order for a computer to imitate the behaviour of a human manually grouping profiles, Machine learning (ML) techniques were used. ML can be defined as a field of study which allows computers to learn without being clearly programmed to do so. Specifically, *unsupervised* ML techniques can be used to produce profile groups for further human analysis. The term “unsupervised” refers to the ML algorithm working with unlabelled data. In this context of this project, a “label” would be a descriptor for a group of profiles with similar behaviours.

This section aims to give a high-level overview of the ML techniques used to produce the profile clusters; the reader should consult other reference material for an in-depth explanation of how the ML techniques work.

The starting point for devising the ML clustering methodology was to consider how a human would group similar demand profiles. An intuitive way of doing this would be to plot the demand profiles, and match those with similar profile shapes. For example, one group may consist of profiles with morning and evening peaks, and another group may consist of profiles with a midday peak. When creating these groups, a human would also consider the “context” of a profile peak. Peaks do not occur at the same time, but fall within a range. For example, the evening peak could fall between 5 pm to 7 pm.

Previous work within the company focused on an unsupervised technique known as k-means clustering and that used Dynamic Time Warping (DTW) as a metric to group the half-hourly data of Electricity Supply Areas (ESAs). Half-hourly data for “representative days” of the year was gathered and clustering performed on these data with the intention of inferring information about the ESAs from the similarity of clustering’s during these representative days.

The k-means algorithm is a hard-clustering process which produces non-overlapping clusters; soft-clustering is slightly different and produces clusters which may overlap. Points are assigned to the most probable cluster group. Soft-clustering can allow better grouping for “non-spherical” data, which is the case for clusters having different variances and co-variances.

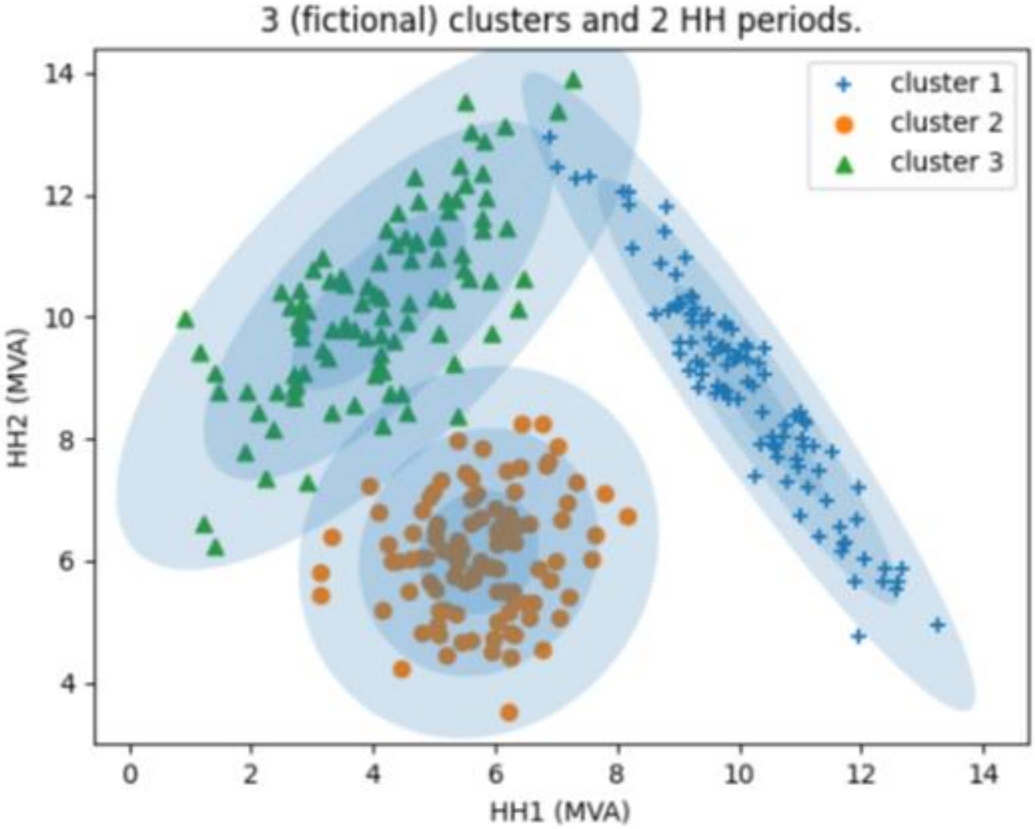


Figure 1: GMM applied to two (fictional) half-hour data.

