Geodemographic analysis and estimation of early plug-in hybrid electric vehicle adoption

Jukka Saarenpää[a,*], Mikko Kolehmainen[a], Harri Niska[a]

[a]*Department of Environmental Science, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland*

Abstract

Electric vehicles and hybrids are expected to become increasingly common in the coming years. The implications of growing adoption depend on its geographical extent. For instance, vehicles that are chargeable from the electrical grid, such as plug-in hybrids, can introduce problems for the distribution network especially if the vehicle adoption is spatially concentrated. In this paper, the adoption of hybrid electric vehicles is analysed in heterogeneous areas. The main purpose is to study the interrelationships between early hybrid electric vehicle adoption and different demographic and socio-economic characteristics of the areas. It is further discussed how the results can be applied to estimate the upcoming plug-in hybrid adoption. As there is a vast amount of information in the various registers of the society, slowly being opened for free usage but not fully utilised so far, it is also of interest to study and demonstrate the usability of public register data in this context. Our analysis suggests that certain characteristics of the areas strongly correlate with the hybrid electric vehicle adoption. The results of this study could be relevant, e.g., for electric distribution network planning, targeting policies to support cleaner vehicle adoption, marketing hybrid vehicles and locating charging stations.

Keywords: Hybrid Electric Vehicle Adoption, Geodemographic Analysis, Plug-in Hybrid, Electric Distribution Network Planning

1. Introduction

The worries over green house gas emissions and limited oil resources have led to a global attempt to increase the energy efficiency and the share of the renewables. According to World Resources Institute, in 2005 the transportation sector produced 14.3 percent of the global greenhouse gas emissions [1]. Electric vehicles (EV) and hybrid electric vehicles (HEV) are expected to improve the situation by increasing the fuel efficiency and by presenting a serious alternative to the use of fossil fuels.

The batteries used in electric vehicles and plug-in hybrid electric vehicles (PHEV) are charged from the grid, and thus, substantial adoption of such vehicles can cause problems for the power-distribution networks. A situation where the adoption of EVs and PHEVs is regionally concentrated and the charging is unmanaged, has been considered especially risky in several studies (e.g. Lopes et al. [2] and Shao et al. [3]).

In the coming years, power-distribution networks are expected to be capable of handling various distributed energy resources (DER) intelligently. As a result of the increased automation, the so called smart grid will potentially improve the efficiency, reliability and security of the electricity distribution. The grid connected vehicles can also be regarded as energy resources, as they do not only cause new increased loads while being charged, but can potentially be used to feed the battery stored energy back to the grid by providing vehicle-to-grid services (V2G). The possible use cases vary from serving the grid with more power during the peak hours to providing ancillary services, such as frequency regulation.

All in all, there is an increasing need for tools to help the planners and policy-makers to assess the diffusion of environmentally friendly technological innovations, such as the electric vehicles. Apart from that, there is a vast amount of data in the various registers of the society, slowly being opened for free usage, but not fully utilised so far. For privacy and practical reasons such data is often constrained in spatial resolution and might lack in coverage. Thus, methods are needed which can deal with the aforementioned constraints.

The general problem in assessing the diffusion of innovations is that proper data, collected from real market transactions, simply does not exist yet or its quantity is low. To deal with the absence of data, it is in some cases possible to relate the new product to another one, which has been in the market for a longer period of time, by analogy. Analogous products have certain similarities, and yet, such distinct features that they cannot be classified just as different generations of the same product. The method is used especially when attempting to predict the sales performance of new products. For instance, Bass et al. [4] used an analogy between a new satellite television prod-

*Corresponding author. Tel.: +358-40-355-2322.
Email address: jukka.saarenpaa@uef.fi (Jukka Saarenpää)

uct and cable TV.

The market for electric vehicles and hybrids is still in its early days. Several models of HEVs have been sold for years but the adoption has been slow. The additional price premium, the limited model selection and the hybrid technology's unfamiliarity are possible culprits for the slow adoption. The hybrid car market is still marginal compared to the conventional car market. As for the PHEV models, several car manufacturers have planned to start selling them during 2012. Pure EVs are still very rare and the adoption is expected to be slower compared to HEVs and PHEVs. This is thought to be the result of their short driving range before having to recharge, slow charging and the high cost of batteries. Because of the range and recharging issues the use cases for an EV are limited at the moment compared to the conventional internal combustion engine vehicles (ICEV), HEVs and PHEVs. Thus, as concluded by Tamor et al. [5], a PHEV is much more plausible replacement for a conventional vehicle than an EV.

One thing that all the electric vehicle technologies certainly have in common is the capability to partly or entirely substitute the fossil-based fuel by electricity. Therefore the technology's selling points could be considered as being environmental friendliness and fuel (cost) savings. Moreover, some consumers seem to perceive the hybrid and EV technology as more advanced and technologically more interesting than the conventional ICEV technology [6]. Turrentine and Kurani [7] recognized that car buyers in the US market are not tracking their fuel consumption accurately. According to the study, a decision to buy environmentally friendly vehicle is more probably based on symbolic value and the perceived efficiency rather than on well calculated fuel savings.

