On the real range-need of electric cars: a telematic-box data-driven analysis

Gianluca Savaia *, **Simone Formentin** *, **Silvia C. Strada** *, **Sergio M. Savaresi** *

* *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy.*
Email: `silvia.strada@polimi.it`

Abstract: Electric vehicles represent an effective weapon against pollution and global warming. In the last years, there has been an increasing effort by scientists, carmakers and governments to encourage the use of electric alternatives to conventional cars. Nevertheless, the market share of electric vehicles was still less than 2% in 2017, the main reason being the perception that battery ranges and infrastructures are not yet ready to satisfy the drivers' needs. In this paper, we exploit a massive dataset of 35M trips for over 60k vehicles in a metropolitan city of Italy, to show instead that electric vehicles are already a feasible solution. Specifically, we show that, even if no public infrastructure is available, only 4% of existing vehicles cannot be turned into an equivalent electric car when considering also high-end vehicles with long battery ranges.

Keywords: electric vehicles, automotive industry, statistical analysis, big data, economics

1. INTRODUCTION

Electric mobility is undergoing a positive trend in the automotive industry: the market of electric vehicles has grown by 500% since 2013. However, this corresponds to a mere market share of approximately 1.2%, that is a world market where still more than 130 million Internal Combustion Engine (ICE) vehicles are sold against only 260 thousand electric vehicles (EVs), AlixPartners [2017].

Since the mass acceptance of EVs strongly depends on the consumers' perception, understanding people major concerns about EVs has been an important research topic during the last years, see Franke and Krems [2013], Rezvani et al. [2015]. An important conclusion of such studies is that EVs are not taken into account as a reliable alternative to conventional cars mainly because their driving range is not accurately predictable, a phenomenon well known as "range anxiety" which express people's concern of remaining stranded, Coffman et al. [2017]. It follows that, despite the various public incentives and the effort made by the carmakers to offer appealing and competitive electric solutions, the average user still sees the purchase of an electric vehicle as a possible option only in the far future. From the customer's perspective, there still is technological challenge concerning the charging infrastructure, Guo et al. [2018], as well as battery autonomy, Eberle and Von Helmolt [2010].

Although any improvement along the above directions would be definitely advantageous, in this paper we show that the existing technology, as far as the *range anxiety* is concerned, would allow a considerable mass switch to electric mobility without need of changing today's habits.

In more detail, we analyse a significant dataset covering the real driving pattern over a calendar year of more than 10% of the circulating vehicles in a metropolitan city of Italy. The sample dataset is remarkably significant in view of the fact that Italy is the world's first market in telematic devices, as shown in Re [2017], which allowed us to deal with a heterogeneous set of drivers' habits and status.

We evaluate the feasibility of switching to electric vehicles under the assumption that only nightly charges are possible, which correspond to assuming no public infrastructure is available yet. Even within this limited framework, we show that the average usage of ICE vehicles is perfectly compatible with the range provided by existing electric alternatives; results show that *more than half of the sampled vehicles could be turned into entry-level EVs without ever running out of battery.*

We make use of a truncated estimator which accounts for trips out of the ordinary (e.g., holiday trips) by discarding extreme observations. The analysis of the truncated dataset shows that user perception is strongly biased by such extreme points and that *only 4% of existing vehicles are often engaged in long distance trips throughout the year and could not switch to electric propulsion.* This number could be further reduced in case energy-efficient path planning policies are adopted, as shown in our recent contribution, Galluppi et al. [2017], or if day-time charging is also considered (18% of charges are not made at home in the United States, Smart and Schey [2012]). Furthermore, Rolim et al. [2012] show how the driving style is significantly affected by the vehicle propulsion, and how drivers could reduce the energy consumption as a sole consequence of driving an EV rather than an ICE vehicle; hence, the figures presented in this article only set a lower bound to electric adoption.

Eventually, some example applications which exploit this dataset and methodology are presented. The estimators previously discussed are employed in order to give an economical perspective to EV adoption from two different points of view: battery range and costs based on a toy scenario.

