sustainability

MDPI

Article

Fuel Consumption Monitoring through COPERT Model—A Case Study for Urban Sustainability

Muhammad Ali, Muhammad Daud Kamal, Ali Tahir * and Salman Atif

Institute of Geographical Information Systems, National University of Sciences and Technology, Islamabad 44000, Pakistan; mali.ms17igis@igis.nust.edu.pk (M.A.); dkamal.phd20igis@student.nust.edu.pk (M.D.K.); salman@igis.nust.edu.pk (S.A.)
* Correspondence: ali.tahir@igis.nust.edu.pk; Tel.: +92-51-9085-4476

Abstract: Trackers installed in vehicles gives insights into many useful information and predict future mobility patterns and other aspects related to vehicles movement which can be used for smart and sustainable cities planning. A novel approach is used with the COPERT model to estimate fuel consumption on a huge dataset collected over a period of one year. Since the data size is enormous, Apache Spark, a big data analytical framework is used for performance gains while estimating vehicle fuel consumption with the lowest latency possible. The research presents peak and off-peak hours fuel consumption's in three major cities, i.e., Karachi, Lahore and Islamabad. The results can assist smart city professionals to plan alternative trip routes, avoid traffic congestion in order to save fuel and time, and protect against urban pollution for effective smart city planning. The research will be a step towards Industry 5.0 by combining sustainable disruptive technologies.

Keywords: COPERT; Global Positioning System (GPS); fuel consumption; apache spark; big data; spatio-temporal queries

check for updates

Citation: Ali, M.; Kamal, M.D.; Tahir, A.; Atif, S. Fuel Consumption Monitoring through COPERT Model—A Case Study for Urban Sustainabilit. *Sustainability* **2021**, *13*, 11614. https://doi.org/10.3390/su132111614

Academic Editor: Fahim Ullah

Received: 5 August 2021
Accepted: 13 October 2021
Published: 21 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As urbanization increases, the issues related to the environment and transportation contribute to route efficiency of the vehicles. A total of 20–30% of Greenhouse Gases (GHG) are released into the environment from transportation claimed by the report of the Intergovernmental Panel on climate change [1]. GHG harms humans indirectly by polluting the environment and therefore many policymakers are interested to estimate the GHG emission in different sectors, which includes energy, transportation, agriculture, environment, etc. Transportation has a significant role in GHG emissions [2]. The increasing population rate is now affecting every aspect of the environment including transportation as people use traffic to reach to their destination easily. It is estimated that 90% of pollution in urban areas is due to emission from vehicles [3]. The most expensive and scarce resource in the world is fuel energy, which is not used effectively and hence wasted. This makes us conclude that there is a need to minimize this energy source as it is an important concern in sustainable engineering and urban design [4].

To encourage people to save fuel energy on their trips, users should know how much energy their vehicles consume on alternative routes. Knowing the traffic conditions not only enables the navigation system on the roads but also improves navigational accuracy especially in routing applications. Hence extra fuel energy can be saved, and the environment can be protected [5]. Fuel consumption is a significant economic index and has an adverse effect on the environment. Hence, it is very essential to estimate the energy consumption of different vehicles under the influential factors [6]. China's traffic pollution is higher than the developed countries due to it being the world's most populous country. Therefore, the country has a higher percentage of transportation use and consumes high percentage of fuel. The traffic congestion in such populated countries is also causing huge financial losses to the economy [4]. The fuel economy can be enhanced, and

the environment can be protected by identifying the factors that affect fuel consumption wastage in the congestion and vehicles standing at traffic signals [7].

The main objective of this study is to estimate energy consumption from high-resolution GPS trajectory data generated by vehicles. The trajectories form geospatial big data. For performance gains and to increase the latency, this research analysis is conducted on Apache Spark, which is a big data platform. Furthermore, COPERT [8] model is used in our research, which helps predict the energy consumption from vehicles. COPERT is a mathematical model, which is implemented using Python language in Apache Spark in order to reduce the time taken for each query. Apache Spark reduces the time taken per job significantly. The model will help generate high speed query results for urban sustainability for issues related to traffic, for instance, for long term planning of urban and roads development within smart cities context. This study is conducted on major cities of Pakistan including Karachi, Lahore and Islamabad. The fuel consumption maps are produced for each city as per its dynamics.