As the penetration of HEVs is still very low, the current owners constitute of innovators or early adopters. They have demonstrated their willingness to pay premium price for environmentally friendly or fuel saving vehicle technology. The same consumer traits should be also applicable to the owners of PHEVs and possibly to the owners of EVs. Moreover, the possible use cases between the current PHEVs and HEVs are mostly similar. Either one of them can be a direct replacement for a conventional vehicle, in contrast to EVs which impose additional restrictions, such as the limited driving range. In comparison to HEVs, PHEVs introduce the additional issue of requiring a place for charging. Even though PHEVs can be charged from a normal electrical wall socket, for instance in urban setting, it might be sometimes difficult to find one that is available on a parking lot. Moreover, specialised charging stations might still not be common enough to depend on. In Finland however, it is commonly thought that the vehicle charging will happen mostly at home, as that is where the vehicle is parked for longest continuous period of time during the day. Also, many of the Finnish apartments and work places already have the infrastructure necessary for vehicle pre-heating. The same vehicle pre-heating infrastructure can be utilised as it is or with small modifications to charge the vehicles [8]. As such, at least in the Finnish setting, the availability of vehicle charging infrastructure is less of an issue and we do not consider it as a big differentiating factor between the adopters of HEVs and PHEVs. Thus, it could be assumed that HEVs and PHEVs are analogous products and the adoption would mostly happen among the same consumer groups.

The literature on the factors affecting the willingness to adopt EVs and hybrids is extensive. A large portion of it is based on stated preference surveys, while those based on revealed preferences, i.e. actual market transactions, are fewer. A large portion of the studies also consider the adoption on a countrywide scale rather than regionally, thus not analysing the phenomenon spatially. Furthermore, the use of already existing and publicly available registers, such as census data has been largely unexplored in this context.

In this paper, we have analysed the adoption of HEVs in distinct area types, characterized by different demographic and socio-economic attributes, by making use of public census data. A big motive for the analysis is in the potential to utilise it for estimating the upcoming PHEV adoption. As predicting the diffusion of new (or still non-existent) products is very difficult, the hypothetical analogy between HEVs and PHEVs could be used to predict the upcoming PHEV adoption in different geographical areas. The emphasis here is on the early adoption, as that is what the HEV adoption data available for this study really represents. This could have useful applications e.g. in policy-making, electric distribution network planning and locating charging stations.

The rest of the paper is structured as follows. Section 2 introduces the used data and methods. Section 3 contains analysis of the data both from temporal and regional perspectives. The results are presented and further discussed in section 4. Finally, section 5 concludes the work.

2. Materials and methods

In this section we present the used materials and methods. The main parts of the analysis constitute of data integration and preprocessing, clustering of the areas according to demographic characteristics and calculation of the HEV adoption distribution between the demographic clusters or segments. Moreover, assuming that HEV and PHEV are analogous products, the adoption behaviour of the clusters should be similar for both vehicle types. The diagram of Figure 1 presents the whole data processing chain. More detailed description is provided in the following sections. The computations were carried out by using Matlab® software and the SOM toolbox [9].

2.1. Data collection and preprocessing

The grid database of year 2010 from Statistics Finland, similar to common census databases, was used to

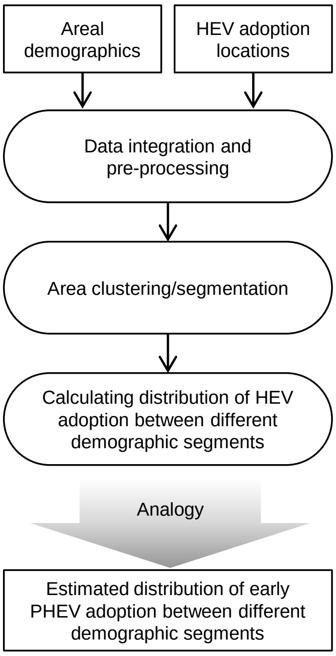

Figure 1: The main elements of the performed analysis.

capture the demographic and socio-economic characteristics of different geographical areas. The database consists of 250 x 250 meter squares covering most of Finland. Some of the more coarsely populated areas however lack the socio-economic information for privacy reasons. The database contains several types of statistics, such as population structure, population's educational level, population's consumer structure, households' size and phase of life, households' consumer structure, buildings and living, jobs by industry and population's main activity. Depending of the statistic, the information in the database represents the situation in the end of year 2009, 2008 or 2007.

The HEV adoption data was obtained from the Finnish Transport Safety Agency's Vehicular and Driver Data Register. The register contains all the vehicles in Finland registered for use on the public roads. The owner and the holder of each vehicle are included, as well as their street addresses. As of writing this paper there are no PHEVs in the Finnish market. However, HEV models that do not have the possibility of charging from the grid have been available for several years already, the most common model being Toyota Prius. The HEV data set contains all the Toyota Prius and Honda Civic Hybrid models registered by a private person in Finland until March 3rd 2011. For this study, the most valuable pieces of information available from the register are the vehicle holder's street address and the vehicle's registration date, which give access to the vehicle owner's location and the vehicle's coarse time of purchase. Even though the hybrid electric vehicles have been in the market for some time, the adoption rate has been low. As a consequence the distribution of HEVs is highly sparse.

The street addresses had to be geocoded first to geographic coordinates, after which they were spatially joined to corresponding statistic squares. Moreover, the non-inhabited squares were removed. The resulting data set contains 36990 statistic squares and 1227 HEVs.

The complete grid database includes 99 demographic variables, many of which are highly cross-correlated. To capture only the relative characteristics of the areas, the variables affected by the amount of people were turned into percentages. Furthermore, all the variables were variance scaled.

The variable amount was reduced to 24 by performing a selection with the aim of having a set of variables that are easy to interpret and not highly correlated with each other. First, variables clearly unrelated to the adoption of HEVs, such as the amount of workplaces by field of business, were removed. Furthermore, a covariance matrix and p-values corresponding the confidence levels of the covariance were calculated for the rest of the variables. Utilising the covariance matrix, a set of variables was selected so that no pair of variables in the set has covariance higher than 0.8 or lower than -0.8 at the 95% confidence level. In case of two or more variables with high covariance, only the variable having the easiest to interpret causal relationship with HEV adoption was kept. The included variables are listed in Table 2.