In our view, the obtained results represent a strong encouragement to a mass switch to electric mobility, and they are backed up by objective and measurable data which confirm some theses already discussed in other papers supported by coarse-grained surveys, see Needell et al. [2016], van Haaren [2011], Haugneland and Kvisle [2015], Axsen et al. [2018]. For scientists and control people, it is also a strong encouragement to keep on working to improve the electric vehicles technology.

2. PROBLEM STATEMENT

The findings of this research are based on the analysis of real trips recorded by in-vehicle devices (i.e., black boxes) whose primary application is in the field of insurance telemetries. The black box device can record various pieces of data which could be exploited to extract useful information on the driving patterns of users.

The 60 thousand vehicle fleet under observation is relative to a metropolitan city located in Southern Italy, and roughly represents 10% of the registered vehicles in the area. The rich footprint of millions of trips is therefore statistically significant and unbiased by subjective scales; notice that this is a major weak point of most literature which builds upon survey-based datasets, whose reliability is strongly affected by human perception, Tversky and Kahneman [1974], and by the way questions are posed, Schwartz [1999].

The goal of this work is to tackle the so called "range anxiety", that is the omnipresent concern that a vehicle has insufficient range to reach its destination, and it is considered to be one of the major barriers to large scale adoption of EVs. Thanks to real driving patterns, it is shown that the current battery technology and the available infrastructure would allow a wider adoption of EVs already today, without interfering with the present (real) driving habits. The scarce penetration of charging stations and their long charging time are often blamed as another key obstacle to the widespread of EVs; hence, in this research we make the assumption that no charging infrastructure is available apart from the own household.

In particular, our working hypothesis consists of assuming *only* nightly charge, whilst no day-time charge is considered; this worst-case scenario turns out to be not so unrealistic, especially in the geographic region under observation. Nonetheless, it is quite a strong limit in a general sense, considering that the Norwegian Vehicle Association indicates that almost 60% of the drivers have access to charging stations at their workplace and supercharging infrastructure is getting implemented throughout Europe, Haugneland and Kvisle [2015].

3. THE EXPERIMENTAL DATASET

The dataset has been kindly granted by a major insurance company practising in Italy. It provides their customers the possibility to install a black box on their vehicles, wired to the battery which operates as power source.

The black box can record various information on the vehicle status at a high sampling frequency. Each record, at each given instant, can store: timestamp, current speed and position, cumulative distance with respect to the previous observation and the event type; this latter attribute allows us to recognise whether the vehicle was started, turned off or it was travelling.

The dataset refers to the calendar year 2016 and contains roughly 300 million records, recorded from more than 60 thousand vehicles. We assume the insurance's customers constitute a representative and consistent sample of the population living in the province under analysis. Before these reasons, the authors can assert the data are statistically relevant and unbiased by subjective perception, on the contrary of surveys-based datasets.

The singular records previously described are associated to a each elementary event (e.g., vehicle started), but we are interested in extracting a comprehensive piece of information about each complete trip. Therefore, singular records were grouped into a key-on and key-off pattern, exploiting the event flag recorded by the black box; this process resulted in a dataset whose entries corresponded to a single trip described by start-end timestamps, start-end locations and overall distance travelled.

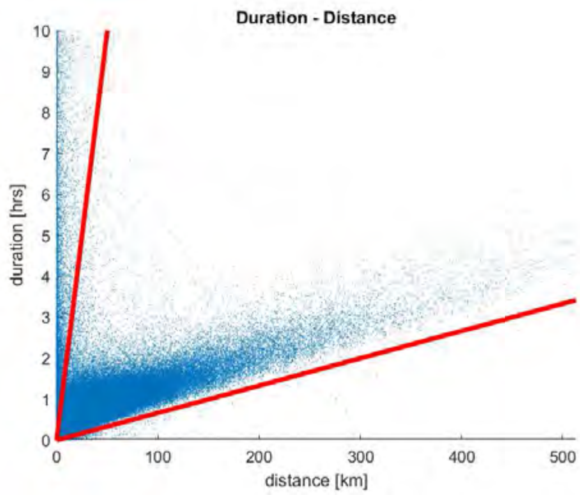

Fig. 1. Trips mapped into a distance - duration plot.

