Dynamic Capture and Estimation of Energy-Emission Benefits for Electric Taxicabs with Spatiotemporal Big Data and Deep-Learning-Based Microscopic Model

Xinyi LI\textsuperscript{ab}, Biyu CHEN\textsuperscript{ab}, Qi LIU\textsuperscript{ab} and Jia TAO\textsuperscript{bc}

\textsuperscript{a} State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei, China
\textsuperscript{b} Geocomputation Center for Social Sciences, Wuhan University, Wuhan 430079, Hubei, China
\textsuperscript{c} School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, Hubei, China

Abstract. Reducing carbon emissions to cope with climate change and short of energy have become a global trend, so it is urgent to accurately measure energy consumption and emissions. As taxi occupies the second highest proportion of domestic roads, it is necessary to study the emission of new energy. However, existing studies often consider the whole region, with low accuracy and no true value, resulting in difficult verification of conclusions. This paper proposed a micro-energy consumption and carbon emission model for taxicabs based on trajectory data and deep learning method to dynamically simulate the real-time energy consumption of taxicabs with different energy sources. First, this paper detected the correlation between driving state and energy consumption carbon emission in portable emissions measurement system (PEMS) environment. Then, a deep learning-based framework was built to learn the vehicle’s energy consumption carbon emission pattern. In particular, Gated Recurrent Unit (GRU) neural network is used to learn current and historical driving habits and the influence of external environment on energy consumption, while life cycle assessment (LCA) method is used to obtain the emission patterns of vehicles with different energy types in the whole life cycle. The measured data are obtained in Wuhan, the precision of our model is higher than that of the existing model. At the same time, we also applied it to the taxicab monthly trajectory dataset to obtain the spatial-temporal energy consumption emission patterns of different types of energy. The results show that pure electric vehicle (PEV) has obvious greenhouse gas emission reduction effect, compared with gasoline and compressed natural gas (CNG) vehicles, the emission reduction is 12.03% and 12.07% respectively, but the total energy consumption achieves little advantage. This model will lay a foundation for the formulation of regional road network emission inventory, so as to provide support for the government to make relevant decisions.

Keywords. Energy-emission benefits estimation model, GRU network, LCA

\textsuperscript{1} Corresponding Author, Xinyi LI, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei, China; Email: xinyi_li@whu.edu.cn.
1. Introduction

Environmental pollution and energy shortage have become one of the most concerned hotspots in the world. Under the background of carbon emissions, China proposed dual carbon plan, develop the new energy industry [1], new energy vehicles become the key to energy conservation and emissions reduction for auto industry in China. Small cars occupy the highest share and turnover on roads in China, and the data show that their emissions of carbon pollution account for 65%-70% of all of the vehicles [2], indicating that it is worthy of research. Moreover, the importance of taxi industry in China cannot be underestimated. In the car-using scene, it accounts for 24.6% of new energy passenger vehicles [3]. Therefore, the emission and energy consumption of the taxi industry should be monitored.

Since the 1990s, studies on carbon emission benefits of new energy vehicles (pure electric vehicles (EVs), compressed natural gas (CNG) vehicles and hybrid electric vehicles (HEVs), etc.) replacing traditional fuel vehicles have been carried out at home and abroad [4]. Most of the existing studies consider the overall benefit from a macro perspective and obtain the true value unconditionally, so the conclusions are difficult to verify. Then, micro models are increasingly used to calculate energy consumption and carbon emissions. It is mainly aimed at the emission simulation of a motor vehicle passing through a specific road section to obtain the actual pollutant emission of the motor vehicle [5]. It is also combined with the micro traffic simulation model to simulate the road pollutant emission of the road section the motor vehicle passes. However, the micro model can only calculate the road data and cannot judge the consumption behind the energy, so the carbon emission of new energy electric vehicles cannot be settled. Life cycle assessment (LCA) is a popular model to study the carbon emission benefit of all kinds of new energy vehicles [6]. Arar’s [7] research shows that combined with the calculation of the power energy structure in the United States, the carbon emissions of American passenger cars and light trucks will be significantly reduced after the conversion from oil to electricity, and the carbon emission will be reduced by 36% in 2020. Ying et al. [8] using LCA theory simulated carbon emission under various emission reduction scenarios, with results showing that replacing traditional fuel vehicles by pure electric buses will reduce CO2 emission by 19.7%. Pourahmadiyan et al. [9] developed a regional LCA framework and used Simcenter Amesim software for dynamic simulation of vehicles. To evaluate compressed natural gas (CNG) and liquefied natural gas (LNG) as fuels to replace diesel on a bus in Victoria, British Columbia, Canada. However, LCA generally uses energy consumption data provided by car manufacturers and cannot guarantee accuracy. Real data cannot be obtained from the above studies, and there are relatively few studies on real road conditions testing, with low credibility.