2.2. Self-Organizing Map

The self-organizing map (SOM), developed by Kohonen [10], is a well known vector quantization and clustering method which can map complex, non-linear relationships inside the data to lower-dimensional space while still approximately retaining the original topological and metric properties.

The map describing the output space consists of neurons which in turn are associated with prototype vectors. When training the SOM, each input vector is measured against the prototype vectors. The most similar prototype vector and its neighbourhood, determined by some kernel function, are updated by moving them closer to the corresponding input vector. When the training is done, the prototype vectors of the SOM have adapted to represent the common characteristics of the training data. The SOM also tends to approximate the probability density of the original data.

The smoothing of the data which occurs due to mapping it into more general prototype vectors has the consequence of producing larger error for the rare data and smaller error to the more commonly occurring data. As such, the SOM can be thought of as a (lossy) data compression method. Here we have used the SOM to compress the data so that it is faster to process by the K-means algorithm as demonstrated by Vesanto and Alhoniemi [11]. Furthermore, they show that the quality of the final clustering is comparable to direct clustering by K-means algorithm. For training the SOM, we used the default parameters of the SOM toolbox.

2.3. K-means clustering

The K-means algorithm [12] is a popular non-hierarchical clustering method, where the data is clustered to a predefined amount of clusters k. Here we have used it to partition the SOM output space into representative groups. As the optimal size of k is unknown, it is estimated by running the algorithm iteratively with different sizes of k and by measuring the Davies-Bouldin index [13] of each clustering. Essentially, the Davies-Bouldin index represents the ratio of intra-cluster scatter and inter-cluster distance.

The K-means algorithm is relatively slow with large amounts of data. Clustering the 36990 data rows directly by K-means would be a time consuming process. By using the SOM to compress the original data, in this case into 959 prototype vectors, the time required for processing is reduced significantly.

3. Data analysis

3.1. Finding an adequate number of clusters

For the analysis, the statistic squares were clustered into demographically representative classes. In order to find a good number of classes, the clustering was performed iteratively 50 times and the mean Davies-Bouldin index was obtained for each clustering. The resulting DB-indices for the different amounts of clusters are presented in Figure 2. The best (minimum) DB-index is achieved with a cluster amount of 14 or possibly more. However, for the sake of interpretability, we decided to limit the cluster amount to a local minimum of 5, which still produced a relatively good result. The smaller number of clusters also mitigates the problem of having so few cases of adoption, resulting in cluster statistics that are more robust.

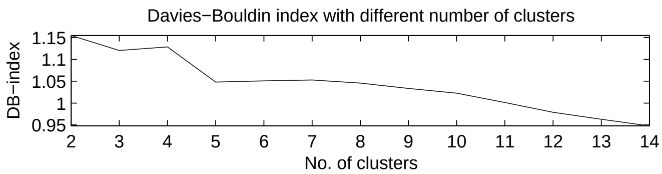

Figure 2: The Davies-Bouldin index for different number of clusters.

3.2. Interpreting the clusters

The 250 x 250 meter squares covering most of Finland's inhabited zones were clustered into 5 groups according to the socio-economic and demographic information found in the grid database. Moreover, the ratio of HEVs per household was calculated for each cluster. The distribution of squares, households and HEVs among the clusters is presented in Table 1.

As can be seen, in general the HEV adoption is still at a very low level. In comparison there are approximately 2 cars per household in Finland[1].

[1] The amounts of households and cars obtained from: http://www.stat.fi

We have further interpreted each cluster's characteristics by observing the cluster means listed in Table 2. Textual descriptions based on the cluster means, are shown in Table 3.

Figure 3 gives a better idea of the interrelations between the variables and the amount of HEVs adopted per household in each cluster. Variables having strong correlation (r>0.8 or r<-0.8) with the HEV adoption ratio are Average age, Secondary education, Bachelor's degree, Higher academic degree, Average household size, Households of under 7 years old children, Households of 7 to 12 years old children, Households of 13 to 17 years old children, Households of 18 to 64 years old with no children, Households of over 65 years old pensioners, Average income of households, high-income households and average floor area.

There seems to be evidence that the characteristics of the areas related to income level, education level, the amount of children and the apartment size are highly interrelated with the amount of HEVs adopted per household. Since the clusters are different by these attributes, also the HEV adoption ratio is different. However, these findings may apply only when looking at the whole country and one point of time. Further analysis is needed to confirm if and how the same patterns are occurring in different regions and time intervals.

3.3. Temporal analysis

In order to confirm whether the same socio-economic patterns have been affecting the HEV adoption behaviour in the past, we analysed the adoption in each cluster year by year. To get better sense of how the total adoption of HEVs has distributed between the clusters each year, we simply scaled the actual adoption ratios of each cluster by the sum of the adoption ratios of all the clusters as shown by equations 1 and 2. The relative adoption ratio RAR_i of cluster i can be denoted as

$$AR = \frac{HEVs}{Households} \tag{1}$$

$$RAR_i = \frac{AR_i}{\sum_{j=1}^{N} AR_j} \tag{2}$$

where AR_i is the adoption ratio of cluster i, N is the total amount of clusters and AR_j is the adoption ratio of cluster j.

As we have used the grid database of year 2010, some of the characteristics of the areas might have been slightly different in the preceding years. Nevertheless, we assume that the characteristics are unchanged for the most part. Thus we compared the relative ratios of HEVs adopted per household in each cluster from January 2004 to March 2011. The results can be seen in Figure 4b and Table 4.