The 300 million events recorded by the black boxes turned into a new dataset containing 53 million trips related to more than 60 thousand vehicles, for an overall measured distance of roughly 300 million kilometres. This sample corresponds to 10% of the overall vehicles registered in the city of interest, which sums up to 600 thousand according to the national records.

Considering the dataset consisted in a raw log relative to thousands of devices, data needed to be filtered from outliers and all of those pieces of data which originated from a fault of the acquisition system. This phase is essential in order to ensure the drawn results are not affected by bad data; the authors, as further elaborated in the following, chose to cut a major part of the dataset to guarantee its validity and significance.

Primarily, every trip which recorded a zero-travelled distance is not relevant to the analysis and was therefore discarded; these trips correspond to those cases when the car is started and stopped without moving or, trivially, when the GPS suffered some fault and could not record a valid path.

In a second step, each trip was assigned an average speed by taking the ratio between the trip distance and duration, as depicted in Figure 1. It has been observed that some trips could not correspond to reality, since their average speed was too little or to high. Hence, two velocity thresholds have been identified as lower (5 km/h) and upper (150 km/h) boundaries for the cluster of feasible trips; those entries which fell outside that region were ultimately discarded.

Eventually, the trips recorded during the first semester of 2016 were discarded because the number of vehicles involved in the data collection in that period was not statistically relevant. Therefore, only the second semester was exploited to extrapolate the information of interest.

Despite the relevant disposal of spurious information, the final dataset consists of about 35 million trips (34% cut from the original) recorded from 60 thousand vehicles.

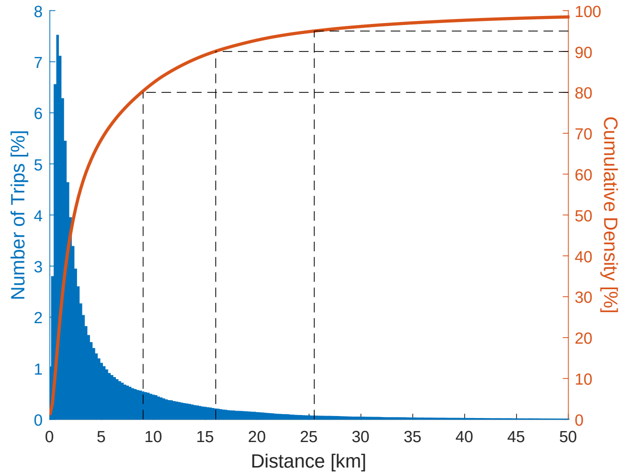

Fig. 2. Empirical density (blue) and cumulative (red) frequency distribution of all trips as a function of the covered distance.

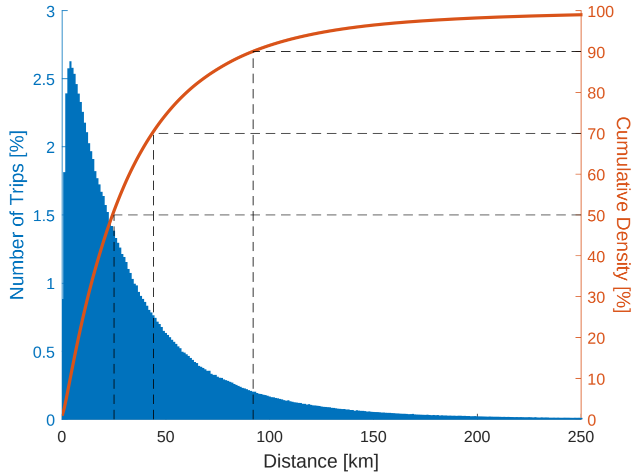

Fig. 3. Empirical density (blue) and cumulative (red) density distribution of all DVKTs as a function of the covered distance.

4. PRELIMINARY DATA ANALYSIS

The available dataset contains over 35 million trips, where each trip is intended as a key-on and key-off sequence. In order to find out how far people drive between the start and stop of their vehicle, the dataset is binned with respect to the measured driven distance per trip, resulting in the empirical density function shown in Figure 2. As one may expect, the distribution is positively skewed, meaning that most trips are within $5km$ with a tail wearing out beyond $50km$.

