

Modelling decarbonisation of the transport sector with method for assessing vehicle driving cycles based on real GPS data

Luka Herc¹, Luka Perković², Tomislav Pukšec¹, Neven Duić¹

1 Faculty of Mechanical Engineering and Naval Architecture University of Zagreb, Ivana Lučića 5, 10002, Zagreb, Croatia

2 University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva ul. 6, 10000, Zagreb, Croatia

Abstract

This research presents a novel method for the statistical evaluation of the synthetic driving cycles for small-to-medium vehicles, based on the real driving cycles recorded with a GPS tracker with a resolution of five seconds. The recorded data is processed so it can be used as input for energy planning, namely the estimation of battery electric vehicles' energy demand and charging strategies in the dump, smart and V2G regimes. Initial statistical analysis shows that hourly distribution among various vehicles is best represented with gamma distribution. However, due to the lower amount of data recorded from the GPS, synthetic driving cycles match the data measurement with a correlation of 0,5 and 0,8 for workdays and weekends, respectively. This drawback can be avoided with more data being recorded during the research on the topic and consequent re-tuning of the distribution parameters. Also, the variations in the process are presented with the use of different combinations of statistical distributions and machine learning.

Keywords

Synthetic driving cycles, Battery electric vehicles, Energy planning, Energy demand

1. Introduction

In this work input data is taken from the GPS trackers mounted on numerous vehicles that have been recording data within Croatia for a period of a few months. Routes were recorded both inside and outside of urban areas and therefore they reflect both the short- and long-distance driving cycles. The input from GPS trackers was then processed in the in-house application that uses the algorithm presented in the methodology part of this research. The main output of data processing are synthetic driving cycles that are inputs for energy planning tools. The main input is battery power consumption and electric vehicle availability for charging. In this way, energy demand from electric transport and intermittent energy generation can be modelled in a coupled way to evaluate different options for sustainable transport, namely dump, smart and V2G charging. In the research, 2 approaches are explored of processing the captured data of which is one statistical analysis approach and the other is with the use of neural networks. Further on, the applicability of the developed methods and their results is tested on the linear energy system optimization model H2RES [1].

1.1. Literature review

The world faces a looming crisis in the form of climate change. Transport systems are among the biggest polluters and are responsible for significant portions of greenhouse gas emissions. The reductions of the emissions from the transport sector are among the plans of the European Union with the final goal of complete decarbonization [2]. Therefore, an action that aims to mitigate the emissions is required. Transport is one part of an energy system and should be considered in combination with the remaining

parts of the system [3], especially in the case of the application of distributed generation [4]. There are tangible benefits of introducing greater interconnectivity of the transport and power sector with technologies such as the smart charge that is applicable on the local scale through distributed generation [5] and on a system-wide scale which results in the greater utilization of renewable energy and reduction of critical excess of electricity production (CEEP) [6]. The application of V2X in the form as Vehicle to Grid (V2G) or Vehicle to Home (V2H) is beneficial in helping the management of the energy system [7]. It is crucial to state that smart charging strategies are required in order to match energy demand for charging with the availability of energy. For example, to account for the peaks in energy generation from renewable energy sources during the night and therefore to prioritize charging during that time [8]. These vehicles rely on batteries and therefore it is important to observe that the prices of batteries and electric vehicles (EV) have been decreasing steadily for the past decade [9] and will eventually become lower than for internal combustion vehicles [10]. Therefore, that makes them attractive in two ways. The first one is to replace fossil fuel powered vehicles, while the second one is to provide balancing capabilities to the energy system.

Hydrogen fuel cell electric vehicles (HFCEV) are also capable of ensuring decarbonization of transport system. Also, they can as well offer balancing services and flexibility [11]. The source of flexibility is in this case in the production of hydrogen since the electrolyzers can be controlled as demanded by the power system and hydrogen can be stored [12]. Also, the systems use hydrogen storage systems to mediate between the demand and generation. Decarbonization of the complete transport sector with only battery electric solutions is almost impossible as is the case with heavy-duty road transport [13], shipping [14] aviation [15] and other options with the use of hydrogen and electrofuels are required in such cases.

When modelled, all of the mentioned technologies, including the battery electric vehicles, fuel cell electric vehicles and application of electrofuels require the inputs for the distribution of transport demand and the distribution of availability of refiling energy storage in the vehicles themselves in the energy models such as EnergyPLAN [16], PLEXOS [17] or H2RES [1]. Therefore, the creation of the transport demand cycles and charging availability cycles from the experimental data is a significant contribution to the realism of modelling the transport sector and its interaction with the remainder of the energy system and economy overall in the energy modelling and optimization models.

Previous research on modelling the transition in the transport sector was performed by Pukšec et al. which presented in the EDT model that conventional vehicles can be phased out in Croatia as soon as 2040 [18]. The energy planning approach can be different and recent studies assume a priori penetration of BEVs at a certain rate, like the analysis from Prina et al. [19]. The approach with principal component analysis and k-means is used to develop a driving cycle for the city of Shenyang, China [20]. This paper relates the model to input data from GPS tracker but does not provide information regarding time-series of energy consumption. The specifics of BEVs, like battery power drain and battery swapping, were included in electric vehicle routing problem solving, as presented in [21]. The first approach to obtaining synthetic driving cycles was made by Novosel et al., where load curves for transport demand were obtained from the agent-based modelling in MATSim [22]. More recent developments of agent-based model, with bottom-up parameter tuning, were performed by Lin et al. in [23]. Arias and Bae presented forecasting EV charging demand for residential and commercial sites for bus networks during the season and presented the result as a charging demand time series [24]. Gallet et al. presented a simplified longitudinal dynamics model for the prediction of energy demand for buses, relying on synthetic driving profiles tuned with real-life data [25]. Su et al. presented a model for estimating daily charging patterns with emphasis on the heterogeneity of driving patterns from different modes of electric transport in the case of New Zealand [26].

1.2. Hypothesis and contribution of this work

The GPS data recorded in this work presents valuable information for the estimation of synthetic driving cycles. The recorded data already inherently contains information about traffic congestion, which represents driving with zero velocity within the drive cycle. The hypothesis is that with the use of correctly calibrated statistical distributions, users can create reasonably well-predicted driving cycles for the purpose of energy planning inputs. Statistical distributions differ in every hour of the day, therefore they are calibrated for each of the 24 hours to represent the intra-day difference in power consumption. Also, the methodology in this paper distinguishes between working days and weekend days, meaning that driving cycles differ substantially between the two. An additional novelty is the use of calculated energy demand data to train a neural network that will on the basis of a set of parameters indicating the day of the week, outside temperature, solar irradiation and precipitation output the energy demand data.

2. Method

The future transport systems will be partly electrified and therefore coupled with the power systems. Taking this into account, the methodology should cover all steps from well-to-wheel and can be divided into modelling the overall energy system, modelling the transport sector in detail, and modelling the driving cycles. Modelling the transport sector in detail comprises modelling energy consumption of various fuels, used as primary energy in the form of fossil fuel or biofuel derivatives, hydrogen or electricity that has to be stored in the so-called tank. Tanks can be fuel tanks, compressed hydrogen tanks or batteries. This tank consumption is included in the overall energy system as final energy consumption in the transport sector. In order to integrate the electrification of transport, i.e. consumption of "tank" electricity, consumption curves are synthetically derived from the input GPS data.

Methods section and the process itself can be divided into a couple of steps that were conducted:

2.1. Estimation of the driving cycles from the GPS data

Estimation of driving cycles is crucial for successful integration of BEV's into the energy systems. Besides the estimation of tank (battery) consumption of electricity throughout the day, the important parameters are trip departures and trip arrivals. They are used to estimate the periods when BEV is parked and possibly plugged into the grid to charge the battery or serve as auxiliary power source to the grid. In order to estimate the driving cycles, GPS measurements are used. Due to the relatively small amount of input data from the GPS measurements, the approach with fitted synthetic driving cycles is used to provide themore realistic data on a fleet of vehicles. Two approaches are used to generate the synthetic driving cycles: statistical approaches with the use of various statistical distributions for modelling energy demand, speed, and departure times and the method that used neural networks to correlate the parameters such as weather conditions and time of use with the energy demand. The main outputs of the statistical process or the neural networks are the energy demand distributions and charging availability distributions. These distributions are then used as an inputs into H2RES model to test their applicability.

2.1.1. Statistical method

The methodology for estimation of the driving cycles consists of several steps, each one represented by its own algorithm. The code representing the methodology is written in Python and currently stands as an in-house tool. The methodology is graphically presented in Figure 1.

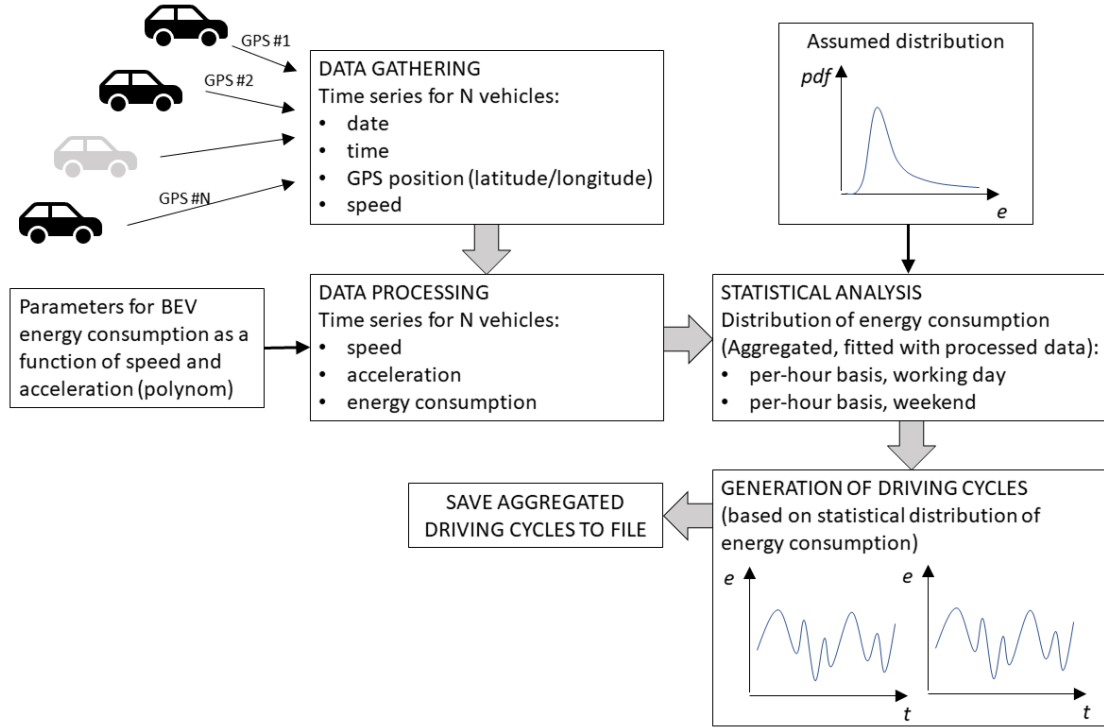


Figure 1 - Methodology for calculating synthetic driving cycles for smart charging of BEV's

The first step is to take the input data from the GPS tracker for all vehicles and gather them into a single database. The database contains all recorded cycles for each vehicle. Vehicles do not have to be BEVs, since it is assumed that eventually, owners will not change driving cycles if BEVs are used instead of conventional vehicles. Nevertheless, the overall methodology would provide better outputs of synthetic driving cycles if GPS data is recorded with BEVs since charging time would implicitly affect the input driving cycles recorded with GPS. The recorded data is gathered and processed into database containing the time-position pairs of each vehicle in each trip, from which speed, acceleration, tank energy consumption and trip-timing data like trip starts and trip ends can be stored for further statistical analysis.

The initial analysis determines the distribution of starting and ending trip hours, determined from the gathered GPS data. This results in a bimodal distribution for both starting and ending hours. The next step is to estimate the energy consumption during the individual drives, which depends on the driving style of individual drives, also gathered in GPS data.