The remainder of the paper is as follows. Section 2 describes all the relevant studies. The methodology and proposed approach are explained in Section 3. Section 4 presents our findings and finally, the conclusions and direction to future work are given in Section 5.

2. Related Work

Fuel energy consumption from vehicles is mainly influenced by the vehicle engine. Fuel consumption can be calculated from different factors such as vehicle engine, engine cycle condition as well as vehicle parameters [9]. Fuel consumption can be estimated from the actual experiments and inspecting the independent effective variables [10,11]. This can further be calculated using the carbon balance technique. In all the above techniques, the experimental assessment is more accurate and reliable technique in estimating vehicle fuel consumption [12]. In the early studies, the fuel consumption/emission was estimated from the aggregated data at a large spatio-temporal scale [9]. Similarly, trajectories are used to calculate travel time and speed for each road segment, and accordingly estimated emission levels as per the study area [13]. The authors took the average of most recent GPS tracks.

Likewise, in a similar study, the authors estimated fuel consumption and emissions using a microscopic model (CMEM) using space-time paths [14]. Other studies consider city dynamics when estimating fuel consumption. For example, the cities are distinguished by the high concentration of vehicles energy consumption, emission, population, and industries. To minimize the GHG and, in particular, the fuel energy consumption by developing alternative energy sources that are inexpensive and which increase the vehicle efficiency, better planning is desired [15]. Emission factor is typically expressed in gram per liter fuel burnt with the contextual information on vehicle fuel, model, and driving conditions [16]. Fuel energy consumption is also been affected by the driving behavior of the driver on the road. There are other studies on the driving behavior evaluation of traffic safety including road accidents and fuel energy consumption [17,18]. More advanced data handling techniques are also being studied. Data analytics, data mining technology and neural network techniques have been applied which classifies the driving styles based on the aggressive character of the driver behavior and its effects on the fuel consumption [19]. Remote sensing using satellite imagery has also been evaluated as an alternative approach to establishing fuel-based mobile source emission inventories [8].

The driving behavior data of the drivers using smart phones were analyzed using discrete wavelet transformation and the signal variance along with safety features using K-nearest neighbor algorithm [20]. Similarly, the fuel energy consumption and safety characteristic of drivers from their driving behavior were described by [21]. In another study, classified natural driving data using different techniques such as the Hidden Markov Model, hierarchical clustering, and dynamic time warping is used. GPS vehicles data analysis has been hot research area, may it be vehicles, train, or any other moving object. The knowledge obtained from movement patterns can help analysts in making informed

decision. For example, vehicle future location prediction using geo-spatial big data context within sustainability and urban design are recently reported [22].

In a study for quantification of instantaneous drive cycle and parameters in high spatio-temporal data, a novel approach was enabled by the authors in [23] which provided the basis for a microscopic emissions model. In an article by [24], where efforts were made using a novel algorithm and architecture to handle big data in smart cities. In a study by [25], several taxis data were decimated to find out why a taxi driver declines to serve taxi customers in Bangkok using an approach that is data driven for taxi probe data and their collection, analysis was carried out for several months. The authors in [26] look for methodological advancement concerning big data where they point out certain important limitations, including data validation and sample representation. The sample representation remains mainly unsolved. The authors look after the theoretical aspects of limitation regarding big data. The aim of [27] was to understand the relationship between the visitor's path and time they take if there is any change among their patterns using GPS data at Yuanmingyuan Park from January 2014 to August 2020. They used Python and ArcGIS for locating the hot spots where there was a high dwell time.

A traffic flow network model was investigated by [28] for congestion problems through taxi GPS data. For validation, an actual scenario was considered identifying the congestion areas. Big data pattern recognition techniques were analyzed and a survey is conducted for ranking of prediction algorithms [29], which could be useful for new researchers who tend to find trajectory algorithms for their specific needs. A novel model was proposed for data driven smart sustainable cities of future by [30] for providing a strategic planning process for transformative changes towards sustainability in urban planning. Their model combines and integrates the whole system of sustainable urbanization and the newly evolving system of smart urbanism, the outcome of which are the four case studies that were analysed for compact cities, eco-cities, data driven smart cities, and environmentally data driven smart sustainable cities.