This preliminary simple analysis shows that the trips length seems highly compatible with off-the-shelf batteries: an electric battery autonomy of $100km$ would in fact cover 99% of the demand.

However, since we assumed only nightly charging, the electric battery range must cover the sum of the lengths of each individual trip undertaken within the same day. For the sake of conciseness, from now on we refer to the aforementioned sum as Daily Vehicle Kilometers Travelled (DVKT), which is the common acronym used in literature to refer to the overall distance driven in a day by a specific vehicle, van Haaren [2011]. Not surprisingly, the DVKT distribution presents a higher density on the leftmost side (shorter daily distances) with a long tail, but the order of magnitude is much higher than in Figure 2, reaching $250km$ as it is shown in Figure 3. The cumulative distribution leads to the following consideration: a battery range of $100km$ could satisfy 90% of the demand and a battery range upgrade to $250km$ is sufficient to cover more than a 99% of the occurrences.

According to the present data, considering a likelihood perspective, there is only a 10% probability for a vehicle to travel further than $100km$ in a day. This interpretation, although correct, tends to return an overconfident output which is biased by a veiled assumption, that is that all the drivers' habits are the same; indeed, as data were laid out, the trips might have belonged to the same vehicle.

In the next section, a more refined analysis is carried out, where each vehicle is profiled separately from the others.

5. VEHICLE PROFILING

If data relative to a single vehicle were extracted from the dataset, its trips length distribution would obviously

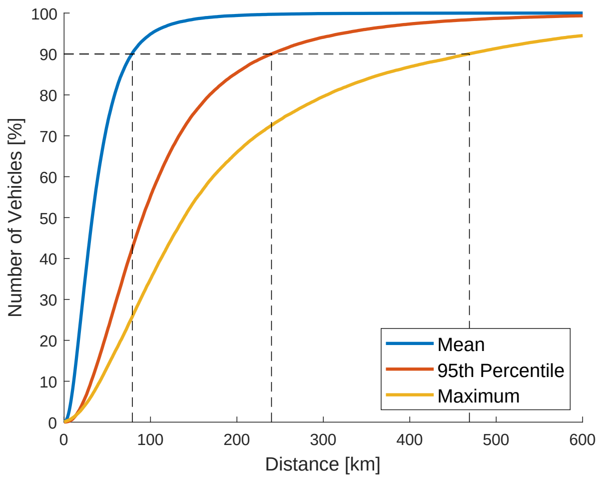

Fig. 4. The cumulative distributions of the metrics describing each vehicle.

resemble the overall density function shown in Figure 3. Of course, each driver has different habits: the trips length distribution may lean towards right if he travels longer than average, or towards left if the vehicle is only used to cover short distances.

Each vehicle is then characterised by means of three metrics, namely condensing its DVKT distribution into three scalar values: mean, 95th percentile (distance covered in 95% of the days) and maximum DVKT travelled. The mean DVKT is the most natural attribute one may think of in order to represent a vehicle, but it cannot give evidence of the long tail which is typical of these distributions. Hence, the 95th percentile DVKT is used to overcome the deficiency of the previous metric, since this value represents the distance travelled in 95% of the observations. At last, the maximum DVKT represents the worst-case scenario and, opposite to the mean DVKT, emphasises the extreme values. The distributions relative to the mentioned three metrics are depicted in Figure 4.

As expected, the distribution of the average trips is the most optimistic: 90% of the vehicles travel on average less than $100km$ a day. This result, albeit remarkable, is contentious since the mean value is a quite weak represen-

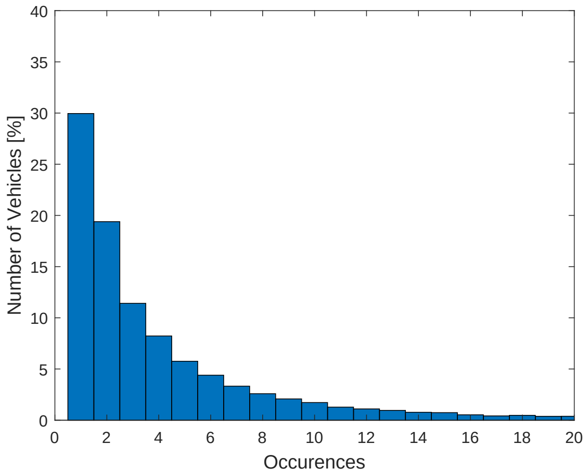

Fig. 5. Empirical density distribution of the number of vehicles as a function of the number of times that a vehicle has ever travelled further than 200km in a day.