Based on the above, we proposed a micro energy consumption & carbon emission model based on observed vehicle information and neural network method, combined with LCA method to evaluate the energy consumption emissions of gasoline, natural gas and electric vehicles. (1) Firstly, the PEMS system is used to measure the road conditions, and the instantaneous observation data of driving state and energy consumption carbon emission of vehicles of the three energy types are obtained. (2) Based on the observed time series data, we use GRU neural network to learn its pattern and obtain the instantaneous energy consumption value of each road. Then the average urban road energy consumption is input into the LCA model as a parameter to obtain a more accurate emission factor. (3) Finally, the trained model is applied to the taxi trajectory dataset in
Wuhan to analyze the temporal-spatial heterogeneity of energy consumption on weekdays and weekends, so as to compare the greenhouse gas emissions of electric vehicles with those of gasoline and natural gas vehicles. This study is beneficial to the sustainable development of ITS.

2. Methodology

The problem can be split as two parts. The first part is aimed to figure out the instantaneous energy consumption $P_{t}^{EC}$ at time stamp $t$ based on the present and past eigen variables $\{X_{t-m}, ..., X_{t}\}$, including vehicle driving condition data $\{DC_{t-m}, ..., DC_{t}\}$ and environment condition data $\{EN_{t-m}, ..., EN_{t}\}$, given the his diachronic observed characteristic variables $\{X_{t-1}, ..., X_{t-1}\}$ and energy consumption $\{P_{t}^{EC}, ..., P_{t}^{EC}\}$. The other takes the CO$_2$ emission $\dot{P}_{t}^{CO2}$ at time $t$ as the target, based on the given condition like energy consumption data $\dot{P}_{t}^{EC}$ and the location emission factors.

2.1. Data Preliminary

Driving condition is a dynamic circumstances that a vehicle is running along the pathway [10]. It includes vehicle motion state variables such like second-by-second velocity $(\nu, a)$, acceleration $(\nu, a)$, VSP $\nu \cdot a$, time step, spatial context variables such as location and slope, and ancillary state variables such as energy category and whether passengers are carried. Environmental condition data indicates the external conditions during driving, often referring to air humidity and temperature. In general, the eigen variables $X$ at time stamp $t$ can be explained as $X_{t} = (\nu, a, vsp, lon, lat, h, s, ec, pass, w)$, where $\nu$ denotes the vehicle velocity, $a$ denotes vehicle acceleration, $vsp$ denotes VSP, $lon$, $lat$, $h$ denote the vehicle location including longitude, latitude and height, $s$ represents vehicle slope, $ec$ represents the energy category, $pass$ indicates whether passengers are carried, and $w$ denotes the weather condition.

As for energy consumption data $\dot{P}_{t}^{EC}$, it refers to how much energy a vehicle consumes at time $t$ (s) in that energy category. The unit of the energy consumption vary according to the energy category, indicating (kWh/100 km) for electric vehicle and (L/100 km) for fuel vehicle. The CO$_2$ emission data $\dot{P}_{t}^{CO2}$ are the amount of emissions at time $t$ (s) from the tailpipe of the vehicle under that energy type.

2.2. Deep-Learning Based Taxicab Energy and Emission Model

As we all know, the vehicle energy consumption and emission are closely related to vehicle driving conditions [12], road environment conditions [13], GPS locations [14]. Moreover, evidences indicate that the time decay pattern is the same for different driving conditions, and the time dependence of traffic flow is also similar [15]. Therefore, the energy consumption emission model we constructed ought to deal with time dependence. The recurrent neural network (RNN) model is designed to deal with sequential information more effectively [16]. However, RNNs are rarely applied in vehicle energy consumption emission models. Only Yu et al. used the LSTM (an enhanced RNN network) to predict NOx emissions [17], Jia et al. also used LSTM to calculate CO$_2$ emissions [15]. On the other hand, GRU network can also process long-dependent data...
and has a simpler structure than LSTM [18]. Above all, we develop a brand new deep-learning energy-emission model (DLEE) which can capture the non-linear associations among the time sequence.

2.2.1. Framework

As shown in Figure 1, the model framework contains two continuous parts. The upper layer is the deep-learning based network model and the below layer is the LCA model. The first part is made up of neural networks.