SoBigData initiative was developed by [31] for virtual research of environment with mobility datasets and urban analytics for city of citizens thematic area of Horizon 2020 where several intuitions around Europe took part. The paper concluded that the urban data science has yet to address smart cities vision. Different areas of interests were analysed by [32] to find hot-spot which were mainly focusing on primary transit hubs, famous cultural venues, and commercial centres. The role and potential urban computing and intelligence in the strategic planning of data-driven was observed by [33]. Moreover, they developed an advanced form of decision support system for urban intelligence framework for planning function, which contributed to their previous research on sustainable smart cities model. Bus lines and stops-based massive data were generated and analyzed by [34] which further provided a link to network representation for sharing route demand in community detection. These links were segmented into similar traveling routes. They concluded with core customization of bus lines and new lines were generated.

A range of methods were reviewed by [35] which focused on big data that illuminates greater patterns of tourist activities as applied by predicting travel demands for sustainable management of destinations. A GPS big dataset in China optics valley in Wuhan of bikes demands was studied by [36] to propose a space-time demand cube framework that accurately covers models that maximizes the space-time demand coverage and minimizes the total distance between riders and bike stations.

The purpose of this research is to use GPS data as movement dataset and predict fuel consumption. The knowledge obtained from this can be used for improved decision making in smart cities using sustainable disruptive technologies.

3. Materials and Methods

3.1. Study Area

GPS vehicle data are acquired from the local tracker company stationed in Pakistan. The vehicles move around all over the country. Mainly data records are from Islamabad,

Lahore, and Karachi as shown in Figure 1. The study was conducted un these three major cities. Islamabad is the capital of Pakistan. Lahore is the provincial capital of Punjab while Karachi is located in the Sindh province. Karachi is also one of the most populated cities in the world.

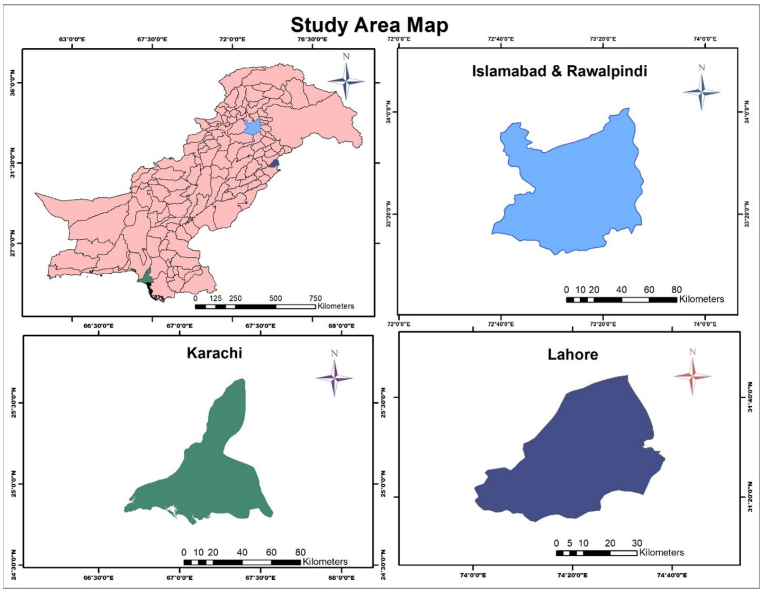


Figure 1. Study Area Map.

3.2. Approach

This section describes our approach such as data collection, data preprocessing, data storage in Apache Spark, Spatio-temporal queries, COPERT Model implementation in Apache Spark using Pyspark. The flowchart of the methodology is shown in Figure 2.