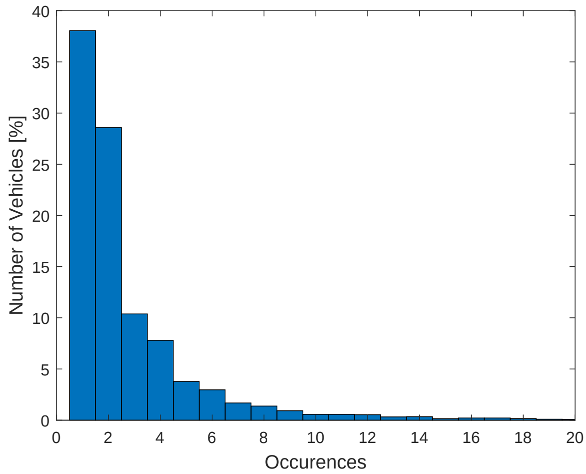

Fig. 6. Empirical density distribution of the number of vehicles as a function of the number of times that a vehicle has ever travelled further than 400km in a day.

tation, because it does not take into account how far the DVKTs are spread out from their average value. When the 95th percentile metric is considered, its distribution is spread out over a wider range of distances and gives a closer representation of reality. Yet, more than 50% of the vehicles (in 95% of the observed cases) did not travel further than 100km, and 90% of them did not go beyond 250km within a day. Although less pretentious, this result confirms a very high potential for the widespread usage of EVs.

When considering the 95th percentile, 5% of the trips are left out and these correspond to the longest distances which might drain the battery. Therefore, it is important to investigate the distribution of the longest distance ever covered in a day by each vehicle. Even this metric, which represents the scenarios out of the ordinary (e.g., holiday trips), generates some interesting figures: 40% of the vehicles have never travelled further than 100km, 75% covered up to 250km and 90% never topped more than 500km. These results confirm the fact that more than half the sample population under observation could have used an entry-level EV without ever running out of battery.

The three metrics characterise each vehicle from the best-case (average DVKT) to the worst-case (maximum DVKT) and help to identify the boundaries of interest in this scenario. In the following, the focus will be given to the outermost distribution, since drivers are interested in a vehicle which *always* satisfies their needs, and not only *on average*. As one may observe in Figure 4, when the longest distance ever travelled within a day is considered, the number of vehicles that could switch to electric propulsion significantly shrinks: 90% of vehicles drives on average less than 100km, but only less than 40% never exceeds that threshold. The information which is missing from these figures is the frequency of these occurrences, since such a long distance might have been travelled only once a year or so.

We investigate two distance thresholds which are of particular interest for market purposes, 200km and 400km, and we count the number of times a vehicle exceeds the threshold. The added value of this point of view is major: whereas 34% of vehicles travelled further than 200km (computed as the complementary distribution shown in Figure 4), more than half of them found themselves in that situation only in two occasions (Figure 5); the same observation is valid for the other threshold (Figure 6). Hence, we can conclude that the distributions analysed in the previous section are significantly biased by the extreme values, shaping the empirical density distributions towards longer distances.