The relative adoption ratios between the clusters have remained rather similar in years 2008-2010, the difference

	C1	C2	C3	C4	C5
Stat. squares	4787	8809	5430	6820	11144
HEVs	248	106	73	476	324
Households	131565	244004	358361	892830	353543
HEVs per 1000 hh	1.89	0.43	0.20	0.53	0.92

Table 1: Cluster statistics

Variable	C1	C2	C3	C4	C5	Tot.
Women (%)	50	51	53	52	50	51
Average age (years)	32	46	51	40	37	41
Secondary education (%)	14	34	46	29	24	29
High school education (%)	7	4	3	10	6	6
Vocational school degree (%)	40	51	36	44	49	46
Bachelor's degree (%)	14	6	4	9	11	9
Higher academic degree (%)	20	4	2	8	10	8
Average hh size (inhabitants)	3.25	2.13	1.53	1.81	2.64	2.28
Hh of -34 yo. singles (%)	1	4	11	17	3	7
Hh of -34 yo. couples, no children (%)	2	3	2	7	4	4
Hh of -7 yo. children (%)	31	9	5	9	17	14
Hh of 7-12 yo. children (%)	30	9	4	7	17	13
Hh of 13-17 yo. children (%)	23	9	4	7	17	12
Hh of 18-64 yo., no children (%)	30	46	50	61	42	46
Hh of 65- yo. pensioners (%)	11	33	38	20	20	25
Average income of hh (k EUR)	82	43	22	36	57	48
Low-income hh -10555 EUR/y (%)	3	12	40	24	7	16
Medium-income hh 10556-36469 EUR/y (%)	36	70	45	64	58	57
High-income hh 36469- EUR/y (%)	57	18	5	12	34	25
Average floor area (m^2)	125	91	58	64	102	89
Average floor area per inhabitant (m^2)	39	44	38	36	39	40
One-family houses (%)	98	96	49	32	95	77
Rented apartments (%)	5	10	48	48	10	22
Employment (%)	92	91	72	89	93	89

Table 2: The cluster means

Cluster	Description
1	High-income, highly educated households of families with children, living in big one-family houses.
2	Medium-income, low or medium level education households of older couples living in medium-size one-family houses.
3	Low-income, low-education households of singles and older couples living mainly in flats. High unemployment rate.
4	Medium-income, medium education level households of young singles and couples without many children living mainly in flats.
5	Medium-to-high-income, medium education level families with children, living in one-family houses.

Table 3: Textual cluster descriptions

between highest and lowest values being 0.094 in cluster 1, 0.046 in cluster 2, 0.047 in cluster 3, 0.005 in cluster 4 and 0.014 in cluster 5. Overall, the greatest HEV adoption has occurred from 2008 onwards, which explains the smaller variation in the relative adoption ratios. The years with lower overall adoption of HEVs, such as 2004-2007, are still fairly similar in their distribution of HEV adoption between the clusters. However, the low amount of adopted HEVs increases the variation which can be seen in greater confidence intervals. The effect of low adoption can be noticed for instance in the case of cluster 3 where, initially in 2004, the share of HEVs was high, but later diminished. Additionally, in 2004 the amount of HEVs per household was still slightly greater in cluster 2 than in cluster 4. From