In order to estimate the energy consumption of BEVs, the approach from Galvin is used [27] for the calculation of power demand at the battery as a function of speed and acceleration in time t :

$$P_{EV,t} = a_1 w_t + a_2 w_t^2 + a_3 w_t^3 + a_4 w_t \frac{w_t - w_{t-1}}{dt} \quad (1)$$

The equation has input parameters a_1 to a_4 , which have to be determined from the measurements on BEVs and which are listed in [27]. The above model does not take into account some features that can furthermore increase the battery power supply, such as change in elevation (uphill/downhill movement), and use of auxiliary services (heating/cooling), which can have a significant influence on the total power demand [28]. The next step is to recalculate power consumption into the energy consumption in time t in kWh:

$$e_{EV,t} = \frac{P_{EV,t}(dt/3600)}{1000} \quad (2)$$

Each time corresponds to a specific hour and day in the week. The next step is to reallocate the energy consumption into corresponding average energy consumption for each hour for all vehicles with recorded data:

$$\text{for each } t = 1, 2, \dots, 24: \begin{cases} \langle e_{EV,workday} \rangle(t) = \sum_{N_{vehicles}} \frac{1}{\sum \Delta t} \sum_t^{t+1h} e_{EV,t,workday} \Delta t \\ \langle e_{EV,weekend} \rangle(t) = \sum_{N_{vehicles}} \frac{1}{\sum \Delta t} \sum_t^{t+1h} e_{EV,t,weekend} \Delta t \end{cases} \quad (3)$$

In the Eq.(3) the averaging is done for each vehicle and every recorded time step, where working days and weekend days are separated due to the expected differences in driving patterns. Based on the average energy consumed by the average battery for each hour during the day and distinguishing between the workdays and weekends, an energy pattern can be used to fit some selected distribution which will be used for generating synthetic driving cycles. Synthetic driving cycles have to follow the pattern set by the recorded average energy consumption of the BEVs battery. For the initial analysis, the gamma distribution was used. The algorithm for fitting the distribution function can be implicitly described as:

$$\text{for each } t = 1, 2, \dots, 24: \begin{cases} F_{fit,workday}(t) (\delta_{fit_parameters}, \langle e_{EV,workday} \rangle(t)) \\ F_{fit,weekend}(t) (\delta_{fit_parameters}, \langle e_{EV,weekend} \rangle(t)) \end{cases} \quad (4)$$

For example, fitting parameters of gamma distribution, the shape g_1 , shift g_2 and scale g_3 have to be found by the fitting procedure based on the calculated values of average consumption in hour t , $\langle e_{EV,workday} \rangle(t)$ or $\langle e_{EV,weekend} \rangle(t)$, which again relies on the recorded GPS data. They are different for each hour in typical work day and weekend, meaning that there are total of 48 parameters to be found in matching procedure between the synthetic and GPS data. Based on the parameters of distribution, synthetic driving cycles can be estimated from selecting the hourly values of energy consumption randomly, following the probability of fitted distribution. For example, if gamma distribution is used, then synthetic driving cycles can be obtained by the

$$\text{for each } t = 1, 2, \dots, 24: \begin{cases} e_{EV,workday}^{Synthetic}(t) = \Gamma[g_{1,workday}(t), g_{2,workday}(t), g_{3,workday}(t)] \\ e_{EV,weekend}^{Synthetic}(t) = \Gamma[g_{1,weekend}(t), g_{2,weekend}(t), g_{3,weekend}(t)] \end{cases} \quad (5)$$

The synthetic driving cycles depend on the resulting parameters g of the gamma distribution (or any other selected distribution) which provide the expected occurrence of energy consumption of the BEV in specific hour t . The accuracy of the proposed method will be enhanced with an increased number of input data from GPS trackers. An increased number of data can be used to find more accurate fitting parameters for the selected distribution, or even change the distribution used as used for the fitting procedure.

The statistical method consists of several steps. The first one is the implementation of the driving cycles calculation with the use of the gamma cycle as described previously. After that, the approach was modified in order to remove the bias towards the excessive values stemming from the outliers in the input data. Finally, the approach testing different statistical distributions and their combinations was used for the calculations of the energy demand, duration and departure times were used. These were Gamma, Weibull and Normal distributions.

2.1.2. Neural network method

The input to this method is the calculated energy demand distributions on hourly levels for the hours when the parameters were tracked. Besides the energy demand, the data on the precipitation, solar irradiation, days in the week and non-working days were provided. An example of the structure of data used is displayed in Table 3. while the remainder of the method description is in the appendix

In short, the code utilizes Google Drive for data storage, The Python package Pandas is used for data manipulation, Scikit-learn for data preprocessing and error evaluation, and Keras for building and training the neural network model.

The neural network architecture consists of a sequential model with three hidden layers containing 128, 64, and 32 neurons, respectively, all using ReLU activation functions to capture nonlinear relationships in the data. Dropout layers are applied after the first two hidden layers to reduce overfitting. The output layer uses a single neuron with ReLU activation to ensure non-negative predictions of energy demand.

The model was trained using the Adam optimizer, Huber loss (to balance sensitivity to outliers), and early stopping to avoid unnecessary overtraining. The feature set included hourly time representations (encoded as sine and cosine to preserve circularity) and a binary weekend indicator, with a decision made to exclude temperature and precipitation since the measured data does not include the periods of time when the temperatures were low and precipitation which would introduce the bias into the predictive model.

In this approach, the distribution of availability for charging is based on the resulting energy demand distribution and experiments that acquired the data that the cars are on average parked 95 % of the time [29]. Therefore, the negative of the relative distribution for energy demand is scaled to match this experimental data.

2.2. Modelling of the overall energy system

Portion of the research includes the implementation and testing of the driving cycle distributions and the distributions of the vehicle availability into the energy planning and optimization model. For this task, H2RES model is used. The model performs the optimization of the capacity additions as well as the dispatch optimization for various units in an energy system. The model encompasses the whole power system sector. This section includes power plants and energy storage capacities. Also, the heating system is modelled, and it differentiates between individual heating and district heating networks. further on, the industry sector is modelled and finally, the transport sector. Additional defining characteristics of the model are the availability of flexibility options in the form of flexible power plant operation, vehicle-

to-grid system (V2G), power-to-heat (P2H), power-to-gas (P2G) and stationary energy storage. The more detail descriptions of the model can be found in the previous publications that use the H2RES model such as [30] that compares the model with an existing tool and the [31] that presents the extensive developments in the model, primarily in the heating and industry section. For the purpose of this research, the transport sector is most critical. The distribution curves are tested for the implementation into the H2RES energy system modelling and optimization software in order to ensure that the model is capable of solving the case with these curves. Meaning that they do not present excessive changes in the demand or the too long periods with no availability of replenishing the batteries.

This part of the model is tasked with ensuring that transport demand is met. The demand is given as a travelled distance demand in each hour through hourly distribution. With the use of efficiency parameters which define the energy demand for kilometres driven for various technologies, the energy demand in the sector is calculated. The model uses internal combustion vehicles (ICE), electric vehicles (EV), and fuel cell electric vehicles (FCEV). The vehicles with ICE engines use fossil fuels, while the electric vehicles use the electricity from the grid. Also, for the purpose of powering FCEVs, electricity from the grid is used to generate hydrogen. In the vehicles itself, the storage is modelled only for electric vehicles. Also, the number of vehicles has to stay consistent through the years. The model invests and decommissions the vehicles.

The battery-electric vehicles have energy storage implemented through their batteries and therefore are capable of providing flexibility services to the system. Fuel cell vehicles also can act as a flexibility option as the generation of hydrogen can be made flexible since the storage system is also used. All of the types of vehicles have defined investment costs in the reference year and corresponding learning curves that modify the prices in later years.

The basic schematics of an H2RES model is displayed in Figure 1.

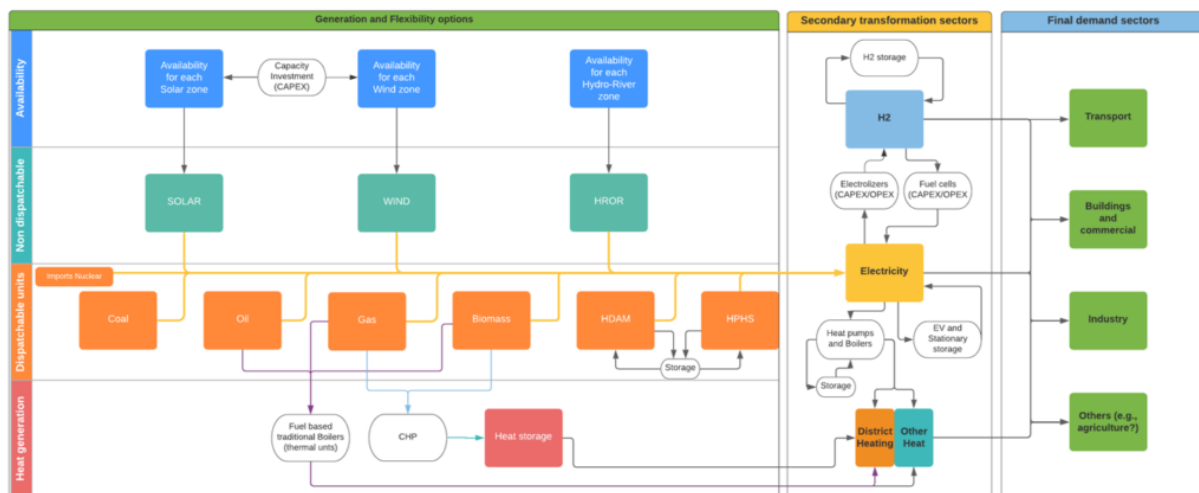


Figure 2. H2RES schematics [1]

3. Case study for Croatia

The GPS trackers were installed on conventional and electric vehicles for a limited amount of time in order to record the data that will be used for the estimation of synthetic driving cycles. Measured data includes the spatial position, speed, acceleration and time.

3.1. Case study setup and initial analysis of the recorded GPS data

GPS data were recorded with the time step of $\Delta t = 5$ s. A total of 1262 individual drives were recorded in the time span between September 2018. and April 2019. Initial analysis of the recorded data from the GPS trackers is presented in the form of histograms in Figure 3.

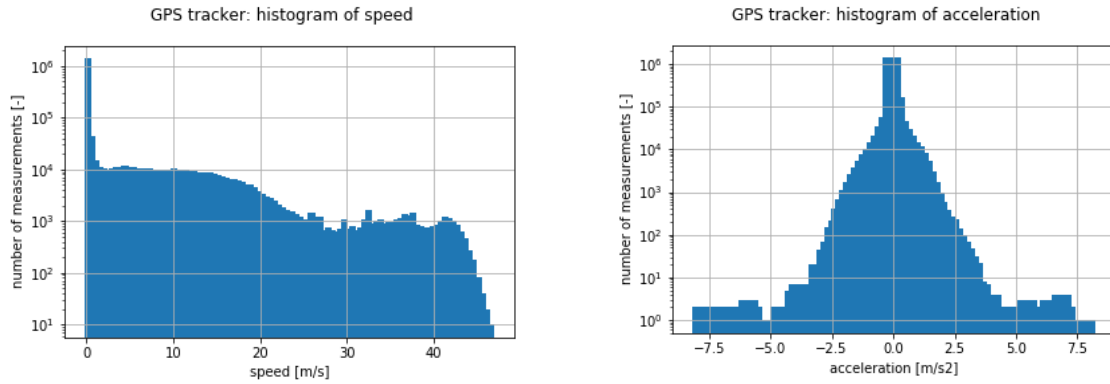


Figure 3 - Histograms of speed and acceleration, recorded with GPS trackers

Histograms indicate that during the driving a significant amount of data is recorded at zero velocity and acceleration of the vehicle. If individual cycles are analysed, in Figure 4, it can be seen that the majority of trip durations last less than one hour.

