



2012-6

# Analysing Domestic Electricity Smart Metering Data Using Self Organising Maps

Fintan McLoughlin

*Dublin Institute of Technology*, [fintan.mcloughlin@dit.ie](mailto:fintan.mcloughlin@dit.ie)

Aidan Duffy

*Dublin Institute of Technology*, [aidan.duffy@dit.ie](mailto:aidan.duffy@dit.ie)

Michael Conlon

*Dublin Institute of Technology*, [Michael.Conlon@dit.ie](mailto:Michael.Conlon@dit.ie)

Follow this and additional works at: <http://arrow.dit.ie/dubencon2>

 Part of the [Applied Mathematics Commons](#), and the [Applied Statistics Commons](#)

## Recommended Citation

McLoughlin, F., Duffy, A. & Conlon, M. (2012). Analysing domestic electricity smart metering data using self-organising maps. *CIREC Workshop*, Lisbon, Portugal, 29-30 May.

This Conference Paper is brought to you for free and open access by the Dublin Energy Lab at ARROW@DIT. It has been accepted for inclusion in Conference Papers by an authorized administrator of ARROW@DIT. For more information, please contact [yvonne.desmond@dit.ie](mailto:yvonne.desmond@dit.ie), [arrow.admin@dit.ie](mailto:arrow.admin@dit.ie), [brian.widdis@dit.ie](mailto:brian.widdis@dit.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)



## ANALYSING DOMESTIC ELECTRICITY SMART METERING DATA USING SELF ORGANISING MAPS

Fintan MC LOUGHLIN  
 Dublin Institute of Technology (DIT) – Ireland  
 fintan.mcloughlin@dit.ie

Aidan DUFFY  
 DIT – Ireland  
 aidan.duffy@dit.ie

Michael CONLON  
 DIT - Ireland  
 michael.conlon@dit.ie

### ABSTRACT

*This paper investigates a method of classifying domestic electricity load profiles through Self Organising Maps (SOMs). Approximately four thousand customers are divided into groups based on their electricity demand patterns. Dwelling and occupant characteristics are then investigated for each group. The results show that SOMs are an effective way of classifying customers into groups in terms of their electrical load profile and that certain dwelling and occupant characteristics are significant factors in determining which group they end up in.*

### INTRODUCTION

Electricity use in EU - 27 member states accounted for 28.6% of final electricity consumption in 2008. However, despite improvements in appliance efficiencies, average electricity consumption for the sector still increased 1.8% compared to 2007, mainly due to higher incomes, smaller households, larger dwellings and increased ownership of electrical appliances [1]. But in order to have a true understanding as to the drivers of electricity consumption and the measures to reduce its use, it is important to have a detailed grasp as to how electricity is consumed in the home.

Self Organising Maps (SOMs) is a technique used to segregate data that show similarities and order them into groups. They were originally conceived by Kohonen [2], a Finnish mathematician who realised their potential for various applications such as speech recognition, image processing and robotics. In the past, SOMs have also been applied to the electricity industry. Rasanen et al. [3] applied SOMs to create comparison groups so that customers who exhibited similar building characteristics could compare electricity use against each other. Dominguez et al. [4] used SOMs to analyse electrical load data from a group of buildings as well as environmental and electricity tariff information in order to achieve economic and energy savings. Sanchez et al. [5] classified domestic customers in terms of their electricity load profile. Dent et al. [6] applied SOMs and C-means clustering in order to classify customers into nine different profile groups based on their electricity consumption patterns.

### METHODOLOGY

The data set used in the analysis below was taken from a population of 345,645 households in Ireland. The population was divided into six groups based on total

annual household electricity consumption to ensure an even spread of electricity consuming customers. An initial sample of 5,574 was drawn on a randomised basis across all profiles. This was subsequently reduced to 5,375 households by targeting certain groups to improve representivity of dwelling and socio-economic variables within the sample size. A final sample size of 3,941 households, for a single day (Wednesday 1<sup>st</sup> July 2009) was used in the analysis below once large outliers and non-continuous data were removed.

SOMs apply a neural network process that uses unsupervised learning to divide a data set into different groups. A rectangular or hexagonal lattice structure of nodes is usually used to segregate the data. Figure 1 presents a hexagonal structure with a 3x3 matrix from Matlab SOMs toolbox which was used in the analysis. The sample number of hits for each group is also shown in Figure 1 at the centre of each hexagonal.