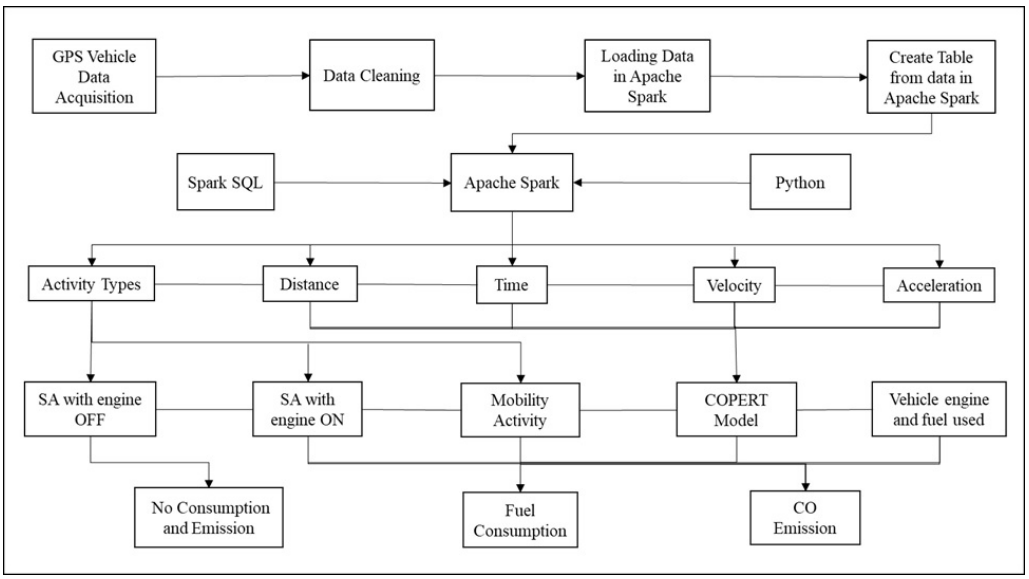


Figure 2. Methodology Flowchart.

3.3. Data Preprocessing

GPS data are obtained from a local tracker company. The data consisted of 32 columns in which vehicle registration, latitude, longitude, and timestamps were used. The data comprised of 12 tables. The data are generated by 107 vehicles. The total size of the data is approximately 30 GB. For analysis, the preprocessing of data is carried out to reduce the data in which unnecessary columns were removed, and the outliers were identified and

removed. The total GPS points are more than 10 million. Initially, the fuel consumption is estimated experimentally gasoline vehicle. Next, the estimated accuracy is validated. The experimental vehicle has a sampling interval of 5 s. A descriptive snippet of the data is shown in Table 1.

Table 1. GPS Sample Data used for One Vehicle.

Vehicle Reg.	Latitude	Longitude	Timestamp
Vehicle A	24.8189	66.98127	01/10/2017 12:47:55 a.m.
Vehicle A	24.8189	66.98137	01/10/2017 12:48:00 a.m.
Vehicle A	24.819	66.9815	01/10/2017 12:48:05 a.m.
Vehicle A	24.81892	66.98167	01/10/2017 12:48:15 a.m.

3.4. Apache Spark

The analytics of this big data on the conventional system is time-consuming and therefore Apache Spark is used for faster processing. Apache Spark is an open-source, general-purpose distributed computing system used for dealing with a huge amount of data. Apache Spark is the fastest tool for completing tasks compared with previous big data tools (such as Apache Hadoop) because of its in-memory caching of data and query execution in an effective way. Apache Spark is written in Scala and provides APIs in Python, R and Java. This further provides API's for machine learning, graph analysis, SQL, and streaming libraries. Apache Spark can access data from diverse data sources such as Hadoop HDFS, S3, Google Cloud, and local data storage. There are three Spark Cluster manager such as standalone cluster manager, Hadoop Yarn, and Apache Mesos [37]. In this study, Python (PySpark) API is used to support Apache Spark. A comparative study was done to find the latency for queries of Apache Spark with other major data processing platform. The results are shown in Table 2 below.

Table 2. Time taken by Spatio temporal queries.

Total Tables	Total Number of Rows	Apache Spark	Apache Hadoop	MySQL
2 tables	15,327,179	33 s	52 s	152 s
4 tables	135,658,227	61 s	92 s	391 s
6 tables	57,386,217	91 s	131 s	696 s
8 tables	78,712,075	120 s	176 s	1128 s
10 tables	93,519,190	147 s	224 s	1635 s
12 tables	105,299,117	201 s	297 s	1935 s

3.5. Spatio-Temporal Queries

For spatial temporal queries, a database was created and then the data were stored in the SparkSQL table. Pyspark API was used for storing the data in Spark database. Spark context was used to define the cores that are used by the local system. The following Python command is used while Table 3 shows a few queries:

```
sc=SparkContext("local[*]","User")
spark = SparkSession.Builder
.master("local")'
.appName("FC_Pak")
.getOrCreate()
```

For estimating the fuel consumption, the GPS data control points were used. These control points have a temporal resolution of 5 s. The control points are divided into three categories, i.e.,

(i) Stationary activity with engine on (if the distance between the control points is less than the threshold value),

(ii) Stationary activity with engine off (the distance between the control point is 0) and
(iii) Mobility activity (if the distance between the control point is greater than the threshold value).