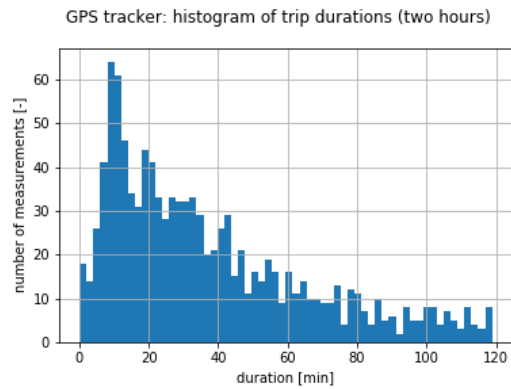


Figure 4 - Histogram of trip durations of individual driving cycles recorded with GPS trackers

The bimodal distribution is more emphasized in workdays days, than in weekends. Bimodal distributions are shown in Figure 5.

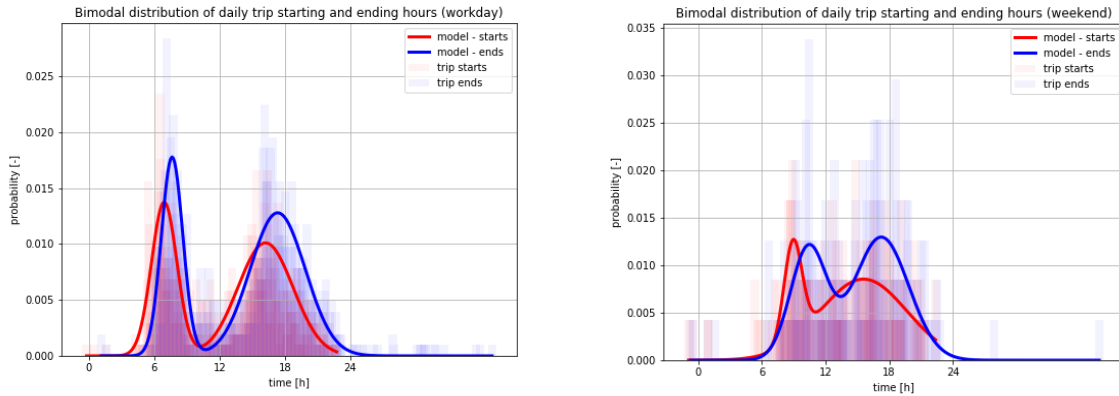


Figure 5 - Bimodal distribution of trip starting and ending hours with fitted bimodal distributions

3.2. Estimation of average battery power consumption from the recorded GPS data

The assumed parameters $a_1 - a_4$ of Eq. (1) for the estimation of battery power demand are derived from the cycle measurements done in [32], where total of eight BEV's were analyzed (NissanEV, Kia, Mitsubishi, BMW, Ford, Chevrolet, Smart, Nissan2012). In this analysis, the ninth vehicle is added – the hypothetical vehicle having parameters that are the average value from all eight vehicles ($a_i = \langle a_i \rangle$). With the use of Eq. (1), battery power consumption, as well as the energy used from the battery can be estimated for each GPS recorded measurement. In order to get the tank consumption per km, for each recorded driving cycle the battery consumption is divided with driven kilometres. Inverse value of tank consumption is tank efficiency. Results and comparing values from the European Commission's Joint Research Centre Tank-to-wheels report v5 [33] for BEV 200 and 400 miles range (named BEV-200 and BEV-400) and likely market-average technology development expected by EURCAR and AVL experts are shown in Figure 6.

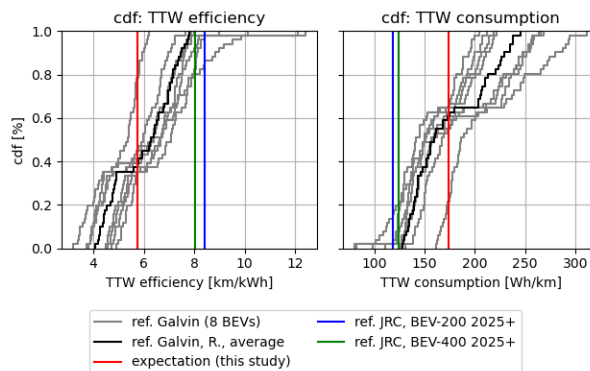


Figure 6 - Distribution of corresponding battery energy consumption based on GPS tracker data and the assumed electric vehicle

The efficiency of the hypothetical BEV with average coefficients $a_1 - a_4$ is 5.75 km/kWh, which is lower than expected values of more than 8 km/kWh for both BEV-200 and BEV-400 vehicles in JRC report. The tank consumption for hypothetical BEV, calculated as an inverse of the efficiency, equals to 173 Wh/km, which is higher than the JRC estimations. In further analysis, the value of 6 km/kWh is taken as an estimated tank efficiency of the BEV.

The distribution of energy consumption shows a bimodal distribution of per-hour averaged energy that follows the bimodal distribution of starting and ending hours of individual drives. The following figure

shows the relationship between battery power consumption and driving parameters, like acceleration and speed.

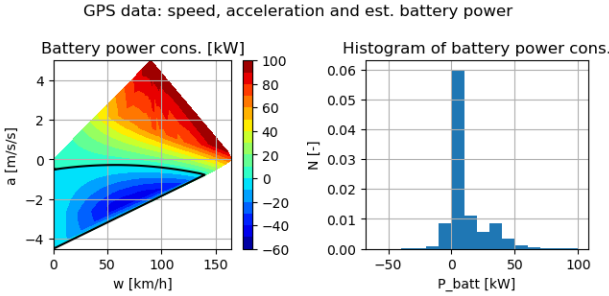


Figure 7 - Relationship between the battery power demand and acceleration for each recorded GPS data, coloured with the corresponding velocity of the vehicle and compared with power consumption along the line for reference vehicle

The speed ranges from 0 to approximately 160 km/h with acceleration ranging from -4 to +5 m/s². The estimated battery power demand ranges from -60 kW to +100 kW at high accelerations. Negative power indicates charging of the battery with regenerative braking.

The results of the initial energy demand assessment are presented in Figure 8. It can be observed that there are excessive variations in energy demand, especially in the second part of the year. This can be in part attributed to the method behind the energy demand determination. For instance, one of the vehicles used in this process was owned by the faculty and shared between the employees. This means that it regularly changed the operating patterns related to where each individual employee lived.

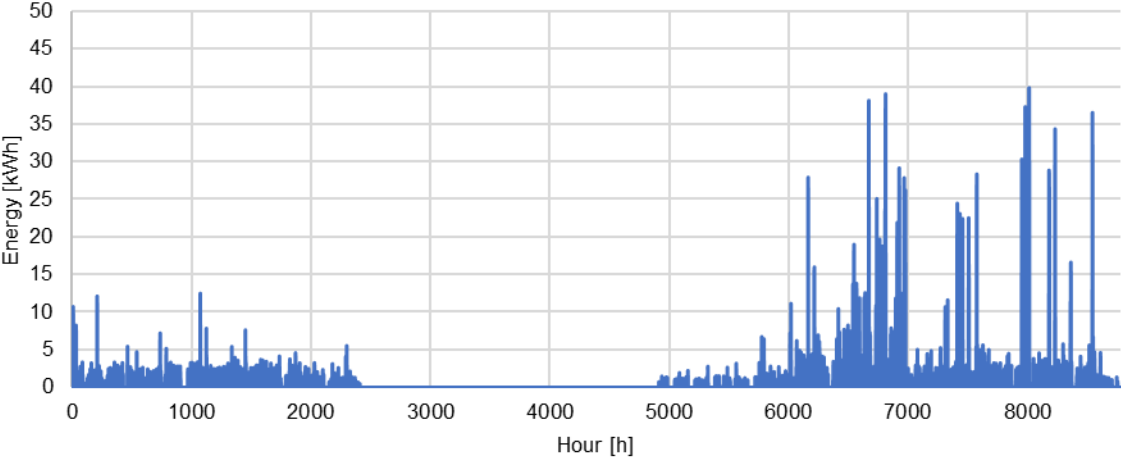


Figure 8. Original data on energy demand

4. Results

4.1 Initial results

The assumed distribution in this part of the work is gamma distribution having three parameters that are fitted. The resulting fitted distributions of battery power demand, grouped with histograms of estimated demand based on the recorded GPS data are presented in Figure 9 and Figure 10 for working days and weekend days, respectively.

Following figures present the hourly distribution of battery power demand, comparing the modelled gamma distribution with the GPS-estimated consumption. Each subplot corresponds to a different hour

of the day, highlighting variations in energy demand over a 24-hour period. The modelled distribution (blue line) is plotted alongside the GPS data histogram (orange bars), showing how well the statistical approach aligns with recorded vehicle activity.

A key observation from these figures is that during early morning hours (e.g., $t = 1\text{h}$ to $t = 6\text{h}$), energy demand is low, and data availability is sparse. This is not due to data collection issues but rather a reflection of actual vehicle usage patterns—fewer vehicles are active during these hours. Conversely, during peak driving periods ($t = 7\text{h}$ to $t = 12\text{h}$ and $t = 21\text{h}$ to $t = 22\text{h}$), the recorded GPS data is more complete, allowing for better alignment between the estimated and modelled distributions.

While the model captures general trends, some discrepancies between the modelled and GPS-based distributions remain, particularly during low-data periods. This indicates that further refinement of the distribution fitting process could improve the accuracy of demand estimations, especially when real-world usage patterns introduce significant variability.

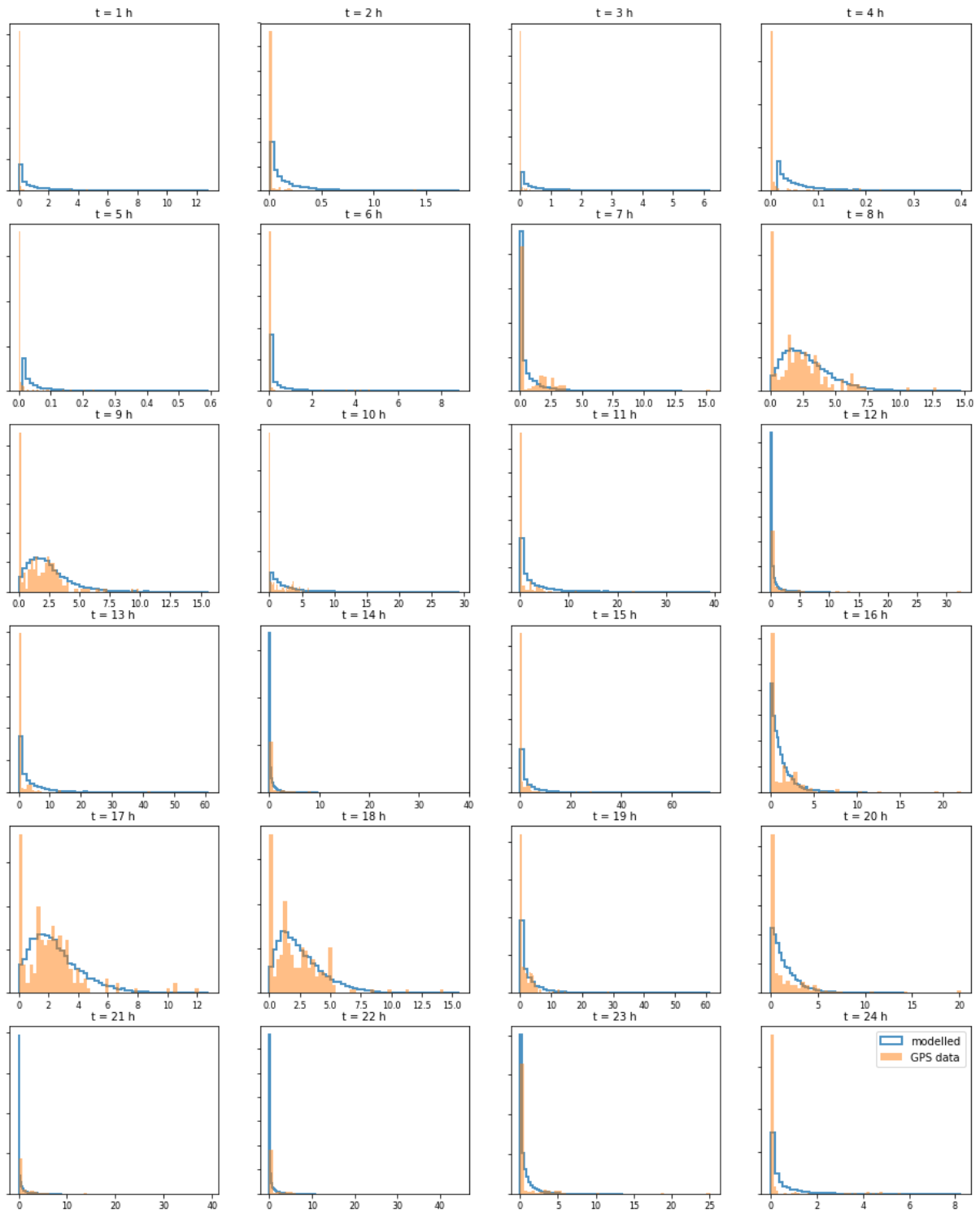


Figure 9 - Comparison between distributions of modelled power consumptions and one estimated from the recorded GPS data for work days with the use of the default normal-Weibull-gamma (nwg) approach