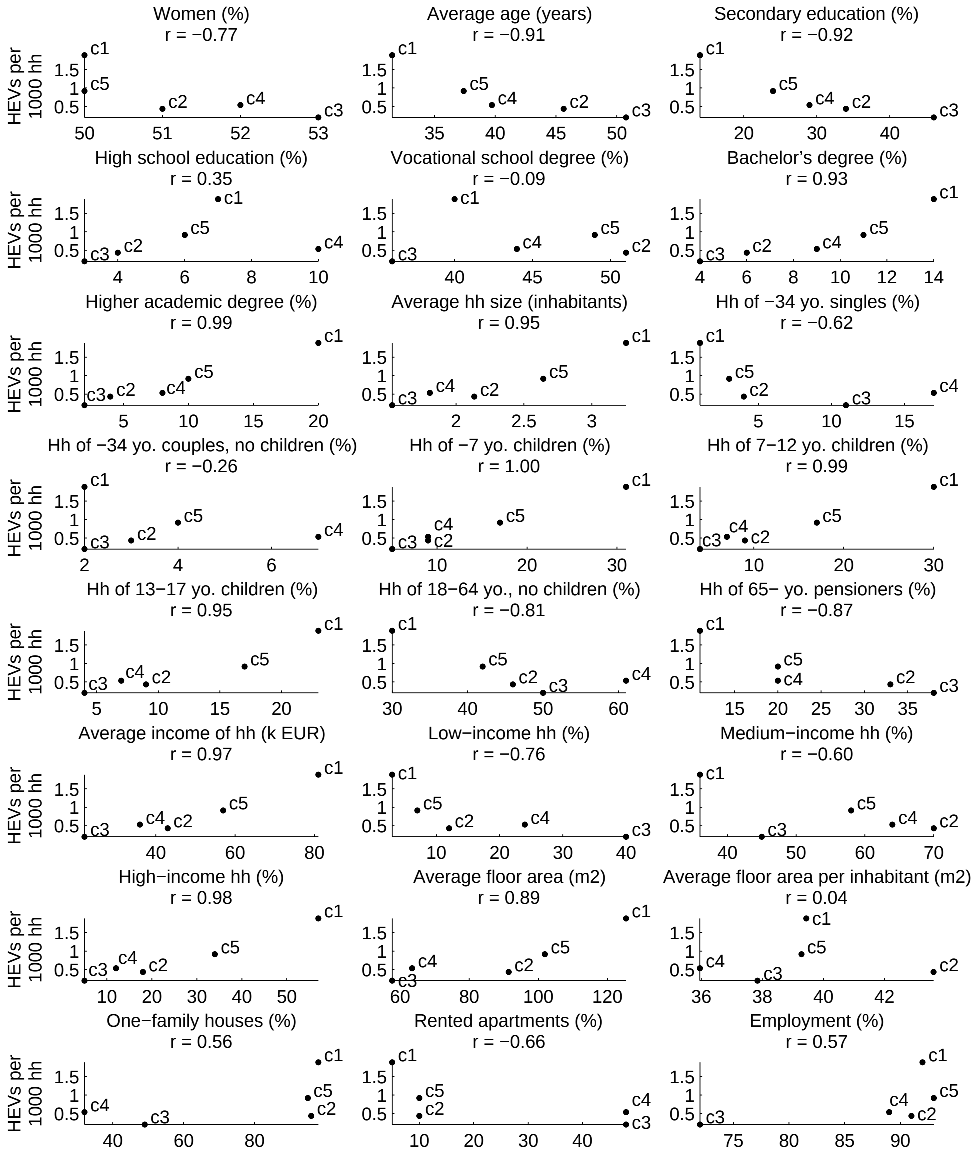

Figure 3: The correlation between each variable and the amount of HEVs adopted per 1000 households. The values are cluster means.

Year	C1	C2	C3	C4	C5
2004	$0.317^{+0.198}_{-0.176}$	$0.085^{+0.117}_{-0.064}$	$0.272^{+0.147}_{-0.124}$	$0.070^{+0.070}_{-0.038}$	$0.256^{+0.166}_{-0.122}$
2005	$0.484^{+0.248}_{-0.251}$	$0.075^{+0.207}_{-0.075}$	$0.025^{+0.145}_{-0.025}$	$0.082^{+0.094}_{-0.047}$	$0.335^{+0.214}_{-0.177}$
2006	$0.355^{+0.223}_{-0.191}$	$0.077^{+0.186}_{-0.077}$	$0.052^{+0.108}_{-0.052}$	$0.199^{+0.111}_{-0.090}$	$0.317^{+0.180}_{-0.145}$
2007	$0.516^{+0.143}_{-0.154}$	$0.111^{+0.118}_{-0.069}$	$0.013^{+0.058}_{-0.013}$	$0.142^{+0.077}_{-0.052}$	$0.218^{+0.113}_{-0.095}$
2008	$0.464^{+0.055}_{-0.060}$	$0.124^{+0.038}_{-0.028}$	$0.042^{+0.023}_{-0.016}$	$0.138^{+0.025}_{-0.021}$	$0.233^{+0.048}_{-0.037}$
2009	$0.449^{+0.076}_{-0.077}$	$0.123^{+0.046}_{-0.037}$	$0.070^{+0.034}_{-0.025}$	$0.138^{+0.030}_{-0.028}$	$0.219^{+0.056}_{-0.045}$
2010	$0.543^{+0.072}_{-0.080}$	$0.078^{+0.042}_{-0.033}$	$0.023^{+0.023}_{-0.013}$	$0.133^{+0.037}_{-0.029}$	$0.222^{+0.054}_{-0.052}$
3/2011	$0.672^{+0.197}_{-0.359}$	$0.121^{+0.387}_{-0.121}$	$0.000^{+0.000}_{-0.000}$	$0.082^{+0.151}_{-0.058}$	$0.125^{+0.257}_{-0.099}$

Table 4: The relative adoption ratios of the clusters in each year with corresponding 95% confidence intervals.

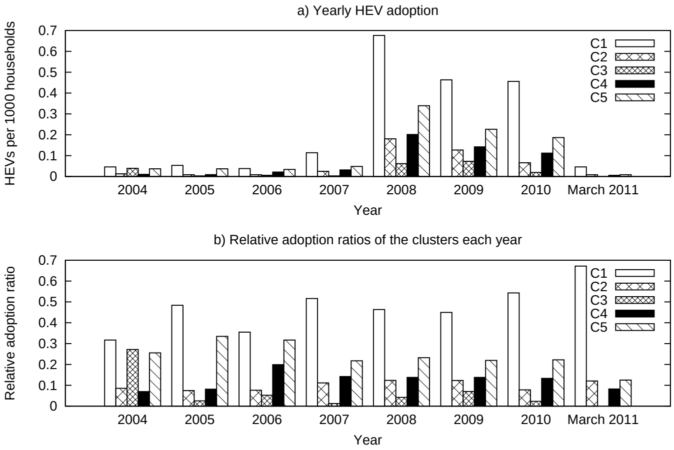

Figure 4: a) HEVs per 1000 households adopted each year in different clusters. b) Relative adoption ratios of the clusters in each year.

then onward the amount of HEVs per household has been greater in cluster 4.

As shown in Figure 4a and Table 5, the overall HEV adoption was rather low until 2008. The great increase in HEV sales was probably due to the 2008 tax reformation, which changed the taxation of new cars to be based on CO_2-emissions and so reduced the purchase price of HEVs substantially. Before 2008 the overall adoption was so low that any sporadic or atypical cases of adoption were noticeable, as seemed to happen for instance in the case of cluster 3 in year 2004, where the adoption ratio was much higher than later on.

The increase in the overall HEV adoption rate between years 2007 and 2008 was 579 percent. As a curiosity, the tax reformation decreased the price of Toyota Prius about 17%. Nevertheless, the yearly adoption ratio of HEVs per household was roughly similar in its proportion between the clusters even before 2008. It seems like the tax reformation did not change much the relative amount of willingness to purchase a HEV. Thus, the tax reformation was rather equally effective in all the clusters.