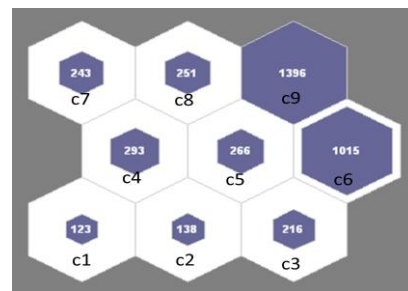


Figure 1: SOM groups and sample number of hits

Each hexagonal node is defined by a weight vector which consists of 48 different dimensions, representing half hourly time intervals across a day. The mapping process is started by initialising weight vectors with random values at each node. As the network progresses each input vector is compared with the weights of each node and the node with the greatest similarity (called the Best Matching Unit) is assigned that particular vector. The weights are then adjusted at the node based on the input vector. The process is repeated until all input vectors have been categorised into groups.

SOM's use a number of different metrics to determine whether an input vector belongs to a particular node or not, however, Euclidean distance is the most commonly used and is applied in the following analysis. The distance is calculated at each node by Equation 1 below, where  $d_{st}$  is the Euclidean distance between the input vector  $x_s$  and the node centre vector  $y_t$ .

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \quad (1)$$

At every iteration,  $y_i$  will change as weights are altered by each new input vector. Figure 2 shows the weight distances calculated between nodes. Brighter colours (yellow) represent nodes closer together whereas darker (black) colours represent nodes further apart. For example nodes 6 and 9 are very close together and are represented by bright yellow whereas nodes 1 and 4 are not very close and are shown in black representing distinctly different profiles.

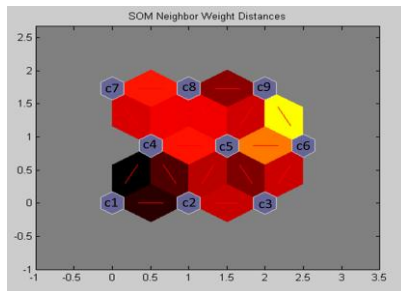


Figure 2: SOM neighbour distances (Euclidean)

Figure 3 shows mean electricity demand at half hourly intervals for each of the nine different nodes. The timing of peak time electricity demand can be seen to be quite different for each group. In particular, groups c6 and c9 appear to have distinctly different patterns of electricity use when compared to the rest of the groups. As shown in Table 1, both groups on average use much smaller amounts of electricity compared to the rest of the groups. However, group c9 appears to have a almost flat average electricity demand profile across the entire day. This is not an accurate reflection of customers within this group as will be explained in the next figure.

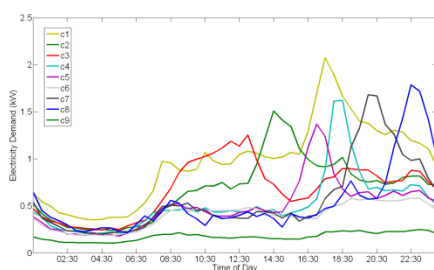


Figure 3: Mean load profile shape by SOM group

Figure 4 shows individual customer profiles plotted by group numbers c1-c9. As well as groups c6 and c9 having very different profile shapes to the rest, these two groups also represent around 60% of the entire sample. Due to the large sample size within these two groups and the highly stochastic nature of their electricity demand profiles (i.e. small intervals of electricity use across different times of the day) the average profile shape appears almost flat.

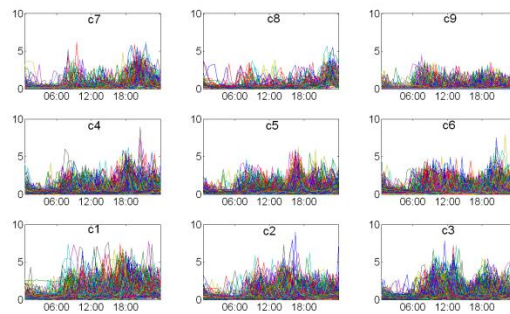


Figure 4: Electricity load profiles by SOM group with time of day on the x-axis and electricity demand in kW on the y-axis

## RESULTS AND DISCUSSION

The data set was organised into nine different groups as discussed above. For each group Table 1 presents four parameters: total electricity consumption, maximum demand, load factor and time of use of maximum electricity demand. Each parameter is determined by calculating the mean for all customers within each group except for time of use where the mode was used. The sample size for each group is shown with overall percentages in brackets.