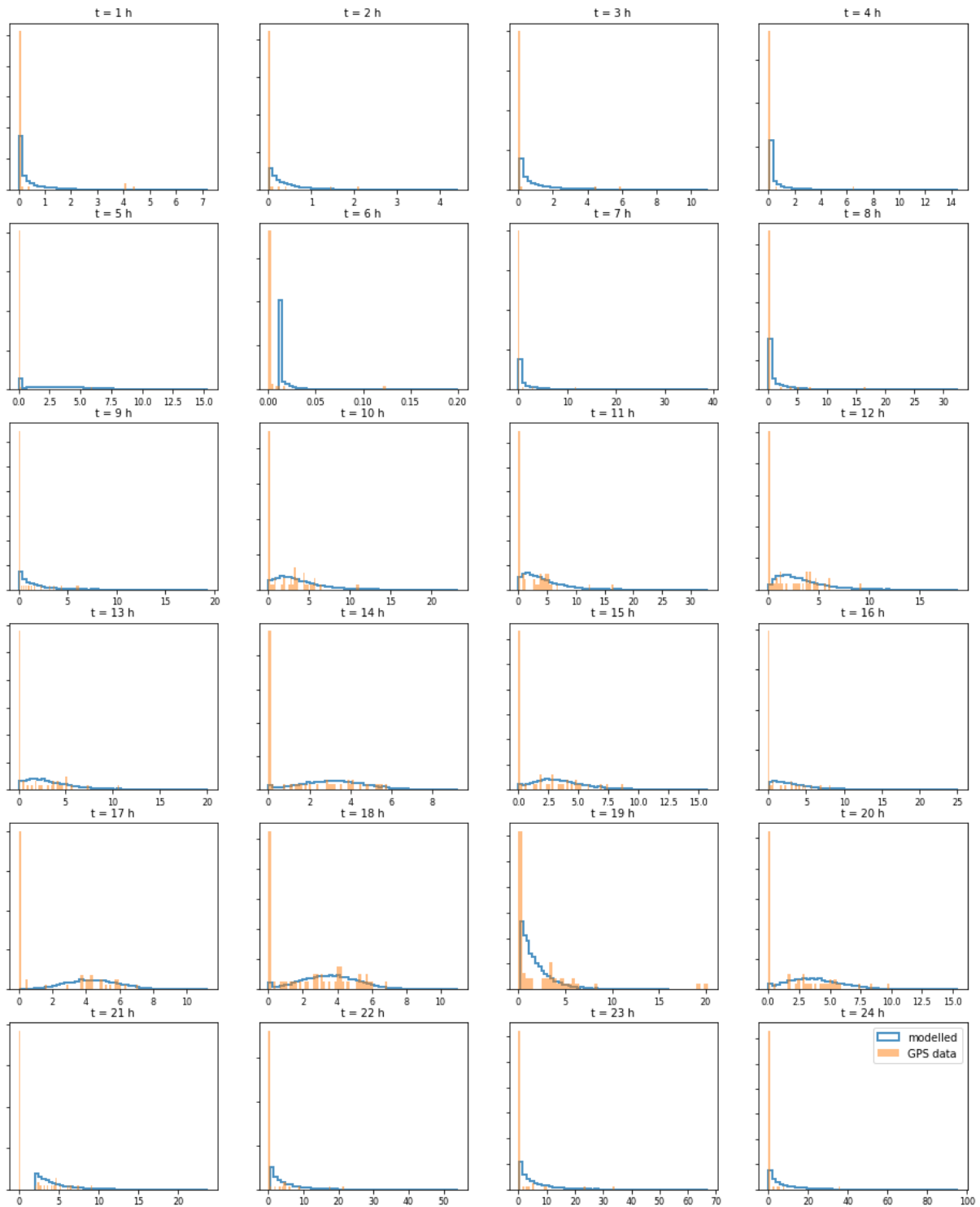


Figure 10 - Comparison between distributions of modelled power consumptions and one estimated from the recorded GPS data for the weekend with the use of the default normal-Weibull-gamma (nwg) approach

To further investigate the variation in energy demand, Figure 11 presents a time-series analysis of synthetic driving cycles over a one-week period. In this plot, the ratio of the energy demand in kWh units in relation to the time unit of hours [h] is displayed. The results reveal significant fluctuations in power demand, with peak consumption periods aligning with high vehicle activity hours.

Although missing GPS data reduces statistical precision in some cases, the overall trends in vehicle energy demand remain visible, showing that synthetic driving cycles can approximate real-world consumption patterns. Captured distributions during the experiment have gaps in the form of missing portion of the data when the vehicles were not used for tracking as well as the hours of low general transport demand as for example during the night. This makes the statistical analysis less accurate. However, during the hours for which more GPS data is available, more accurate fitting of the gamma function can be performed.

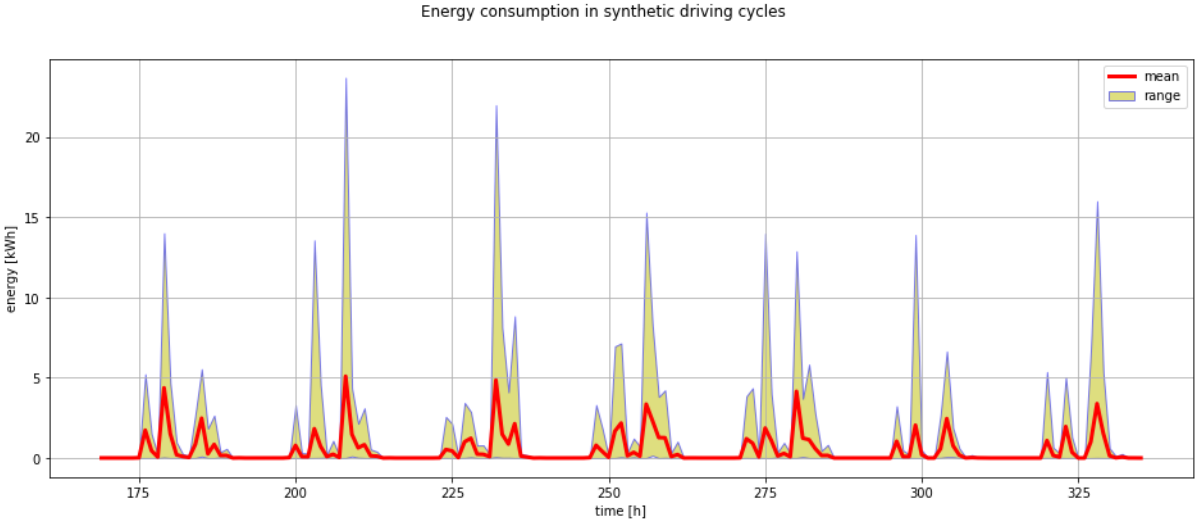


Figure 11 - Graphical representation of generated driving cycles, with median and range

A critical aspect of evaluating the accuracy of synthetic driving cycles is their correlation with real-world GPS-estimated energy consumption. Figure 12. presents a scatter plot comparing these values, with separate regression lines for workdays and weekends.

The results indicate a stronger correlation for weekend driving ($R^2 = 0.84$), meaning the synthetic cycles closely match real-world vehicle energy consumption on weekends. However, the correlation for workdays is significantly weaker ($R^2 = 0.26$), suggesting that weekday driving behaviour is more complex and less predictable using the current modelling approach. This discrepancy could be attributed to factors such as varied trip purposes, congestion patterns, and irregular travel schedules during the workweek. The reason for a worse match in the weekday is also probably due to an insufficient amount of data available that is required to correctly estimate and tune the statistical distribution used for the generation of synthetic driving cycles. GPS data are scarce, with significant amounts of gaps in statistical distribution, as can be seen in Figure 9 and Figure 10.

Despite these limitations, the results indicate that the synthetic cycles successfully replicate energy demand trends under more consistent driving conditions, such as those typically observed on weekends. Further refinement of the modelling approach, potentially by incorporating more detailed driving behaviour classifications, could improve weekday estimations.

Correlation between estimated and synthetic cycles

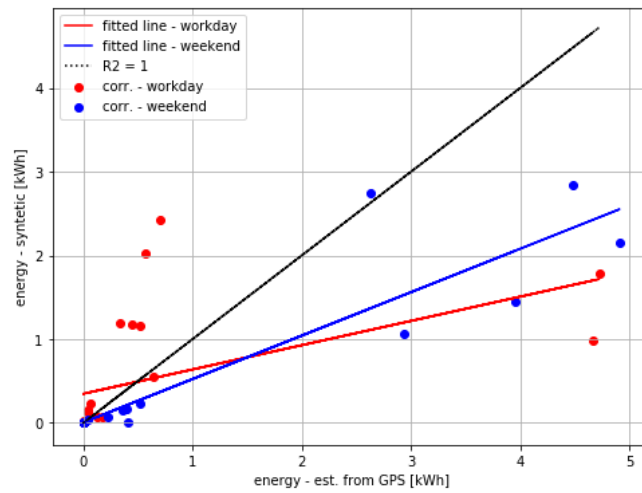


Figure 12 - Correlation between the battery energy consumption estimated from the GPS recorded data and generated by the synthetic driving cycles

Also, from the hourly results it can be seen that the resulting distribution significantly varies and has high peak loads, especially in the beginning of the year. In order to mitigate the influence of extreme values that are present in the input data, the implementation of multiple runs of the same system has been implemented. This in fact introduces variability into the statistical model which provides additional detail. In the model, this was represented by the number of vehicles for which the simulation was run. The cases presented in Figure 13, Figure 14 and Figure 15 display the normalized results in accordance with the maximum energy demand value that the model provided. Each plot is made for the different number of vehicles used that are actually the number of runs. It can be observed that the cases with a higher number of vehicles provided more smoother results without excessive values in single cases such as the case with 10 vehicles in Figure 13. This is expected as in this case that less of the vehicles are used, the probability of generating an outlier with excessive data is present.