The input $X$ of the model consists of driving conditions and external environment conditions. The environment includes the type of energy (Petrol, CNG, Electricity), whether it is empty, and the weather (sunny, rainy). The above variables need to be independently encoded to represent each token as a more expressive feature vector. The network takes the characteristic $X$, which is sequentially arranged in time series, as the input, and each time inputs the current time variable and the previous variable of time step $M$. The multi-layer GRU network is arranged to learn the vehicle energy consumption pattern. Finally, the energy consumption at the current time $\bar{Y}_{\text{EC}}$ is obtained after going through a full connection layer. GRU learning process formula [19] is as follows:

$$H_t = f(h_{t-1}, x_t)$$  \hspace{1cm} (1)
\[ R_t = \sigma(X_t W_{xx} + H_{t-1}W_{hr} + b_r) \]  
\[ Z_t = \sigma(X_t W_{zx} + (R_t \odot H_{t-1})W_{hh} + b_h) \]  
\[ \tilde{H}_t = \text{tanh}(X_t W_{xh} + H_{t-1}W_{hh} + b_r) \]  
\[ H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t \]  

The LCA process includes five parts: raw material acquisition, production, transportation, maintenance and scrap recovery. LCA framework consists of four stages: determination of target and boundary scope (determination of functional units), process inventory analysis (data collection and modeling), impact assessment (calculation) and result interpretation [20]. The results obtained in the first part—energy consumption value \( \hat{Y}_{EC} \) were input into the model as the parameters of vehicle operation and energy use stage, and the customized single-vehicle CO\(_2\) emission result \( \hat{Y}_{single}^{CO2} \) was obtained.

2.2.2. Training and Testing of the Network

For training, vehicle driving state variables and their variations are inputs. Secondly, forward and back propagation algorithms and Adam optimization methods are used, ReLu function has been used as the activation method, and the input data is gradually input into our model to minimize the mean square error (E) between \( Y_{ec} \) and \( \hat{Y}_{ec} \).

\[ E = \| Y_{t}^{EC} - \hat{Y}_{t}^{EC} \|^2 \]  

For the test model, we used the Root Mean Square Error (RMSE) to measure the accuracy of energy consumption and emissions.

\[ RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (Y_{t}^{EC} - \hat{Y}_{t}^{EC})^2} \]  

3. Results

3.1. Dataset and Preprocessing

There are two data sets used in this study, one is observed dataset, the other is Wuhan taxi monthly trajectory dataset.

The observed data set is used to obtain instantaneous energy consumption data through OBD and GPS device for taxicabs driving on real road driving conditions. We collected a total of 63 rides containing 179,471 data points, where about 69.7% were collected from CNG vehicles, about 20.1% were collected using petrol vehicles (PVs), about 10.2% were collected from PEV. After data washing and alignment, training and test dataset have been obtained.

The second dataset is the monthly GPS trajectory data of Wuhan taxicabs. Overall, 17,199 vehicles contributed 1,891,138,793 points in August 2018. The models used in this dataset are the same as those measured by us. As shown in Figure 2a, one-day taxi track data covers most roads in Wuhan. In addition, each trajectory data contains longitude, latitude, altitude and velocity information. Through map matching with SRTM DEM elevation dataset [20], track data with slope were obtained, as shown in Figure 2b.
3.2. Model Settings

3.2.1. DL Model

The model parameter settings are shown in Table 1. In the measured data, there are three different datasets: PV, CNG and EV datasets. We divide them into training set and testing set respectively. 70% of the data is selected for training and 30% is selected for testing, and the 10-fold cross validation method is used in the training set. In this way, our model can be trained and tested. The precision of the model is $R^2 = 0.95$, and the energy consumption value obtained can be used for subsequent calculation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time step length</td>
<td>18</td>
<td>Training epochs</td>
<td>5000</td>
</tr>
<tr>
<td>Batch size</td>
<td>60</td>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>GRU cells</td>
<td>64</td>
<td>GRU layers</td>
<td>2</td>
</tr>
</tbody>
</table>

3.2.2. LCA Model

In this study, the GREET (Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation) model of Argonne Laboratory was used to calculate LCA [21]. The functional unit is the vehicle running 1km. Research boundary from vehicle material production to vehicle scrap recycling process involved in each stage of carbon emissions, it includes the upstream stage of fuel (such as the production and transportation of fuel raw materials and fuel production and transportation), the downstream stage of fuel (fuel consumption), the production and processing of taxi raw materials (battery system included in pure electric buses), assembly and manufacturing, operation, supporting facilities operation and scrap recovery. Except for energy consumption, other parameters refer to the Evaluation Report of Greenhouse Gas and Air Pollutant Emission in Automobile Life Cycle issued by China Society of Automotive Engineering [22]. The wheelbase of the taxi subject in this experiment is 2550 mm and belongs to Class A. Through calculation results of fuel cycle and material cycle integration, combined with the taxi lifecycle mileage of 600000 km and use fixed number of years for eight years, we choose to central China power model, grade A passenger car material cycle emissions listing on different kinds of energy (PEV, PV and CNG) modeling, and eventually get the vehicle CO$_2$ emissions.
According to the above section, the average energy consumption of Wuhan electric taxi is 16.7kwh/100 km, which exceeds the official data but is within the credible range. The average energy consumption of oil and gas taxicabs is 8.1L/100 km and 5.3L/100 km. Substitute this into the calculation of the LCA model. We obtained the energy consumption and carbon emission values of three different energy types of taxicabs in the whole life cycle.