Groups c2 and c3 have the earliest electricity demand peaks at 14:30 and 13:00 respectively, showing high electricity use patterns around lunch time. Groups c6 and c9 use the least amount of total electrical energy, with the latter consuming less than half that of the former. The unusually high load factor for group c9 is caused by a high percentage of customers who consumed very little electricity across the day.

| SOM No. | Sample Size  | Total Electricity Consumption (kWh) | Maximum Demand (kW) | Load Factor (%) | Time of Use |
|---------|--------------|-------------------------------------|---------------------|-----------------|-------------|
| c1      | 123 (3.1%)   | 23.04                               | 4.14                | 24.34           | 17:30       |
| c2      | 138 (3.5%)   | 16.27                               | 3.30                | 22.03           | 14:30       |
| c3      | 216 (5.5%)   | 16.25                               | 3.26                | 22.42           | 13:00       |
| c4      | 293 (7.4%)   | 12.45                               | 2.97                | 19.01           | 18:00       |
| c5      | 266 (6.7%)   | 11.92                               | 2.79                | 19.17           | 17:00       |
| c6      | 1015 (25.8%) | 9.59                                | 2.11                | 21.75           | 20:00       |
| c7      | 243 (6.2%)   | 13.55                               | 3.19                | 18.75           | 20:30       |
| c8      | 251 (6.4%)   | 12.89                               | 2.88                | 19.56           | 22:30       |
| c9      | 1396 (35.4%) | 4.18                                | 1.02                | 27.57           | 18:00       |

Table 1: Total electricity consumption, maximum demand, load factor and time of use for each SOM group

A number of dwelling and occupant characteristics were investigated for each group, with the results presented in terms of percentages in Table 2. Sample size for each characteristic group is also shown with overall percentages in brackets.

| Dwelling/Occupant Variable | Sample Size (penetration) | c1 (%) | c2 (%) | c3 (%) | c4 (%) | c5 (%) | c6 (%) | c7 (%) | c8 (%) | c9 (%) |
|----------------------------|---------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Detached                   | 1043 (26.5%)              | 34.29  | 28.57  | 36.42  | 26.54  | 21.60  | 27.12  | 32.37  | 36.25  | 20.95  |
| Semi-detached              | 1230 (31.2%)              | 21.90  | 29.44  | 23.15  | 30.21  | 38.76  | 31.37  | 31.65  | 26.25  | 33.97  |
| Terraced                   | 569 (14.4%)               | 8.10   | 10.39  | 10.49  | 13.04  | 14.79  | 14.94  | 7.19   | 15.00  | 18.49  |
| Bungalow                   | 1025 (26.0%)              | 35.71  | 31.60  | 29.63  | 29.52  | 23.67  | 24.72  | 28.06  | 21.25  | 22.59  |
| Apartment                  | 67 (1.7%)                 | 0.00   | 0.00   | 0.00   | 0.46   | 0.89   | 1.66   | 0.72   | 1.25   | 3.83   |
| No. bedrooms - 1           | 42 (1.1%)                 | 0.00   | 0.43   | 0.00   | 0.23   | 0.59   | 0.74   | 1.44   | 1.25   | 2.46   |
| No. bedrooms - 2           | 333 (8.5%)                | 0.95   | 2.16   | 2.47   | 5.03   | 5.33   | 7.66   | 5.04   | 5.00   | 16.76  |
| No. bedrooms - 3           | 1748 (44.4%)              | 26.67  | 37.23  | 36.11  | 40.05  | 46.45  | 48.71  | 33.09  | 33.75  | 50.64  |
| No. bedrooms - 4           | 1367 (34.7%)              | 47.62  | 41.99  | 42.59  | 40.73  | 37.28  | 33.39  | 47.48  | 45.00  | 24.04  |
| No. bedrooms - 5           | 442 (11.2%)               | 24.76  | 18.18  | 18.83  | 13.27  | 10.06  | 9.32   | 12.95  | 13.75  | 5.92   |
| No. bedrooms - 6           | 9 (0.2%)                  | 0.00   | 0.00   | 0.00   | 0.69   | 0.30   | 0.18   | 0.00   | 1.25   | 0.18   |
| No. occupants - 1          | 804 (20.4%)               | 1.90   | 6.49   | 6.79   | 8.92   | 9.47   | 14.76  | 10.07  | 10.00  | 46.45  |
| No. occupants - 2          | 1272 (32.3%)              | 14.29  | 19.05  | 24.69  | 28.15  | 36.09  | 41.42  | 28.06  | 30.00  | 32.88  |
| No. occupants - 3          | 687 (17.4%)               | 18.10  | 17.75  | 20.68  | 25.63  | 21.30  | 19.37  | 20.86  | 26.25  | 8.83   |
| No. occupants - 4          | 689 (17.5%)               | 30.48  | 27.27  | 23.15  | 22.43  | 19.82  | 17.25  | 25.18  | 23.75  | 7.38   |
| No. occupants - 5          | 330 (8.4%)                | 20.48  | 18.61  | 17.28  | 11.21  | 9.17   | 4.70   | 11.51  | 7.50   | 3.19   |
| No. occupants - 6          | 123 (3.1%)                | 12.86  | 6.93   | 4.94   | 2.75   | 3.25   | 2.12   | 2.16   | 2.50   | 1.18   |
| HoH age - 18-25            | 16 (0.4%)                 | 0.00   | 2.16   | 0.00   | 0.46   | 0.59   | 0.18   | 0.00   | 0.00   | 0.46   |
| HoH age - 26-35            | 374 (9.5%)                | 6.67   | 9.96   | 8.64   | 10.76  | 6.21   | 9.13   | 20.86  | 21.25  | 8.74   |
| HoH age - 36-45            | 826 (21%)                 | 21.90  | 24.68  | 26.54  | 25.17  | 20.12  | 18.45  | 22.30  | 23.75  | 19.03  |
| HoH age - 46-55            | 950 (24.1%)               | 36.67  | 29.00  | 20.68  | 27.46  | 31.07  | 22.14  | 32.37  | 30.00  | 18.67  |
| HoH age - 56-65            | 841 (21.3%)               | 20.00  | 19.05  | 25.31  | 21.74  | 23.37  | 21.13  | 15.11  | 11.25  | 21.86  |
| HoH age - 65+              | 912 (23.1%)               | 14.29  | 14.72  | 18.21  | 14.42  | 18.34  | 28.14  | 9.35   | 12.50  | 30.60  |
| HoH Social Class - AB      | 593 (15.1%)               | 17.62  | 19.91  | 20.06  | 15.79  | 13.61  | 13.01  | 23.74  | 27.50  | 12.20  |
| HoH Social Class - C       | 1697 (43.1%)              | 49.52  | 41.56  | 46.91  | 50.80  | 45.27  | 40.41  | 55.40  | 46.25  | 38.07  |
| HoH Social Class - DE      | 1505 (38.2%)              | 29.05  | 35.06  | 28.70  | 30.89  | 38.46  | 42.25  | 18.71  | 23.75  | 45.72  |
| HoH Social Class - F       | 107 (2.7%)                | 3.33   | 2.60   | 3.09   | 1.14   | 1.78   | 3.32   | 2.16   | 1.25   | 3.01   |