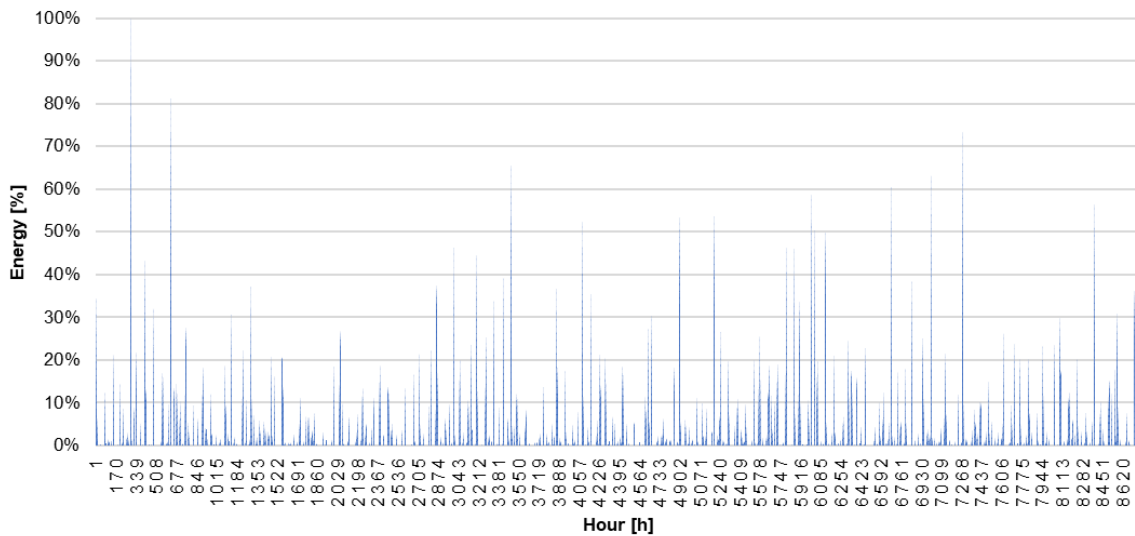


Figure 13. Results for 10 vehicles

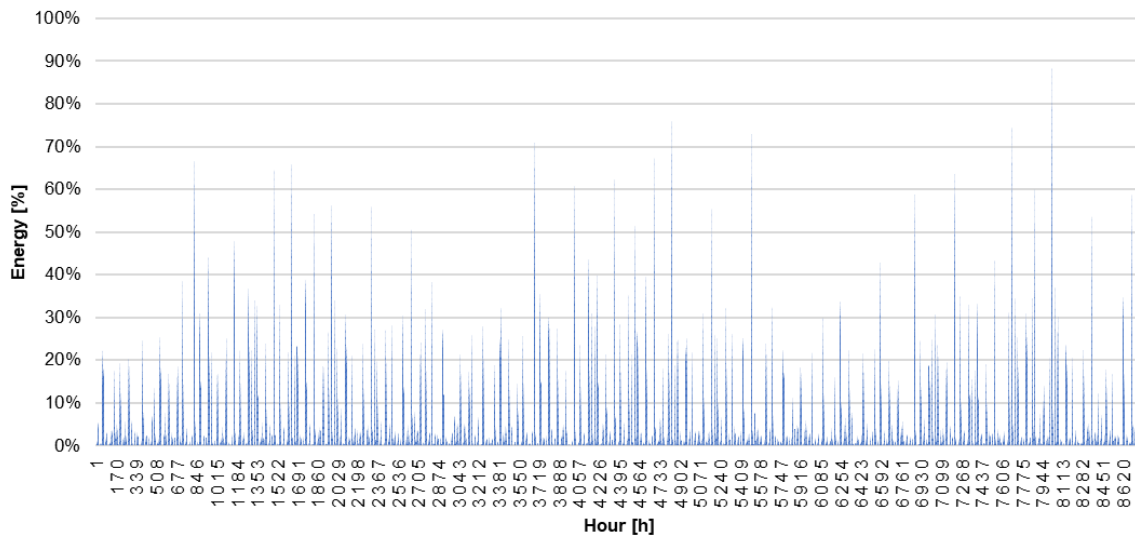


Figure 14. Results for 100 vehicles

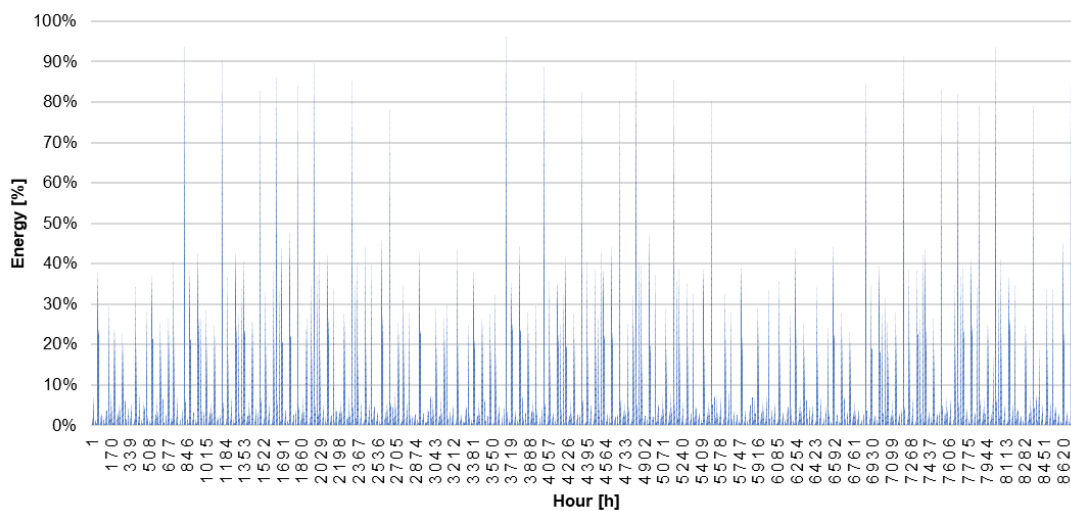


Figure 15. Results for 1000 vehicles

Still, the results even in the case with 1000 vehicles present significant unrealistic variations. It can be concluded that the excessive values originate from the same excessive values in the original data.

4.2. Modified approach with the exclusion of the outliers in input data

In an effort to mitigate excessive variations observed in energy demand simulations, an additional step was introduced to preprocess the original dataset and exclude extreme outlier values. The procedure involved identifying days where the maximum daily energy demand exceeded a predefined threshold and adjusting the dataset accordingly. Specifically, for any day where an outlier was detected, the entire day's energy demand was reset to zero to prevent isolated extreme values from distorting the results.

This approach resulted in a refined distribution, which is presented in Figure 16. Compared to the previous distribution (Figure 8), the modified dataset effectively removes extreme peaks, leading to a more representative energy demand profile that aligns better with expected variations.

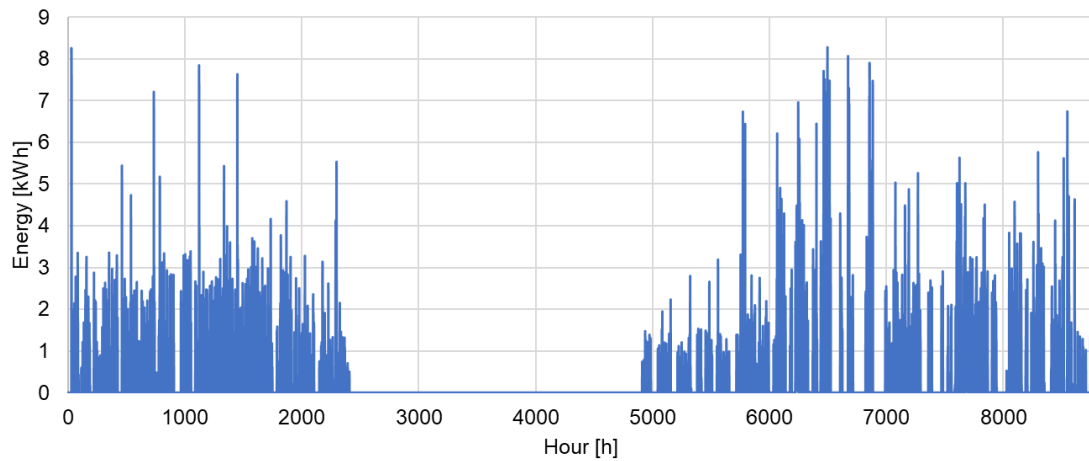


Figure 16. Modified energy demand distribution with the condition of $E_{\text{max_day}} \leq 10$ kWh

Following this modification, the simulated energy demand patterns exhibited reduced fluctuations in maximum values, as illustrated in Figure 17. The updated results now highlight a clearer weekly pattern, with distinguishable peak periods in energy demand that correspond to real-world vehicle usage trends.

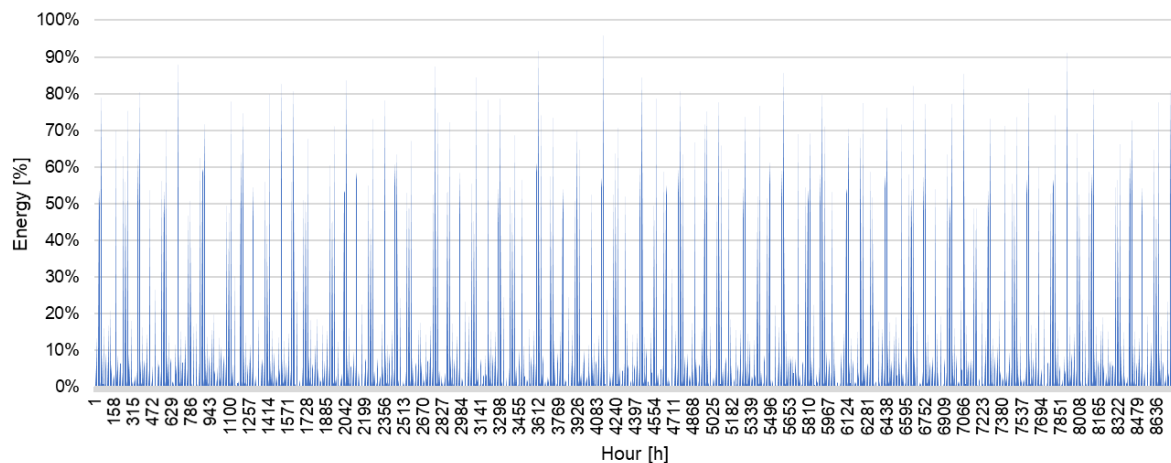


Figure 17. Results of energy demand using n-w-g distribution with the limit of 10 kWh

Alongside energy demand distribution, an important aspect of the analysis is the share of vehicles that remain plugged into charging stations at any given time. The results of this analysis, shown in Figure 18., indicate that the minimum share of plugged vehicles is 74.9%, while the average share is 95.2%. These values closely match real-world observations, confirming that the simulated plugging behaviour aligns with recorded data on parked vehicle availability.

A key assumption in the model is that plugged vehicles remain available for interaction with the power grid. This means that during these periods, the vehicles can participate in demand response strategies, such as managed charging or vehicle-to-grid (V2G) operations. The high average plugged share suggests that a large proportion of vehicles have the potential to contribute to grid services, reinforcing the importance of integrating such demand-side flexibility measures into energy planning.

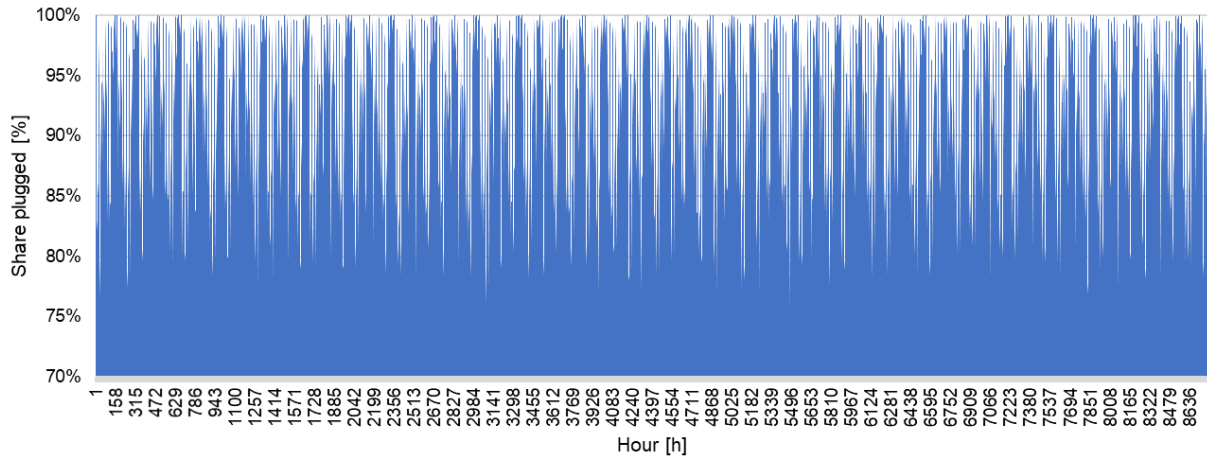


Figure 18. Results of plugged share of vehicles using n-w-g distribution with the limit of 10 kWh

4.3. Results for the different combinations of statistical distributions

To evaluate the influence of different statistical assumptions, simulations were conducted using various combinations of probability distributions for three key parameters:

- Trip start times
- Trip durations
- Energy demand per trip

The tested combinations included normal (n), Weibull (w), and gamma (g) distributions, leading to a total of 27 possible configurations. However, only 16 configurations are presented, as the remaining 11 resulted in extreme cases where vehicles were either always plugged (100%) or never plugged (0%), making them unsuitable for further analysis.