In this approach a Gaussian Mixture Model (GMM) used the available HH-data for each ESA to produce clustering's. To avoid overfitting the number of clusters, k, a measure of cost/error is used: Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) are typical (DTW method used the within-cluster-sum-of-squares measure for this purpose). The AIC or BIC is calculated for each fitted GMM model; because the initial step of GMM randomly assigns cluster centres there are small difference in final cluster centres, to account for this randomness the model is recalculated 10,000 times for each k and the minimum cost model noted. When cluster numbers are increased the AIC/BIC scores reduce and then increase, the optimal number of clusters (and model) will correspond to the minimum value of AIC/BIC.

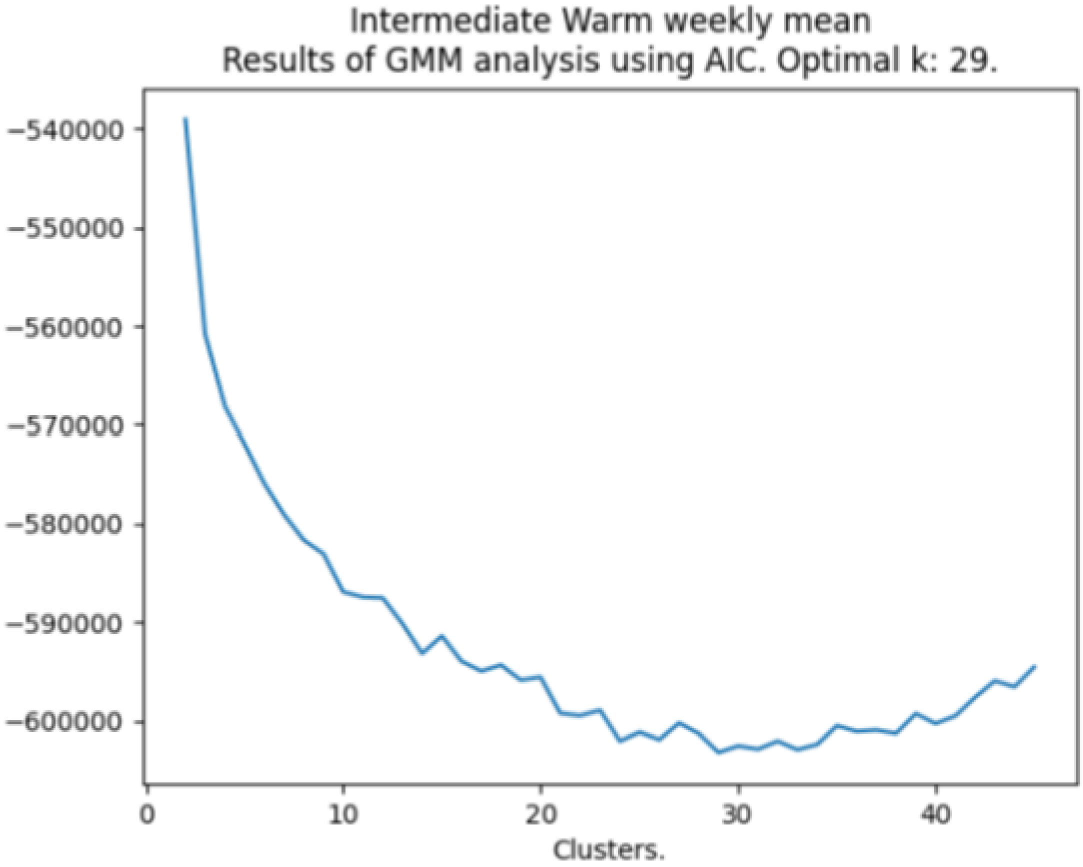


Figure 2: AIC values used to estimate k with GMM.