Table 2: Dwelling and occupant characteristics in terms of percentage penetration by SOM group

Figure 5 shows the influence of dwelling type on each group. Apartments, which make up a very small proportion of the overall sample (less than 3%) tend to be greatest in group c9 which is not surprising as this group consumes the least amount of electricity overall. Conversely groups c1-c3 and c7 consume the most electricity, with detached and bungalow dwellings making up more than 60% in each of these groups.

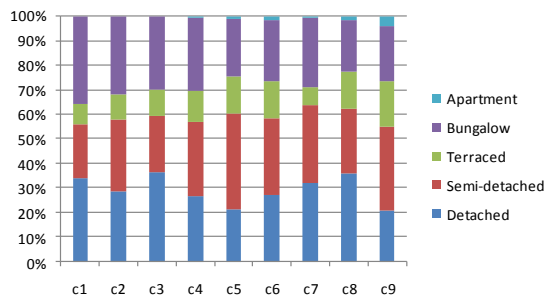


Figure 5: SOM groups by dwelling type

Figure 6 shows the number of bedrooms within a dwelling by each SOM group. Group c9 has the largest percentage of one and two bedrooms representing smaller dwelling types. As already discussed, this group consumes the least electricity overall, with larger four, five and six bedroom dwellings making up over 60% in groups c1-c3 and c7-c8.

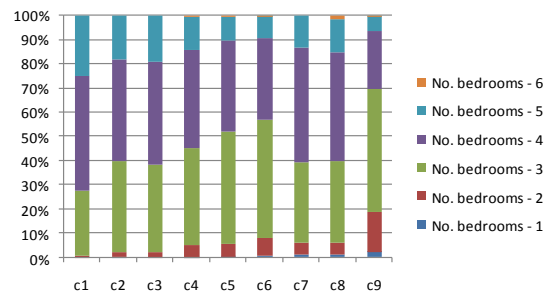


Figure 6: SOM groups by dwelling number of bedrooms

Figure 7 shows number of occupants by each group. Group c9 is largely made up of one and two occupant households, more so than the any of the other groups. As already shown in Figure 3 above, groups c6 and c9 have very similar patterns of electricity demand but differ mainly in terms of the magnitude. Figure 7 possibly explains why this is the case with group c9 having a very different occupant composition compared to c6.