Table 1. summarizes the results, including the maximum, minimum, and average plugged vehicle share, as well as the maximum, minimum, and average energy demand.

- The plugged share of vehicles falls between 80% and 95%, consistent with real-world parking and charging behaviour.
- The average energy demand across all scenarios varies between 0.01 kWh and 0.32 kWh, indicating that some distributions lead to significantly higher or lower charging activity.

Certain distribution combinations, such as nnn and nng, result in higher variability in energy demand, suggesting that normal distributions may introduce more extreme values compared to gamma or Weibull alternatives. Additionally, Weibull and gamma-based configurations (e.g., wgg, gwg, ngg, ggg) tend to produce more stable energy demand distributions, with lower peak values and smoother variations over time. Simulations with higher minimum plugged share values (e.g., nwg, ngn, gwn) indicate cases where most vehicles are consistently connected, making them more suitable for V2G or managed charging applications.

Table 1. Results of the simulations

Distribution	Max plugged share [%]	Min plugged share [%]	Average plugged share [%]	Max energy [kWh]	Min energy [kWh]	Average energy [kWh]
wnn	1	0,01	0,87	1,45	0	0,07
wng	1	0	0,87	0,25	0	0,01
wgn	1	0,11	0,92	1,5	0	0,06
wgg	1	0,11	0,91	0,5	0	0,01
nwg	1	0,75	0,93	1,35	0	0,14
nwn	1	0,74	0,93	0,47	0	0,11
nnn	1	0,42	0,82	0,96	0	0,29
nng	1	0,41	0,82	2,5	0	0,32
ngn	1	0,66	0,9	0,59	0	0,15
ngg	1	0,65	0,9	1,68	0	0,19
gwn	1	0,69	0,94	0,75	0	0,16
gwg	1	0,7	0,94	1,6	0	0,15
gnn	1	0,35	0,82	1,13	0	0,27
gng	1	0,35	0,82	1,37	0	0,29
ggn	1	0,52	0,9	0,8	0	0,14
ggg	1	0,54	0,9	1,96	0	0,21

4.4. Results for the application of the neural networks

The energy demand distribution obtained using the neural network approach is presented in Figure 19. Compared to the original dataset, the results exhibit more regular patterns on daily and weekly level with no pronounced peaks. Due to the relatively low number of captured drives in the original data and the fact that the captured data does not feature all the possible weather conditions, the links to the weather data were not made as they would introduce unbiased bias.

To evaluate the accuracy of this approach, two error metrics were calculated:

- Mean Absolute Error (MAE): 2,3967
- Root Mean Squared Error (RMSE): 3,3881

These results suggest that the neural network approach provides a reasonably accurate estimation of energy demand, with error values remaining within an acceptable range.

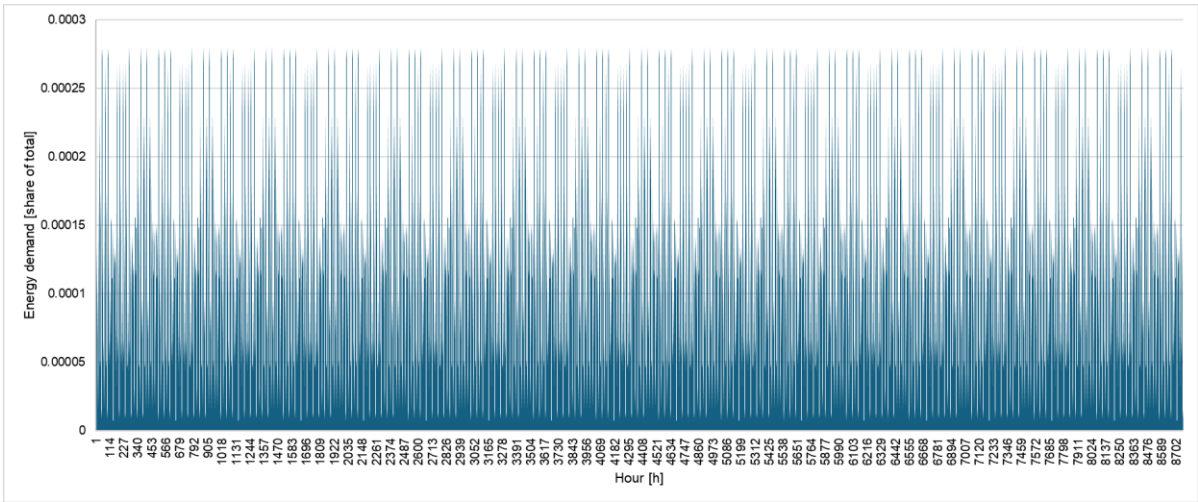


Figure 19. Results of energy demand with the application of neural networks

As shown in Figure 20. The results indicate that the average charging availability is 95%, with a minimum of 75% and a maximum of 100%. Since the impacts of the weather conditions were not examined here, the charging availability is the same in all the seasons. The differences in the availability are present on daily basis and weekly basis. This daily availability effect creates an opportunity for increased battery utilization during periods of reduced energy consumption, particularly in vehicle-to-grid (V2G) and smart charging applications. In practice, this relates to the vehicles being able to perform charging or discharging operations in the nighttime or between the morning and afternoon commutes with an assumption that the vehicles are able to connect to the grid during that time.

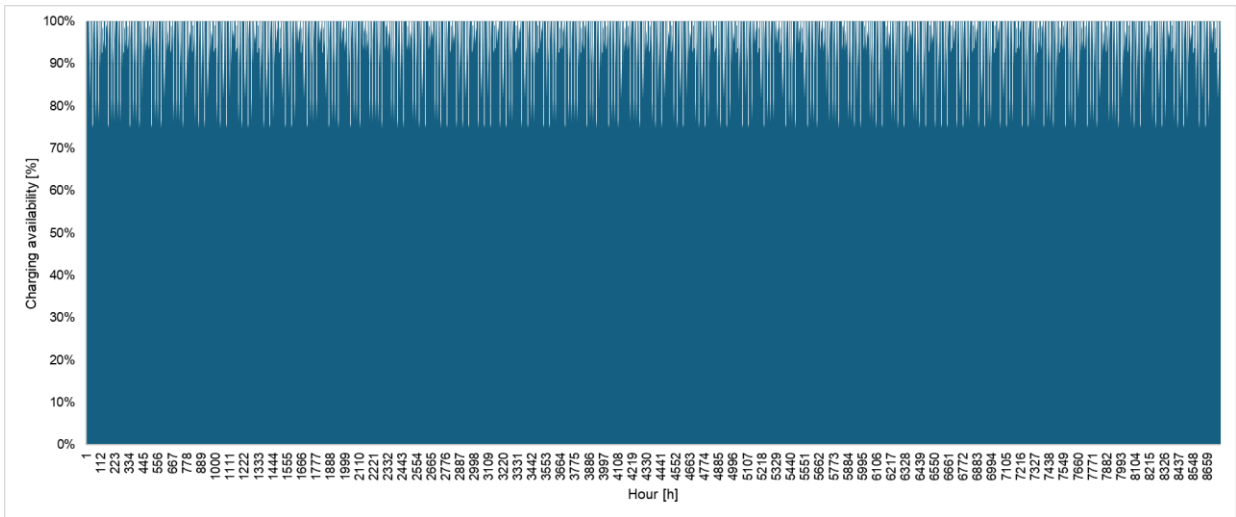


Figure 20. Availability for charging with the use of neural networks

Additionally, in Figure 21, the original data and the predicted data with the use of neural networks is presented. The results depict the weekly patterns that mostly coincide with the captured data as for example the times of morning and afternoon commutes. The captured data is highly focused on the small number of hours due to relatively small number of captured drives. The model managed to generalize this data and give the wider spread of the starting and ending times.

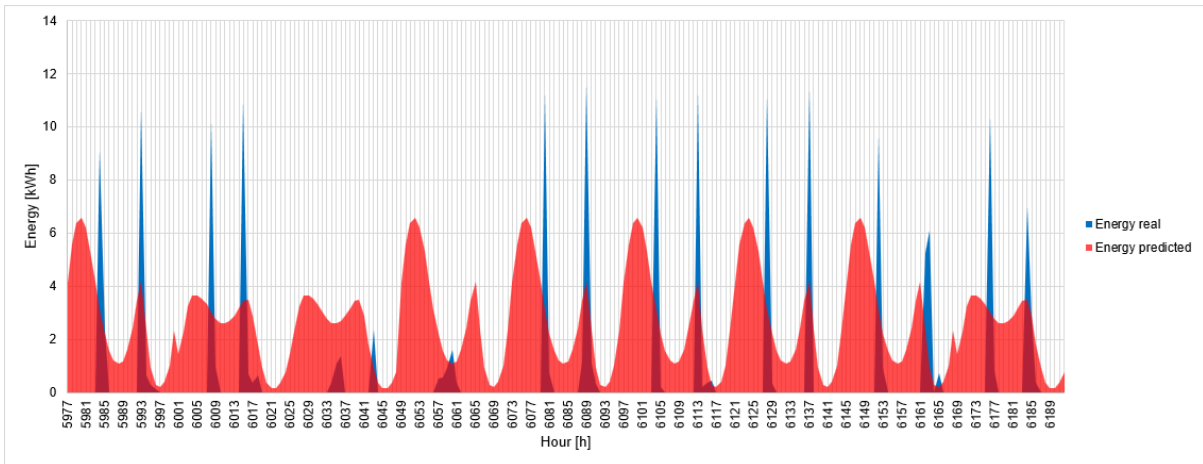


Figure 21. Comparison of the original data and the results of neural network

4.5. Testing of the obtained distributions in the H2RES model

The application of the generated energy demand distributions was tested in the H2RES energy model, with the results summarized in Table 2. The objective of this test was to evaluate whether the generated distributions could be successfully integrated into the model and whether the system could provide a feasible energy balance. The results indicate that not all distributions were applicable within the H2RES framework, as some cases resulted in infeasible outputs. This infeasibility was primarily caused by:

- Inadequate charging availability, meaning that the model could not ensure a sufficient number of plugged-in vehicles to balance the energy demand.
- High variability in energy demand, where large fluctuations made it impossible for the available generation capacity to meet the modelled energy consumption.

Since the primary focus of this study was not the H2RES model itself, the testing was limited to examining the applicability of the generated distributions within the model. However, a general trend can be observed:

- Some distributions resulted in infeasibility across multiple cases, indicating that they do not align well with the system's constraints.
- The distributions that led to successful model runs displayed similar patterns in all cases, regardless of the specific probability distributions used.
- Neural network-based demand distributions were successfully implemented, demonstrating their potential compatibility with energy system models.

The detailed results of this testing process are displayed in the appendix, while a summary of the findings is provided in Table 3.