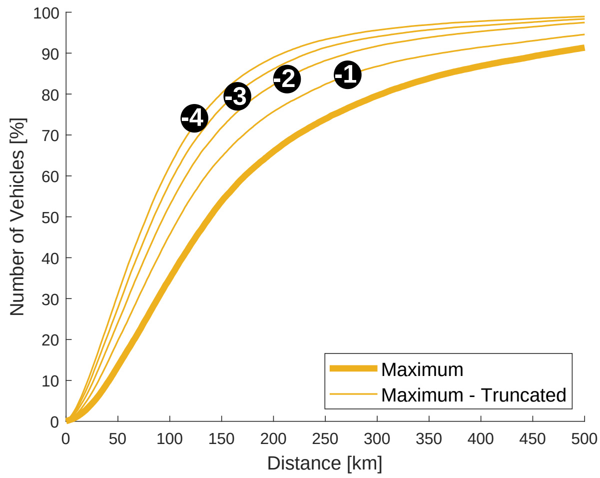

Fig. 7. Empirical cumulative density distribution of the number of vehicles as a function of the maximum distance ever travelled. The number of truncated trips is indicated on each curve.

The effect of the extreme values is investigated by means of an iterative "sample discarding" methodology: at each iteration, for each vehicle, its longest travel is removed from the dataset and the worst-case metric is recomputed; this procedure is repeated 4 times which corresponds to excluding 4 trips (i.e., 2 round-trips). The outcome shown in Figure 7 is that, by just removing a few samples, the distributions shift to the left, that is regressing towards the average.

The wide gap between the metrics which could be observed in Figure 4 is narrowed once unusually long travels are omitted. This confirms the fact that most drivers' routines are compatible with the electric range.

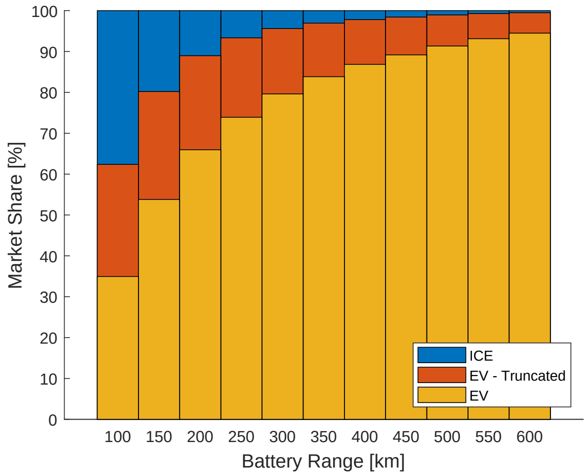

Fig. 8. Number of vehicles which could be replaced by EVs (yellow when range is always satisfied, red when dataset is truncated by 4 trips).

6. POTENTIAL EV ADOPTION IN A TOY SCENARIO

In this section, the penetration capabilities of EVs are analysed based on the previous results.

We concentrate on two significant classes of EVs characterised by different battery range: the first, groups all the vehicles with a 250km battery range, whilst the other, those with a 500km range. Indeed, these numbers refer to optimal driving style and environmental conditions; for this reason, we assume more realistic battery ranges to be 200km and 400km.

The previous results are rearranged in order to give a clearer view with respect to a market perspective: for each battery range, its market share is evaluated by investigating how many vehicles have a compatible driving behaviour, namely they do not travel longer than their battery capability within a day.

The toy scenario is then further elaborated in an attempt to develop a trivial decision making tool which can objectively define whether the adoption of an EV is economically advantageous for a particular vehicle.

Market share evaluation

Hereafter, the maximum daily distance travelled by each vehicle over one year is mainly considered because a potential customer is prone to purchase an EV only in case it can satisfy his needs even in the worst-case. In view of the observations regarding the extraordinary occurrences (Figures 4, 5, 6), the truncated dataset is superimposed; by considering these additional vehicles, we are actually assuming that during a long trip the driver has the possibility to make use of another mean of transportation rather than his EV.

Finally, the market size is estimated as the number of vehicles falling within the distance thresholds indicated in the abscissa of Figure 8; the results are clear: 66% of vehicles could switch to a 200km vehicle class without ever dealing with a battery drain. When the higher range capability is considered, the electric coverage is almost total, leaving only a small fraction to the rival ICE vehicles. In particular, only 4% of the vehicles have driving habits which are not compatible with the current electric ranges available on the market.