All in all, each cluster's share of HEVs adopted per household has been most of the time constant, indicating that the same socio-economic factors have been behind the HEV adoption. However, it should be kept in mind that when the diffusion process develops further, the factors behind HEV adoption or their importance can change, thus affecting also the adoption ratios of the areas represented by the clusters.

3.4. Regional analysis

To examine the distribution of the HEV adoption regionally, in contrast to the entire country, we compared three different regions with highest amount of HEVs per household. Figure 5a and Table 6 show the amount of HEVs per household in each region and cluster. The adoption ratio is highest in the region of Uusimaa while the region of Varsinais-Suomi has slightly higher amount of HEVs per household than the whole country. The Pirkanmaa region however has overall adoption ratio lower than the whole country.

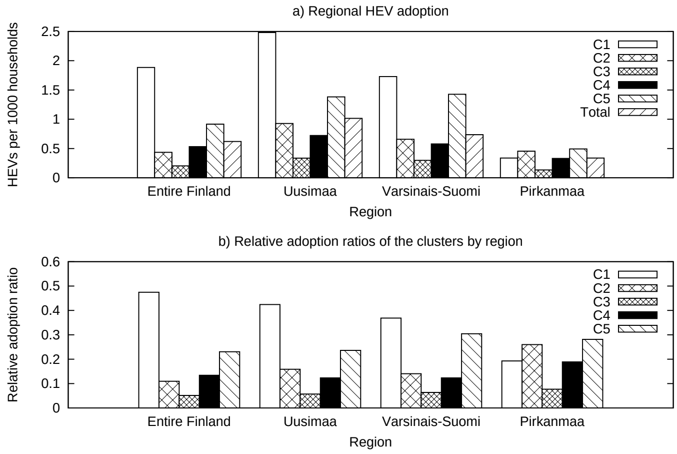

Figure 5: a) HEVs adopted per 1000 households in each region. b) Relative adoption ratios of the clusters in each region.

Calculating the relative adoption ratios as in equations 1 and 2, gives a better idea of the inter-cluster distribution of the HEVs. Figure 5b and Table 7 present how the distribution is roughly equal across the regions of Uusimaa and Varsinais-Suomi, while Pirkanmaa stands out in comparison to other regions with higher proportion of HEVs per household adopted in cluster 2 and lower proportion in cluster 1.

4. Results and discussion

The analysis shows that characteristics related to income level, education level, the amount of children and the size of the residence are highly interrelated with the amount of HEVs adopted per household in the studied areas. Supporting evidence can be found from the existing literature. For instance, Erdem et al. [14] found that higher income and education levels are characteristics of the consumers who are interested to buy hybrid vehicles. The strong correlation of the HEV adoption ratio with the amount of child families could be related to the predominant HEV model in the study data (Toyota Prius) being marketed more as a family car. Thus data with more varied selection of HEV models would be needed to verify whether this is the case.

The performed analysis gives an indication of the general HEV adoption tendency of different types of areas represented by the five clusters. The more specific temporal and regional analyses suggest that the inter-cluster relationships in the case of HEVs adopted per household have, for the most part, remained the same both annually and regionally. Yet, the regions or years with small amount of adoption events or low population are sometimes exceptions, which could be explained by the fact that in these cases the atypical adoption events have more chance of

Year	C1	C2	C3	C4	C5
2004	$0.046^{+0.053}_{-0.030}$	$0.012^{+0.021}_{-0.008}$	$0.039^{+0.025}_{-0.017}$	$0.010^{+0.009}_{-0.006}$	$0.037^{+0.025}_{-0.017}$
2005	$0.053^{+0.067}_{-0.031}$	$0.008^{+0.021}_{-0.008}$	$0.003^{+0.014}_{-0.003}$	$0.009^{+0.008}_{-0.005}$	$0.037^{+0.025}_{-0.017}$
2006	$0.038^{+0.047}_{-0.023}$	$0.008^{+0.020}_{-0.008}$	$0.006^{+0.014}_{-0.006}$	$0.021^{+0.011}_{-0.008}$	$0.034^{+0.024}_{-0.017}$
2007	$0.114^{+0.073}_{-0.046}$	$0.025^{+0.028}_{-0.016}$	$0.003^{+0.014}_{-0.003}$	$0.031^{+0.014}_{-0.010}$	$0.048^{+0.028}_{-0.020}$
2008	$0.676^{+0.152}_{-0.124}$	$0.180^{+0.058}_{-0.048}$	$0.061^{+0.033}_{-0.023}$	$0.202^{+0.031}_{-0.028}$	$0.339^{+0.064}_{-0.058}$
2009	$0.464^{+0.139}_{-0.107}$	$0.127^{+0.050}_{-0.040}$	$0.073^{+0.034}_{-0.025}$	$0.142^{+0.028}_{-0.024}$	$0.226^{+0.054}_{-0.046}$
2010	$0.456^{+0.127}_{-0.106}$	$0.066^{+0.040}_{-0.025}$	$0.020^{+0.020}_{-0.011}$	$0.112^{+0.024}_{-0.021}$	$0.187^{+0.051}_{-0.040}$
3/2011	$0.046^{+0.052}_{-0.030}$	$0.008^{+0.021}_{-0.008}$	$0.000^{+0.000}_{-0.000}$	$0.006^{+0.007}_{-0.003}$	$0.008^{+0.017}_{-0.006}$

Table 5: The yearly amount of HEVs adopted per 1000 households in each cluster with corresponding 95% confidence intervals.