Table 2. Results of the application of the generated distributions in the energy model

Case	Start	Duration	Energy	Cycles successfully calculated	H2RES model successfully run

1	Normal	Normal	Normal	Yes	infeasible
2	Normal	Normal	Weibull	No	-
3	Normal	Normal	Gamma	Yes	Yes
4	Normal	Weibull	Normal	Yes	infeasible
5	Normal	Weibull	Weibull	No	-
6	Normal	Weibull	Gamma	Yes	Yes
7	Normal	Gamma	Normal	Yes	infeasible
8	Normal	Gamma	Weibull	No	-
9	Normal	Gamma	Gamma	Yes	Yes
10	Weibull	Normal	Normal	Yes	infeasible
11	Weibull	Normal	Weibull	No	-
12	Weibull	Normal	Gamma	Yes	infeasible
13	Weibull	Weibull	Normal	No	-
14	Weibull	Weibull	Weibull	No	-
15	Weibull	Weibull	Gamma	No	-
16	Weibull	Gamma	Normal	Yes	infeasible
17	Weibull	Gamma	Weibull	No	-
18	Weibull	Gamma	Gamma	Yes	infeasible
19	Gamma	Normal	Normal	Yes	infeasible
20	Gamma	Normal	Weibull	No	-
21	Gamma	Normal	Gamma	Yes	Yes
22	Gamma	Weibull	Normal	Yes	infeasible
23	Gamma	Weibull	Weibull	No	-
24	Gamma	Weibull	Gamma	Yes	Yes
25	Gamma	Gamma	Normal	Yes	infeasible
26	Gamma	Gamma	Weibull	No	-
27	Gamma	Gamma	Gamma	Yes	Yes
28	Neural network			Yes	Yes

5. Conclusion

The method presented in this paper can be used for the estimation of driving cycles with qualitatively reasonable accuracy, but quantitatively it is still not possible to determine the accuracy of the methodology due to scarce inputs from the GPS tracker. Also, the data was obtained only for the portion of the year. As can be seen from Table 1, the results are highly dependent on the type of statistical approach that is used. Further analysis is required to assess if there is a relation between the fitting of the resulting distribution and the type of curve that is used as input to the model. Additionally, when comparing the conducted approaches, similar results are obtained both in the approach with statistical method and in the neural network approach. Additional details such as the impact of the weather conditions could not be reliably replicated since not all the possible circumstances are captured in the training data. Using the neural network approach, the error MAE was brought to the value of 2,39 and RMSE to 3,38. The assumption is that, as more data will be gathered from the GPS trackers in the future, the quantitative accuracy of the methodology can be improved. The accuracy itself can be measured with the use of R-squared values representing the fitting quality between measurements and modelled driving cycles. Moreover, increasing the number of recorded data should remove the gaps in the statistical distribution of the input data leading to more accurate determination of distribution fitting parameters. Also, the method of capturing the data may be changed in a way that the personal vehicles are equipped with the trackers instead of equipping the faculty vehicle which follows very different driving cycles. A possible improvement for future work is the decoupling of the input dataset and corresponding modelling of driving cycles into more than just two groups (workdays and weekends) in order to capture seasonal variations throughout the year. Further work will also focus on improving the neural networks model with examination of different values in the number of layers and neurons. The final goal is higher quality data provision for the energy planning and optimization models. This will reduce the number of hard to come by input parameters to these types of tools which will in turn enable easier energy planning in tools such as EnergyPLAN and H2RES. Finally, when the entire spectrum of the possible conditions is captured, the involvement of the meteorological conditions into the neural network will be possible without creating the bias.

Acknowledgement

Financial support from the RESFLEX project funded by the Programme of the Government of Republic of Croatia, Croatian Environmental Protection and Energy Efficiency Fund with the support of the Croatian Science Foundation for encouraging research and development activities in the area of Climate Change for the period from 2017 to 2019 is gratefully acknowledged.

This work has received support and funding from the Croatian Science Foundation through the project IP-2019-04-9482 INTERENERGY.

Nomenclature

P	Power [W, kW]
e	energy [kWh]
dt	time step [s]
w	speed
a	coefficient for battery power consumption
t	reference time step data point

REFERENCES

- [1] University of Zagreb, “H2RES model.” Accessed: May 11, 2024. [Online]. Available: <https://h2res.org/>, Croatia
- [2] European Council, “Clean and sustainable mobility,” <https://www.consilium.europa.eu/en/policies/clean-and-sustainable-mobility/>.
- [3] J. Michalski, M. Poltrum, and U. Bünger, “The role of renewable fuel supply in the transport sector in a future decarbonized energy system,” *Int J Hydrogen Energy*, vol. 44, Dec. 2018, doi: 10.1016/j.ijhydene.2018.10.110.
- [4] T. Unterluggauer, J. Rich, P. B. Andersen, and S. Hashemi, “Electric vehicle charging infrastructure planning for integrated transportation and power distribution networks: A review,” *eTransportation*, vol. 12, p. 100163, 2022, doi: <https://doi.org/10.1016/j.etrans.2022.100163>.
- [5] R. Fachrizal, M. Shepero, D. van der Meer, J. Munkhammar, and J. Widén, “Smart charging of electric vehicles considering photovoltaic power production and electricity consumption: A review,” *eTransportation*, vol. 4, p. 100056, 2020, doi: <https://doi.org/10.1016/j.etrans.2020.100056>.
- [6] A. Pfeifer, L. Herc, I. Batas Bjelić, and N. Duić, “Flexibility index and decreasing the costs in energy systems with high share of renewable energy,” *Energy Convers Manag*, vol. 240, p. 114258, 2021, doi: <https://doi.org/10.1016/j.enconman.2021.114258>.
- [7] K. Darcovich, S. Recoskie, H. Ribberink, and C. Michelet, “The impact of V2X service under local climatic conditions within Canada on EV durability,” *eTransportation*, vol. 9, p. 100124, 2021, doi: <https://doi.org/10.1016/j.etrans.2021.100124>.
- [8] J. Dixon, W. Bukhsh, K. Bell, and C. Brand, “Vehicle to grid: driver plug-in patterns, their impact on the cost and carbon of charging, and implications for system flexibility,” *eTransportation*, vol. 13, p. 100180, 2022, doi: <https://doi.org/10.1016/j.etrans.2022.100180>.
- [9] H. Bajolle, M. Lagadic, and N. Louvet, “The future of lithium-ion batteries: Exploring expert conceptions, market trends, and price scenarios,” *Energy Res Soc Sci*, vol. 93, p. 102850, 2022, doi: <https://doi.org/10.1016/j.erss.2022.102850>.
- [10] J. Schwab, C. Sölch, and G. Zöttl, “Electric Vehicle Cost in 2035: The impact of market penetration and charging strategies,” *Energy Econ*, vol. 114, p. 106263, 2022, doi: <https://doi.org/10.1016/j.eneco.2022.106263>.
- [11] B. Li, J. Li, and B. Jian, “Improve multi-energy supply microgrid resilience using mobile hydrogen trucks based on transportation network,” *eTransportation*, vol. 18, p. 100265, 2023, doi: <https://doi.org/10.1016/j.etrans.2023.100265>.
- [12] O. Ruhnau, “How flexible electricity demand stabilizes wind and solar market values: The case of hydrogen electrolyzers,” *Appl Energy*, vol. 307, p. 118194, 2022, doi: <https://doi.org/10.1016/j.apenergy.2021.118194>.
- [13] B. Li, J. Li, and B. Jian, “Improve multi-energy supply microgrid resilience using mobile hydrogen trucks based on transportation network,” *eTransportation*, vol. 18, p. 100265, 2023, doi: <https://doi.org/10.1016/j.etrans.2023.100265>.

- [14] A. Pfeifer, P. Prebeg, and N. Duić, “Challenges and opportunities of zero emission shipping in smart islands: A study of zero emission ferry lines,” *eTransportation*, vol. 3, p. 100048, 2020, doi: <https://doi.org/10.1016/j.etrans.2020.100048>.
- [15] D. L. Greene, J. M. Ogden, and Z. Lin, “Challenges in the designing, planning and deployment of hydrogen refueling infrastructure for fuel cell electric vehicles,” *eTransportation*, vol. 6, p. 100086, 2020, doi: <https://doi.org/10.1016/j.etrans.2020.100086>.
- [16] Aalborg University, “EnergyPLAN,” <https://www.energyplan.eu/>.
- [17] Energy Exemplar, “PLEXOS,” <https://www.energyexemplar.com/plexos>.
- [18] T. Pukšec, G. Krajačić, Z. Lulić, B. V. Mathiesen, and N. Duić, “Forecasting long-term energy demand of Croatian transport sector,” *Energy*, vol. 57, pp. 169–176, 2013, doi: <https://doi.org/10.1016/j.energy.2013.04.071>.
- [19] M. G. Prina, M. Lionetti, G. Manzolini, W. Sparber, and D. Moser, “Transition pathways optimization methodology through EnergyPLAN software for long-term energy planning,” *Appl Energy*, vol. 235, pp. 356–368, 2019, doi: <https://doi.org/10.1016/j.apenergy.2018.10.099>.
- [20] Z. Chen and R. Xiong, “Driving cycle development for electric vehicle application using principal component analysis and k-means cluster: with the case of Shenyang, China,” *Energy Procedia*, vol. 142, pp. 2264–2269, Dec. 2017, doi: [10.1016/j.egypro.2017.12.628](https://doi.org/10.1016/j.egypro.2017.12.628).
- [21] S. Zhang, M. Chen, and W. Zhang, “A novel location-routing problem in electric vehicle transportation with stochastic demands,” *J Clean Prod*, vol. 221, pp. 567–581, 2019, doi: <https://doi.org/10.1016/j.jclepro.2019.02.167>.
- [22] T. Novosel et al., “Agent based modelling and energy planning – Utilization of MATSim for transport energy demand modelling,” *Energy*, vol. 92, pp. 466–475, 2015, doi: <https://doi.org/10.1016/j.energy.2015.05.091>.
- [23] H. Lin, K. Fu, Y. Liu, Q. Sun, and R. Wennersten, “Modeling charging demand of electric vehicles in multi-locations using agent-based method,” *Energy Procedia*, vol. 152, pp. 599–605, Oct. 2018, doi: [10.1016/j.egypro.2018.09.217](https://doi.org/10.1016/j.egypro.2018.09.217).
- [24] M. B. Arias and S. Bae, “Electric vehicle charging demand forecasting model based on big data technologies,” *Appl Energy*, vol. 183, pp. 327–339, 2016, doi: <https://doi.org/10.1016/j.apenergy.2016.08.080>.
- [25] M. Gallet, T. Massier, and T. Hamacher, “Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks,” *Appl Energy*, vol. 230, pp. 344–356, Nov. 2018, doi: [10.1016/j.apenergy.2018.08.086](https://doi.org/10.1016/j.apenergy.2018.08.086).
- [26] J. Su, T. T. Lie, and R. Zamora, “Modelling of large-scale electric vehicles charging demand: A New Zealand case study,” *Electric Power Systems Research*, vol. 167, pp. 171–182, Feb. 2019, doi: [10.1016/j.epsr.2018.10.030](https://doi.org/10.1016/j.epsr.2018.10.030).
- [27] R. Galvin, “Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers,” *Transp Res D Transp Environ*, vol. 53, pp. 234–248, Jun. 2017, doi: [10.1016/j.trd.2017.04.020](https://doi.org/10.1016/j.trd.2017.04.020).

- [28] J. Asamer, A. Graser, B. Heilmann, and M. Ruthmair, "Sensitivity analysis for energy demand estimation of electric vehicles," *Transp Res D Transp Environ*, vol. 46, pp. 182–199, 2016, doi: <https://doi.org/10.1016/j.trd.2016.03.017>.
- [29] MIT, "unparking," <https://senseable.mit.edu/unparking/>.
- [30] L. Herc, A. Pfeifer, F. Feijoo, and N. Duić, "Energy system transitions pathways with the new H2RES model: A comparison with existing planning tool," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 1, 2021, doi: 10.1016/j.prime.2021.100024.
- [31] L. Herc et al., "The management of an energy system in the realm of rapid energy transition and degasification as a consequence of energy crisis, examination in H2RES energy model," *Energy Convers Manag*, vol. 315, p. 118782, Sep. 2024, doi: 10.1016/J.ENCONMAN.2024.118782.
- [32] A. Kıyaklı and H. Solmaz, "Modeling of an Electric Vehicle with MATLAB/Simulink," *International Journal of Automotive Science And Technology*, vol. 2, pp. 9–15, Jan. 2019, doi: 10.30939/ijastech..475477.
- [33] European Comission, "Joint Research Reports - TANK-TO-WHEELS Report - Version 5: Passenger Cars."