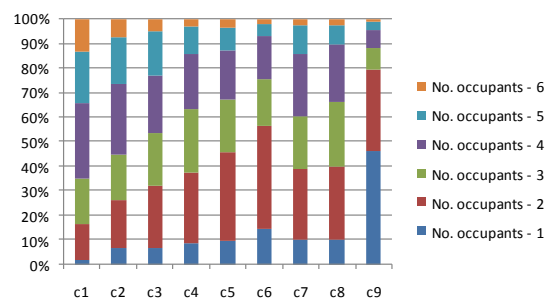


Figure 7: SOM groups by dwelling occupants

Figure 8 shows each group by head of household (HoH) age. Groups c6 and c9 have a higher proportion of older, 65+ HoH's when compared against the rest of the groups. It is also interesting to note that group c7 and c8 have a higher proportion of younger HoH 26-35 which also corresponds with a later use of maximum electricity demand as shown in Figure 3.

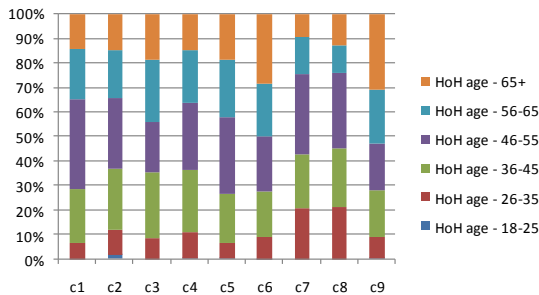


Figure 8: SOM groups by dwelling HoH age

Figure 9 shows groups by HoH social class. Groups c6 and c9 have the lowest percentage of higher classes and the highest percentage of lower classes. The middle HoH's social classes are evenly distributed across all electricity groups.

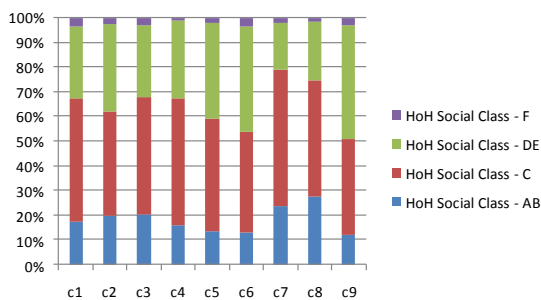


Figure 9: SOM groups by dwelling HoH social class

## CONCLUSION

SOMs were successfully applied to segregate individual domestic electricity load profiles into nine different groups. This resulted in two large sample groups (c6 and c9) with similar stochastic profile shapes but differed substantially in electricity consumption. The remaining seven groups differed in their profile shape mostly in terms of the timing of maximum electricity demand.

Dwelling and occupant characteristics were investigated to determine whether there was any significance to these characteristics within each group. The results indicated that these were significant and that, in particular, the number of bedrooms and occupants influenced which group a dwelling belonged to. These results also show that an approximate load profile shape and electricity consumption profile can be generated for customers based on their dwelling and owner characteristics.

## ACKNOWLEDGEMENT

Dublin Institute of Technology would like to acknowledge the support of Electric Ireland for access to the dataset.

## REFERENCES

- [1] E. E. Agency, "Average annual growth rate in electricity consumption by sector 2007-2008, EU-27." [Online]. Available: <http://www.eea.europa.eu/data-and-maps/figures/average-annual-growth-rate-in-3>. [Accessed: 06-Feb-2012].
- [2] T. Kohonen, "The Self-Organizing Map (Kohonen)," *Proceedings of the IEEE*, Vol. 78, No. 9, September 1990.
- [3] T. Räsänen, J. Ruuskanen, and M. Kolehmainen, "Reducing energy consumption by using self-organizing maps to create more personalized electricity use information," *Applied Energy*, vol. 85, no. 9, pp. 830-840, Sep. 2008.
- [4] M. Dominguez, J. Fuertes, I. Diaz, A. Cuadrado, S. Alonso, and A. Moran, "Analysis of electric power consumption using Self-Organizing Maps .," *Preprints of the 18th IFAC World Congress Milano (Italy)*, pp. 12213-12218.
- [5] I. B. Sanchez, I. D. Espinos, L. M. Sarrion, A. Q. Lopez, and I. N. Burgos, "Clients Segmentation According to their Domestic Energy Consumption by the Use of Self-Organizing Maps," *6th International Conference on the European Energy Market*, pp. 1-6, 2009.
- [6] I. Dent, U. Aickelin, and T. Rodden, "Application of a clustering framework to UK domestic electricity data," *UKCI 2011, the 11th Annual Workshop on Computational Intelligence, Manchester*, pp. 161-166, 2011.