Region	C1	C2	C3	C4	C5	Total
Entire Finland	$1.885^{+0.261}_{-0.225}$	$0.434^{+0.091}_{-0.078}$	$0.204^{+0.051}_{-0.039}$	$0.533^{+0.052}_{-0.055}$	$0.916^{+0.112}_{-0.095}$	$0.620^{+0.037}_{-0.036}$
Uusimaa	$2.481^{+0.377}_{-0.343}$	$0.928^{+0.498}_{-0.397}$	$0.335^{+0.207}_{-0.134}$	$0.722^{+0.100}_{-0.082}$	$1.382^{+0.229}_{-0.225}$	$1.016^{+0.082}_{-0.090}$
Varsinais-Suomi	$1.729^{+1.065}_{-0.724}$	$0.658^{+0.438}_{-0.278}$	$0.297^{+0.242}_{-0.144}$	$0.578^{+0.186}_{-0.170}$	$1.427^{+0.413}_{-0.393}$	$0.736^{+0.150}_{-0.125}$
Pirkanmaa	$0.337^{+1.671}_{-0.337}$	$0.455^{+0.636}_{-0.276}$	$0.135^{+0.317}_{-0.135}$	$0.330^{+0.341}_{-0.189}$	$0.492^{+0.581}_{-0.254}$	$0.338^{+0.159}_{-0.127}$

Table 6: The regional amount of HEVs adopted per 1000 households in each cluster with corresponding 95% confidence intervals.

Region	C1	C2	C3	C4	C5
Entire Finland	$0.474^{+0.038}_{-0.036}$	$0.109^{+0.025}_{-0.019}$	$0.051^{+0.014}_{-0.010}$	$0.134^{+0.016}_{-0.014}$	$0.231^{+0.027}_{-0.027}$
Uusimaa	$0.424^{+0.056}_{-0.053}$	$0.159^{+0.073}_{-0.060}$	$0.057^{+0.031}_{-0.022}$	$0.124^{+0.019}_{-0.019}$	$0.236^{+0.042}_{-0.036}$
Varsinais-Suomi	$0.369^{+0.136}_{-0.119}$	$0.140^{+0.082}_{-0.063}$	$0.063^{+0.049}_{-0.032}$	$0.123^{+0.050}_{-0.037}$	$0.304^{+0.099}_{-0.079}$
Pirkanmaa	$0.193^{+0.402}_{-0.193}$	$0.260^{+0.245}_{-0.176}$	$0.077^{+0.185}_{-0.077}$	$0.189^{+0.168}_{-0.132}$	$0.281^{+0.235}_{-0.180}$

Table 7: The relative regional adoption ratios of the clusters with corresponding 95% confidence intervals.

affecting the HEV adoption ratio. The above-mentioned issue is also more generally known as the "small numbers problem". The greater uncertainty of the low adoption or low population cases can be noticed from the greater confidence intervals. The regional differences can also point to the existence of spatial dependency in the adoption ratios. As such, the areas should either be considered individually or other means of handling the possible spatial dependency should be utilised.

The adoption ratios obtained from the clustering represent the current situation regarding the HEV adoption. As an example, Figure 6 shows a map visualisation of the HEV adoption from Helsinki metropolitan area. The different demographic/socio-economic areas represented by the clusters are illustrated with colours so that darker colour represents higher HEV adoption per household and lighter colour stands for lower adoption per household. Provided that the analogy between HEVs and PHEVs holds up and the same type of consumers obtain both, the same adoption ratios of the clusters can be used by decision makers to anticipate the PHEV adoption. However, as the adoption ratios are based on the early adoption phase of HEVs, they should be only considered to estimate the early PHEV adoption period. Even more meaningful representation of the adoption ratios could be achieved if the above-mentioned per household estimates were scaled by the number of households with cars in each area. Commonly the amount of cars owned by households changes drastically inside cities or metropolitan areas. For instance, downtown areas usually have less cars than suburbs and surrounding areas.

As for further improving the accuracy of the clustering, fuzzy membership could be utilised. The areas are similar to their respective cluster means only to some extent. In reality some areas could be closer to cluster border, in which case it would be more intuitive to rather use distance weighted mean of adoption ratio calculated from all the clusters. To facilitate the analysis, the number of clus-

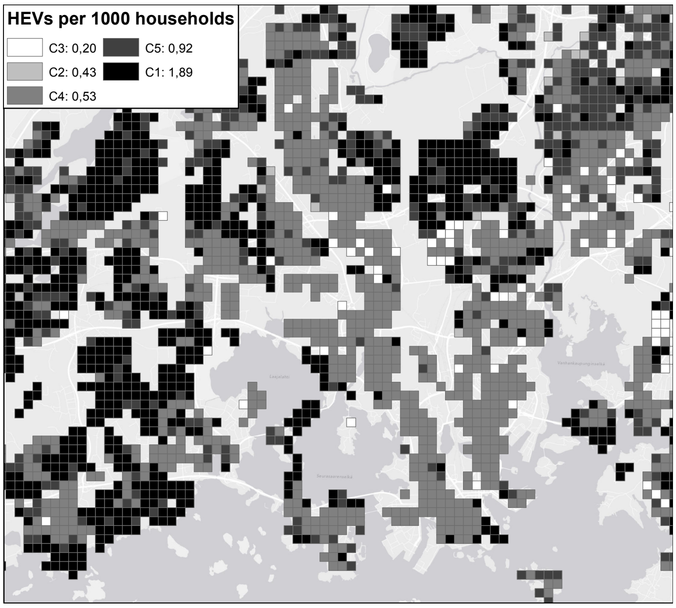

Figure 6: The current HEV adoption ratios of different demographic area types (clusters) in Helsinki metropolitan area. The rate of adoption per household is lower in squares with lighter colour and higher in squares with darker colour.

ters was limited to five. In reality the optimal number of clusters is higher. Producing larger amount of clusters would allow higher spatial granularity for the purpose of ranking areas by their HEV adoption potential. On the other hand, the point until which the granularity can be increased, i.e. until how many clusters the adoption ratios can be still assessed confidently, depends on the used data. The low quantity of HEVs adopted so far requires the granularity to be kept rather low. In the regional analysis, for instance, some regions with low adoption ratio (or population) incorporated less confidence on the cluster adoption ratios. An additional improvement would be to use spatially constrained clustering to present the clusters as spatially contiguous regions. A better spatial precision could potentially be achieved this way.