Prior to clustering the HH real power data for each site was converted to per-unit values based upon the maximum annual value for each site and GMM applied to the data with randomised initial cluster centres and the best models selected. Using AIC/BIC to calculate the optimal number of clusters resulted in a large number of seasonal groupings and the sparsity of ESAs withing similar groupings made regression using MPAN types, LCT types or generation types meaningless. A fixed value of 5 for each seasonal clustering was used.

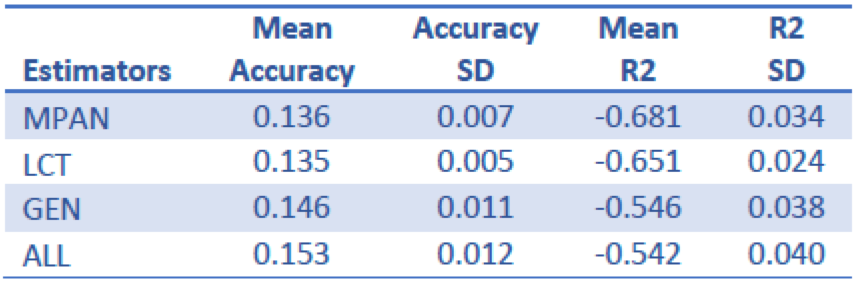
The Adjusted Rand Index (ARI) can give some insight into how ESA behaviour is consistent across seasons, the ARI is calculated between pairs of seasons. If ESAs are mapped to the same clusters (in different seasons) there is good reason to believe that similar factors drive that behaviour. An ARI score of +1 indicates that the groupings are identical, whereas a score of 0 indicates a random assignment to clusters. The results ranged from 0.197 between Winter and Intermediate Cool through to 0.067 between Intermediate Warm and Summer minimum.

Once clustering was completed ESAs were grouped into similar five-seasonal groups (or classes) and regression was carried out for each using the proportion of MPAN types, LCT and generation types as estimators. Classes may be labelled, for example, “W2,S1,Sgen4,IC4,IW3” indicating an ESA is in the following groups: Winter 2, Summer 1, Summer generation 4, Intermediate Cool 4 and Intermediate Warm 3. 215 unique classes were discovered with the data set used; this number will vary should other HH data be used.

Stratified k-fold cross-validation was used to explore any underlying factors, and whether regression models might have some predictive use; the limited-memory Broyden-Fletcher-Goldfarb-Shannon (LBFGS) solver, using 5 folds and 1000 iterations was used for cross-validation. The proportions of MPANS, LCTs, Generators connected to each ESA were used as variables in the multinomial regression. Table 1 shows the results of the cross-validation of the models to predict the seasonal class of each ESA.

Both of the cross-validation measures have low standard deviation indicating that it is possible to construct linear regression models to predict the seasonal classes. The mean accuracy results range from 0.14 to 0.15 which is about 30 times better than chance. The Coefficient of determination scores (R2) are negative, indicating that the estimators are not too good at explaining the variance of the five-season-classes; but it should be noted that during the cross-validation many of the test sets were very small and this could result in differences between the means of the test and training data producing poor R2 scores. It is unclear whether a different regression model, or more pre-processing of the estimators and HH data, or another set of estimators, would explain the clustering classes better. The simplest linear regression model would use MPAN data as a predictor for seasonal classes.

Table 1: Results of cross-validation on regression models.



Estimators	Mean Accuracy	Accuracy SD	Mean R2	R2 SD
MPAN	0.136	0.007	-0.681	0.034
LCT	0.135	0.005	-0.651	0.024
GEN	0.146	0.011	-0.546	0.038
ALL	0.153	0.012	-0.542	0.040