3.3. Evaluation of deep-learning energy-emission (DLEE) Model

A widely accepted Comprehensive Modal Emissions Model (CMEM) energy consumption model was used for comparison with our model. CMEM calculates energy consumption and emissions per second through vehicle driving mode and engine running parameters, which can be well combined with traffic simulation model [23]. However, the CMEM model, whose parameters are fixed and cannot be modified, is not able to measure the energy consumption and emissions of EV types. Therefore, the energy consumption of Petrol and CNG can be compared here. It should be pointed out that our complete model is not suitable for comparison with CMEM because it calculates the emission value in the whole life cycle.

It can be seen that the RMSE of our model is higher than that obtained by CMEM.

![Figure 3. Accuracy evaluation. (a): RMSE of DLEE and CMEM with Petrol, CNG; (b): The test performance of DLEE on EV.](image)

4. Discussion

4.1. Temporal-Spatial Heterogeneity Analysis of Energy Consumption and Emission

After obtaining a high-precision instantaneous energy consumption model, we applied the model to Wuhan trajectory dataset to analyze its temporal spatial heterogeneity. We take 9:00 am on weekday (2018.08.13) and 9:00 am on weekend (2018.08.18) for comparative analysis (assuming CNG is used), as shown in Figure 4.

As can be seen from the figure, under the same classification method, road energy consumption on weekdays is generally higher than that on rest days. The key area is Henglong Shopping Mall in Wuhan, where the road energy consumption on weekdays is higher than that on rest days, and the emission on natural weekdays is also higher.
4.2. Comparative Analysis of Different Energy Types

As can be seen from Table 2, in terms of energy consumption, the advantages of PEV taxicabs are not dominant. As the energy volume of PEV and the construction of external equipment require a larger amount of fuel consumption, the energy consumption of PEV in the whole life cycle is similar to that of Petrol vehicles and CNG vehicles.

In terms of CO₂ emission, the life-cycle GHG emission level of PEV taxi is 12.03% lower than that of Petrol and 12.07% lower than that of CNG. Overall, PEV taxicabs have a life-cycle GHG reduction effect.

<table>
<thead>
<tr>
<th>Type</th>
<th>Fuel cycle</th>
<th>Vehicle cycle</th>
<th>Life cycle assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PV</td>
<td>CNG</td>
<td>PEV</td>
</tr>
<tr>
<td>EC (gce/km)</td>
<td>145.67</td>
<td>137.77</td>
<td>134.66</td>
</tr>
<tr>
<td></td>
<td>18.98</td>
<td>26.89</td>
<td>32.43</td>
</tr>
<tr>
<td></td>
<td>164.65</td>
<td>164.66</td>
<td>167.09</td>
</tr>
<tr>
<td>CO₂ (gCO₂ eq/km)</td>
<td>342.58</td>
<td>327.38</td>
<td>274.56</td>
</tr>
<tr>
<td></td>
<td>69.78</td>
<td>85.14</td>
<td>93.51</td>
</tr>
<tr>
<td></td>
<td>412.36</td>
<td>412.52</td>
<td>368.07</td>
</tr>
</tbody>
</table>

5. Conclusion

In this study, a deep-learning energy-emission (DLEE) model and life cycle assessment are proposed to measure the energy consumption and emission benefits under different energy types. The experimental accuracy is 0.95 and performs better than CMEM. The innovations of this paper are listed below. Firstly, we creatively combined time-dependent deep learning method and LCA method to propose a micro energy consumption carbon emission model. After the training, the model can simulate multi-energy data based on simple basic driving state data, which is relatively easy to obtain and high precision of calculation results. Secondly, this paper evaluates the social benefits from the perspective of spatial and temporal heterogeneity, and makes comparative analysis of different energy consumption types by hour, which conducts more comprehensive results. Then, since Taxicab occupies a large proportion of roads but rarely has new energy consumption emission analysis, the customized energy consumption emission calculation method for taxicab is proposed.

The refined analysis provides technical support for government decision-making. However, this paper only discusses the current energy consumption and emissions, and has not yet explored the future continuous energy consumption emissions of vehicles with different energy types. Subsequent studies will consider simulations under different carbon reduction scenarios for dual carbon target years. Finally, this study can provide
technical support for the elaboration of urban emission inventory, and provide reference for the next step of Wuhan energy strategy development and the replacement of clean energy for motor vehicles.

References