Moreover, the comparison between the 400km autonomy and 200km truncated of extreme values is appealing, since data indicate a range upgrade is not crucial when the assumptions are slightly relaxed. In such situations, renting an ICE vehicle might be a valid alternative and should be further investigated from an economic perspective.

Economic evaluation

In this section, we show how the evidence depicted in this paper could be exploited to take into account also some economic considerations and to investigate the transition towards electric mobility. Albeit the authors are well aware of the greater complexity of this matter, the following is to be intended as an example of the potentiality of the presented methodology.

Assume the scenario summarised in Table 1 where an EV is compared against its ICE counterpart: the former has a higher price tag than the latter, whilst the cost per kilometre advantages the electric propulsion. In this evaluation, we only consider a very simple comparison based on two attributes only: travelled distance and extraordinary trips exceeding a given distance; this setup could be further developed to include more features (e.g., public incentives, CO_2 emissions) whereas the problem statement would remain the same.

Each vehicle is profiled with respect to the total distance travelled throughout the year, and the number of occurrences it drove further than a given threshold, which in this example has been set to 200km.

Table 1. Toy scenario

Vehicle	EV	ICE
Fixed Cost	30'000 €	20'000 €
Energy Cost	0.2 €/kWh	1.5 €/l
Consumption	180 Wh/km	15 km/l

Each vehicle is assigned a total cost per year for the two propulsion systems, ICE and EV, computed as the sum of the following terms: i) fixed cost of the vehicle, ii) cost of the energy consumed with respect to the travelled distance and iii) penalty assigned for every occurrence of an extra-length trip (100 €, for EV only) which shall represent the cost of an alternative mean of transport (e.g., rental) when the battery range is not enough. The fixed cost per year has been obtained by dividing the fixed cost in Table 1 for a typical vehicle lifetime, assumed to be 10 years.

In Figure 9, each vehicle is assigned a cost per year for both the EV and ICE case; the black thick line is the bisecting line representing those vehicles whose driving style suits both propulsion systems, whilst the ICE-suited and EV-suited vehicles are found at its left (in yellow) and right (in blue) sides respectively. The dashed lines locate the origin of the points cloud and correspond to the fixed cost of the vehicle; a vehicle positioned in the origin has a zero travelling distance and hence no energy costs.

This visualisation makes it easy to see how the fixed costs play an essential role in the discrimination: the points cloud move over the plane with respect to the fixed cost and determines the convenience of a switch to electric mobility. The other attributes give a trivial result: the higher the travelled distance within a year, the more convenient the switch to electric, unless the number of extra-length occurrences out-values the energy cost in favour of ICE.

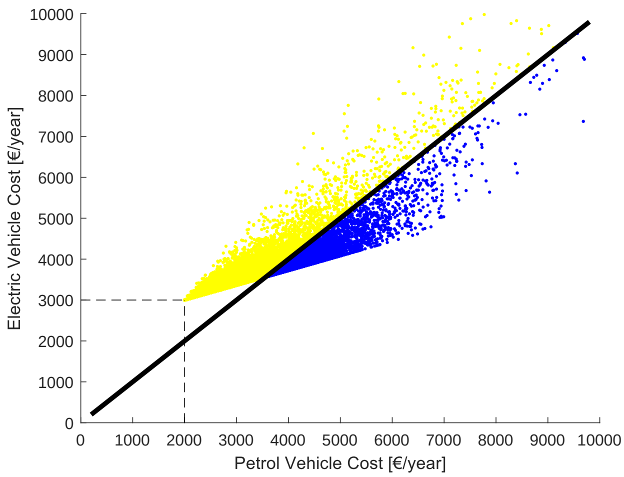

Fig. 9. Cost of ownership for each vehicle when considering ICE and EV propulsion.

In this brief application, the authors do not want to indicate any numeric figure but rather highlight that this methodology could be employed by different stakeholders in order to determine a tailored strategy driven by data itself.

Considering the fact that the fixed costs play a significant role, with the help of this tool, governments could identify the exact amount of public incentives needed to address a specific target population. Car manufacturers could exploit this information to adjust the battery range needed to address the market of a specific region or, on the contrary, they may address tailored advertisements to those vehicle's owners whose driving style is suited for EV adoption.