APPENDIX

```
# =====  
# 1. Import libraries  
# =====  
import pandas as pd  
import numpy as np  
from sklearn.model_selection import train_test_split  
from sklearn.preprocessing import MinMaxScaler  
from sklearn.metrics import mean_squared_error  
from keras.models import Sequential  
from keras.layers import Dense, Dropout  
from keras.callbacks import EarlyStopping  
from keras.losses import Huber  
from google.colab import drive  
drive.mount('/content/gdrive')  
  
# =====  
# 2. Load and preprocess the dataset  
# =====  
# Load dataset  
file_path = '/content/gdrive/MyDrive/NN_data_new1.csv'  
df = pd.read_csv(file_path)  
  
# Time features  
df['sin_hour'] = np.sin(2 * np.pi * df['sat'] / 24)  
df['cos_hour'] = np.cos(2 * np.pi * df['sat'] / 24)  
  
# Weekend indicator  
df['is_weekend'] = (df['sub'] + df['ned']).clip(0, 1)  
  
# Filter nonzero samples  
df_nonzero = df[df['energy real [kWh]'] > 0.01]  
print(f'Training samples after filtering: {len(df_nonzero)}")  
  
# Choose features (temperature removed)  
selected_columns = ['sin_hour', 'cos_hour', 'is_weekend']  
X = df_nonzero[selected_columns].values  
Y = df_nonzero['energy real [kWh]'].values  
  
# =====  
# 3. Train/validation/test split
```

```

# =====
X_train, X_val_and_test, Y_train, Y_val_and_test = train_test_split(
    X, Y, test_size=0.3, random_state=42)
X_val, X_test, Y_val, Y_test = train_test_split(
    X_val_and_test, Y_val_and_test, test_size=0.5, random_state=42)

# =====
# 4. Scale and reweight features
# =====
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_val_scaled = scaler.transform(X_val)
X_test_scaled = scaler.transform(X_test)

# Reweight hour influence
X_train_scaled[:, 0:2] *= 1.2
X_val_scaled[:, 0:2] *= 1.2
X_test_scaled[:, 0:2] *= 1.2

# =====
# 5. Build the model
# =====
model = Sequential([
    Dense(128, activation='relu', input_shape=(3,)), # 3 features
    Dropout(0.2),
    Dense(64, activation='relu'),
    Dropout(0.2),
    Dense(32, activation='relu'),
    Dense(1, activation='relu') # ReLU ensures non-negative output
])

model.compile(optimizer='adam',
              loss=Huber(delta=5.0),
              metrics=['mae'])

# =====
# 6. Train the model
# =====
early_stopping = EarlyStopping(monitor='val_loss', patience=20, restore_best_weights=True)

history = model.fit(
    X_train_scaled, Y_train,
    batch_size=32,
    epochs=1000,

```

```

validation_data=(X_val_scaled, Y_val),
callbacks=[early_stopping],
verbose=1
)

# =====
# 7. Evaluate the model
# =====
loss, mae = model.evaluate(X_test_scaled, Y_test, verbose=0)
print(f"Mean Absolute Error on Test Data: {mae:.4f}")

# Get predictions on test set to calculate RMSE
Y_test_pred = model.predict(X_test_scaled).flatten()
rmse = np.sqrt(mean_squared_error(Y_test, Y_test_pred))
print(f"Root Mean Squared Error (RMSE) on Test Data: {rmse:.4f}")

# =====
# 8. Load new input for 2018 and predict
# =====
new_data_file_path = '/content/gdrive/MyDrive/drives_input_2018_1.csv'
new_data_df = pd.read_csv(new_data_file_path)

# Add same features
new_data_df['sin_hour'] = np.sin(2 * np.pi * new_data_df['sat'] / 24)
new_data_df['cos_hour'] = np.cos(2 * np.pi * new_data_df['sat'] / 24)
new_data_df['is_weekend'] = (new_data_df['sub'] + new_data_df['ned']).clip(0, 1)

# Extract same features
new_data_features = new_data_df[selected_columns].values
new_data_features_scaled = scaler.transform(new_data_features)

# Reweight hour influence
new_data_features_scaled[:, 0:2] *= 1.2

# Predict
predictions = model.predict(new_data_features_scaled).flatten()
predictions_fixed = np.where(predictions < 0, 0, predictions)

# Save results
predictions_df = pd.DataFrame({
    'Sample': range(1, len(predictions_fixed) + 1),
    'Predicted Energy [kWh]': predictions_fixed
})

```

```

predictions_df.to_csv('/content/gdrive/MyDrive/predicted_energy_results_fixed_1.csv',
index=False)
print("Predictions saved to Google Drive!")

```

Neural network description

Table 3. Example of the data for neural network

Day	Hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Holiday	Precipitation [mm/hour]	Irradiation surface [W/m ²]	Energy [kWh]
253	0	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	1	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	2	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	3	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	4	0	1	0	0	0	0	0	0	0.000	8.811	0.0000
253	5	0	1	0	0	0	0	0	0	0.000	121.471	0.0000
253	6	0	1	0	0	0	0	0	0	0.000	296.813	0.0000
253	7	0	1	0	0	0	0	0	0	0.000	470.253	0.0000
253	8	0	1	0	0	0	0	0	0	0.000	614.639	0.5072
253	9	0	1	0	0	0	0	0	0	0.000	707.300	0.5480
253	10	0	1	0	0	0	0	0	0	0.000	751.578	0.9157
253	11	0	1	0	0	0	0	0	0	0.001	734.400	1.6150
253	12	0	1	0	0	0	0	0	0	0.001	673.826	0.3445
253	13	0	1	0	0	0	0	0	0	0.001	569.756	0.0000
253	14	0	1	0	0	0	0	0	0	0.001	421.427	0.0000
253	15	0	1	0	0	0	0	0	0	0.000	255.802	0.0000
253	16	0	1	0	0	0	0	0	0	0.000	89.025	0.0000
253	17	0	1	0	0	0	0	0	0	0.000	2.406	0.0000
253	18	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	19	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	20	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	21	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	22	0	1	0	0	0	0	0	0	0.000	0.000	0.0000
253	23	0	1	0	0	0	0	0	0	0.000	0.000	0.0000

The process in the creation of the neural network and obtaining the results data is described in this next section.

1. Mount Google Drive:

The script begins by mounting Google Drive in the Google Colab environment using the google.colab.drive package. This provides access to stored datasets and enables saving model outputs directly to Google Drive.

2. Import Packages:

Several key Python libraries are imported for later use:

pandas for data handling and manipulation.

NumPy for numerical operations, especially array handling.

train_test_split from sklearn.model_selection to split the dataset into training, validation, and test sets.

MinMaxScaler from sklearn.preprocessing to normalize feature values for neural network training.

mean_squared_error from sklearn.metrics to compute the Root Mean Squared Error (RMSE).

Sequential from keras.models to define a sequential neural network.

Dense and Dropout from keras.layers to construct the neural network architecture.

EarlyStopping from keras.callbacks to halt training early if validation loss stops improving.

Huber from keras.losses as the selected loss function for robust regression.

3. Load Data:

A CSV file containing energy consumption data (NN_data_new1.csv) is loaded using Pandas. This dataset includes hourly and weekend indicators, along with the target energy consumption values.

4. Feature Engineering:

New time-based features are calculated:

sin_hour and cos_hour to encode the hour of the day in a cyclic manner (so hour 23 and 0 are close).

is_weekend as a binary indicator marking weekends.

These features are added to the dataset to improve temporal learning by the model.

5. Filter Nonzero Samples:

The dataset is filtered to exclude entries where energy consumption is very low or zero (below 0.01 kWh), ensuring the model focuses on meaningful consumption patterns.

6. Select Input Features and Targets:

The input features (X) include sin_hour, cos_hour, and is_weekend. The target (Y) is the actual measured energy consumption.

7. Split Data into Training, Validation, and Test Sets:

The dataset is split in two steps:

First, 70% training data, 30% combined validation + test.

Then, the 30% is evenly split into validation and test sets.

This ensures robust model evaluation and avoids testing on data the model has already seen.

8. Scale Features:

A MinMaxScaler is applied to the input features to scale them to the range [0, 1], improving

the neural network's convergence. Additionally, the hour-based features are manually reweighted by multiplying by 1.2, increasing their relative influence during training.

9. Build Neural Network Model:

A sequential neural network is defined with:

An input layer (3 input features).

Three hidden layers with 128, 64, and 32 neurons, respectively, all using ReLU activation and dropout regularization.

An output layer with 1 neuron and ReLU activation to predict non-negative energy consumption.

10. Compile the Model:

The model is compiled with:

adam optimizer for adaptive learning.

Huber loss (robust against outliers).

Mean Absolute Error (MAE) as the evaluation metric.

11. Set Early Stopping:

An early stopping mechanism is set up, monitoring validation loss. If no improvement is seen over 20 consecutive epochs, training halts, and the best weights are restored.

12. Train the Model:

The model is trained using the training set, validated against the validation set, over up to 1000 epochs, using a batch size of 32. Training stops early if needed.

13. Evaluate the Model on Test Data:

After training, the model is evaluated on the held-out test set to compute the MAE, giving a first estimate of predictive performance.

14. Calculate RMSE on Test Data:

In addition to MAE, predictions on the test set are compared to true values using Root Mean Squared Error (RMSE), providing a complementary error metric that penalizes larger errors more heavily.

15. Load New Input Data:

Another CSV file (`drives_input_2018_1.csv`) containing new input data is loaded. This is the dataset the trained model will be used to predict on.

16. Feature Engineering for New Data:

The same transformations (`sin_hour`, `cos_hour`, and `is_weekend`) are applied to the new dataset to match the structure used during training.

17. Scale New Input Data:

The new data is scaled using the same `MinMaxScaler` instance that was fit on the original training data to ensure consistency.

18. Make Predictions on New Data:

The trained model generates predictions for the new inputs. Negative predictions, if any, are corrected to zero.

19. Save Predictions to CSV:

The predictions are saved as a new CSV file (`predicted_energy_results_fixed_1.csv`) in

Google Drive. The results are further analyzed and compared to the input data as well as to generate the charging availability curves.

Results for the application of the H2RES model

The results for the evolution of the transport sector composition are displayed in the Table 4. It can be observed that all the performed cases display similar results with the share of electrification reaching between 78, and 78,2 % in 2050, FCEV between 21,8 and 22 while the share of ICE is 0 in all cases in 2050 since one of the goals was to reach 0 Mt CO2 emissions.

Table 4. Results for the evolution of the transport sector. The nomenclature for the statistical distribution is in line with the nomenclature used in Table 2

Year/used approach or statistical distribution	Electric vehicles						
	GGG	GNG	GWG	NGG	NNG	NWG	Neural network
2020	0,011	0,011	0,011	0,011	0,011	0,01	0,01
2025	0,032	0,032	0,032	0,032	0,032	0,032	0,031
2030	0,182	0,182	0,182	0,182	0,182	0,182	0,181
2035	0,332	0,332	0,332	0,332	0,332	0,332	0,331
2040	0,482	0,482	0,482	0,482	0,482	0,482	0,481
2045	0,632	0,631	0,632	0,632	0,631	0,631	0,631
2050	0,782	0,781	0,782	0,781	0,781	0,781	0,781
Year	Hydrogen fuel cell vehicles						
2020	0	0	0	0	0	0	0
2025	0	0	0	0	0	0	0
2030	0,047	0,055	0,048	0,057	0,046	0,048	0,046
2035	0,096	0,105	0,098	0,107	0,096	0,098	0,096
2040	0,146	0,154	0,147	0,156	0,145	0,147	0,145
2045	0,146	0,154	0,147	0,156	0,145	0,147	0,145
2050	0,218	0,22	0,219	0,219	0,219	0,22	0,219
Year	Internal combustion engine vehicles						
2020	0,989	0,989	0,99	0,989	0,989	0,99	0,99
2025	0,968	0,968	0,968	0,968	0,968	0,968	0,968
2030	0,771	0,763	0,77	0,761	0,772	0,77	0,773
2035	0,572	0,564	0,571	0,562	0,572	0,571	0,573
2040	0,372	0,364	0,371	0,362	0,373	0,371	0,374
2045	0,222	0,216	0,221	0,212	0,223	0,222	0,224
2050	0	0	0	0	0	0	0