Being mobile by their very nature, HEVs and PHEVs can have multiple locations during the day. Only being able to assess their principal location (i.e. the vehicle owner's home area) might not be enough for all the planning applications. For instance, detailed electricity network planning requires also to consider the other potential places for charging, the time when charging takes place and the power used for charging. We realise this and therefore propose the presented analysis method as only one step in the whole planning process. The work presented here gives a good starting point for making more detailed statistical analyses, e.g., in conjunction with travel surveys. Nevertheless, while being a possible direction for further research, the question of how to utilise the presented analysis in different planning cases is not in the scope of this paper.

5. Conclusions

The main intent of this study was to analyse the adoption of HEVs in demographically and socio-economically distinct areas. We identified five different area types among which the relative HEV adoption ratio was observed to remain mostly similar when analysed both regionally and temporally. The analysis also suggests that the income and education level, the amount of families with children and the average size of the residences are highly interrelated with the amount of HEVs adopted per household in the studied areas. However, some bias might exist because of the limitations of the data used, such as the restricted amount of HEV models.

Since HEV and PHEV are similar in many of their attributes and can potentially be characterised as analogous products, the inter-cluster differences in HEV adoption ratios should resemble those of the upcoming PHEV adoption. Thus, the current cluster HEV adoption ratios can be utilised by planners and decision makers to anticipate the upcoming PHEV adoption in demographically distinct areas. However, HEV adoption still being in its early phase, any assessment based on the results should only concern the early PHEV adoption period.

Furthermore, we have demonstrated a possible way to utilise data from public registers in this context. The free usage of data collected by public funding is a growing trend at the moment, increasing the need for methods taking advantage of such data. Despite the compromises that have to be done regarding the data accuracy when using public data, a bigger problem here is the low amount of data generated so far from the actual market transactions of HEVs. The small number of HEVs makes it more difficult to perform analysis with high spatial granularity.

This study makes a contribution to the existing body of literature on (hybrid) electric vehicle adoption. The results could be utilised in a variety of practical applications, such as planning electric distribution networks, targeting policies to support cleaner vehicle adoption, marketing hybrid electric vehicles and locating charging stations.

Acknowledgements

The research was done as a part of Smart Grids and Energy Markets (SGEM) research program of the Finnish Cluster for Energy and Environment (CLEEN), whom we would like to thank for supporting this publication. The funding from Finnish Funding Agency for Technology and Innovation (Tekes) is also gratefully acknowledged.

[1] Herzog, T.. World greenhouse gas emissions in 2005. WRI Working Paper 2009;Available online at http://www.wri.org/publication/navigating-the-numbers.

[2] Lopes, J.A.P., Soares, F.J., Almeida, P.M.R.. Identifying management procedures to deal with connection of electric vehicles in the grid. In: PowerTech, 2009 IEEE Bucharest. 2009, p. 1–8.

[3] Shao, S., Pipattanasomporn, M., Rahman, S.. Challenges of phev penetration to the residential distribution network. In: Power & Energy Society General Meeting, 2009. PES '09. IEEE. 2009, p. 1–8.

[4] Bass, F.M., Gordon, K., Ferguson, T.L., Githens, M.L.. Directv: Forecasting diffusion of a new technology prior to product launch. Interfaces 2001;31(3):S82–S93.

[5] Tamor, M.A., Gearhart, C., Soto, C.. A statistical approach to estimating acceptance of electric vehicles and electrification of personal transportation. Transportation Research Part C: Emerging Technologies 2013;26:125–134.

[6] Heffner, R.R., Kurani, K.S., Turrentine, T.S.. Symbolism in california's early market for hybrid electric vehicles. Transportation Research Part D: Transport and Environment 2007;12(6):396–413.

[7] Turrentine, T.S., Kurani, K.S.. Car buyers and fuel economy? Energy Policy 2007;35(2):1213–1223.

[8] Lassila, J., Haakana, J., Partanen, J., Koivuranta, K., Peltonen, S.. Network effects of electric vehicles - case from nordic country. In: CIRED 2011 - The 21st International Conference and Exhibition on Electricity Distribution, Frankfurt. 2011,CD-ROM, art. 773.

[9] Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J.. Som toolbox for matlab 5. Tech. Rep.; Helsinki University of Technology; 2000.

[10] Kohonen, T.. Self-organizing maps; vol. 30 of *Springer series in information sciences*. Berlin: Springer; 1995.

[11] Vesanto, J., Alhoniemi, E.. Clustering of the self-organizing map. IEEE Transactions on Neural Networks 2000;11(3):586–600.

[12] MacQueen, J.B.. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J., editors. Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability; vol. 1. University of California Press; 1967, p. 281–297.

[13] Davies, D.L., Bouldin, D.W.. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1979;PAMI-1(2):224–227.

[14] Erdem, C., Şentürk, I., Şimşek, T.. Identifying the factors affecting the willingness to pay for fuel-efficient vehicles in turkey: A case of hybrids. Energy Policy 2010;38(6):3038–3043.