7. CONCLUSIONS

In this work, we employed a massive dataset of more than 35 million real trips collected from 60 thousand vehicles in a metropolitan city of Italy to show that the so-called "range anxiety" phenomenon comes from a misperception of reality. The authors do not assert that switching to electric mobility is a trivial matter, and recognise the multiple implications (e.g., economical) which affect this process.

However, this article shows that, from a technological point of view, a significant percentage of vehicles could be turned into their electric alternatives without running into any range shortage. This statement holds true even in the severe scenario where only night charging is available. Furthermore, the authors presented some example applications of the methodology which could help interested stakeholders in their decision-making policies.

The authors intend to further investigate in future works the relaxation of the two strongest assumptions which this analysis is based on: the possibility for every vehicle to be charged during the night, and the energy consumption model independent from ambient temperature and driving style.

ACKNOWLEDGEMENTS

This work has been supported by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE SpA and the Ministry of Economic Development - General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency. The authors would like to thank Filippo Colzi and Silvia Celaschi for their help and suggestions.

REFERENCES

AlixPartners (2017). Electrification index. `https://www.alixpartners.com/media/9854/second_quarter_2017_auto_electrification_index_nov_2017.pdf`.

Axsen, J., Cairns, J., Dusyk, N., and Goldberg, S. (2018). What drives the pioneers? applying lifestyle theory to early electric vehicle buyers in canada. *Energy Research & Social Science*, 44, 17 – 30. doi:https://doi.org/10.1016/j.erss.2018.04.015. URL `http://www.sciencedirect.com/science/article/pii/S2214629618303670`.

Coffman, M., Bernstein, P., and Wee, S. (2017). Electric vehicles revisited: a review of factors that affect adoption. *Transport Reviews*, 37(1), 79–93. doi:10.1080/01441647.2016.1217282.

Eberle, U. and Von Helmolt, R. (2010). Sustainable transportation based on electric vehicle concepts: a brief overview. *Energy & Environmental Science*, 3(6), 689–699.

Franke, T. and Krems, J.F. (2013). What drives range preferences in electric vehicle users? *Transport Policy*, 30, 56–62.

Galluppi, O., Colzi, F., Formentin, S., and Savaresi, S.M. (2017). A vehicle-user matching tool to encourage electric mobility. In *Intelligent Vehicles Symposium (IV), 2017 IEEE*, 1625–1630. IEEE.

Guo, F., Yang, J., and Lu, J. (2018). The battery charging station location problem: Impact of users' range anxiety and distance convenience. *Transportation Research Part E: Logistics and Transportation Review*, 114, 1–18.

Haugneland, P. and Kvisle, H.H. (2015). Norwegian electric car user experiences. *International Journal of Automotive Technology and Management*, 15(2), 194–221.

Needell, Z.A., McNerney, J., Chang, M.T., and Trancik, J.E. (2016). Potential for widespread electrification of personal vehicle travel in the united states. *Nature Energy*, 1(9), 16112.

Re, S. (2017). Unveiling the full potential of telematics. how connected insurance brings value to insurers and consumers: An italian case study. `http://media.swissre.com/documents/unveiling_the_full_potential_of_telematics_italy_case_study.pdf`.

Rezvani, Z., Jansson, J., and Bodin, J. (2015). Advances in consumer electric vehicle adoption research: A review and research agenda. *Transportation research part D: transport and environment*, 34, 122–136.

Rolim, C.C., Gonçalves, G.N., Farias, T.L., and Rodrigues, Ó. (2012). Impacts of electric vehicle adoption on driver behavior and environmental performance. *Procedia-Social and Behavioral Sciences*, 54, 706–715.

Schwartz, N. (1999). Self reports: how the questions shape the answers. *American Psychologist*.

Smart, J. and Schey, S. (2012). Battery electric vehicle driving and charging behavior observed early in the ev project. *SAE International Journal of Alternative Powertrains*, 1(1), 27–33.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

van Haaren, R. (2011). Assessment of electric cars range requirements and usage patterns based on driving behavior recorded in the national household travel survey of 2009. *Report*.