The concept is derived from [1]. The threshold value depends on the positional accuracy of the GPS data. In our data, the positional accuracy of the GPS is 10 m (threshold value). This threshold value affects the classifications of activity type and subsequently effects the fuel consumption. Algorithm 1 explains the steps involved in the queries and Algorithm 2 explains the steps for emissions from the vehicles.

Algorithm 1 Query results

Input: Vehiclereg, datetime, latitude, longitude

```
b ← zip(latitude, longitude)
c ← list(zip(b[:: 1], b[1 :]))
distance_meters ← distance.distance(s[0], s[1]).metersforsinc
for <x in distance_covered> do
    if x > 15 then
        (mobilityactivity)
    end
    if x > 0 AND x < 15 then
        (stationaryactivitywithengineon)
    end
    if x = 0 then
        (stationaryactivitywithengineoff)
    end
    dataframe ← activities
    time ← parser.parse(datetime)
    prior_time ← time.shift(−1)
    time_difference ← time − prior_time
    dataframe ← timedifference
    end
    for <d,t in zip(distance,time_difference> do
                                                          ▷ for velocity
        velocity ← distance/time_difference
        dataframe ← velocity

        for <v,t in zip(velocity,time_difference)> do
                                                          ▷ for acceleration
            acceleration ← velocity/time_difference
            dataframe ← parseracceleration
            fueldensity
            vehicleengine

            for <d,v,t in zip(distance_covered,velocity,time_difference)> do
                if distance >= 15 then
                    (mobilityactivity AND fuelconsumption)
                    FCma ← (217+0.253*velocity+0.00965*(velocity)**2)
                    /(1+0.096velocity0.000421(velocity)**2)end
                    if distance > 0 AND distance < 15 then
                        (stationaryactivitywithengineon)
                        (FCsa ← 0.361 ∗ timeinterval)
                    end
                    else
                        (stationaryactivitywithengineoff)
                        FCsa with engine off ← 0
                        end
                        dist=distance covered in mobility activity ▷ Total fuel consumption
                        is equal to
                        fuel density ← 0.77g/ml
                        FCtotal ← FCma ∗ dist + FCsa ∗ fueldensity
                        end
                    end
                end
```

Algorithm 2 Emissions from the vehicles

procedure EMISSIONS OF CARBON MONOXIDE FROM THE VEHICLES(steps)
COa $\leftarrow$ 71.7
$Cob \leftarrow 35.4$
$COc \leftarrow 11.4$
$COd \leftarrow -0.248$
$COe \leftarrow 0$
$Coα \leftarrow 13.889$

for $d, vinzip(distance, velocity)$ **do**
  if $distance >= 15$ **then**
    $(mobilityactivity)$
    $(EMISSIONS(ma) = (71.7 + 11.4 * Velocity + 0 * (Velocity * *2))/(1 + 35.4 * Velocity + (0.248) * (Velocity * *2)))$
  end
  else
    if $distance > 0$ *AND* $distance < 15$ **then**
      $(stationaryactivitywithengineon)$
      $(EMISSIONS(sa) \leftarrow 13.889 * time)$
    end
    else
      $(stationaryactivitywithengineoff)$
      $(noemissions)$
      $(totalemissionfromthevehicle)$
      $(Totalemissions \leftarrow emissions(ma) * distance + emissions(sa) * 13.889e)$
    end
  end
end
▷ This equations and methodology will give emissions of carbon monoxide.

end procedure

Table 3. Queries.

Query Type	PySpark Queries
Simple Query	sqlcontext.sql ("select * from table where vehicleReg=123").toPandas()
Temporal Query	sqlcontext.sql("select * from table vehicleReg=123 and Timestamp between 'lower timestamp' and 'higher timestamp"').toPandas()
Spatio-temporal Query	sqlcontext.sql("select * from table where timestamp Between 'ower timestamp' and 'higher timestamp' and Furthermore, latitude between 'lower latitude' and 'higher latitude' Furthermore, longitude between 'lower longitude' and 'higher longitude"').toPandas()

3.6. COPERT Model

In this study, the COPERT model is used for estimating vehicle fuel consumption. The GPS data are divided into three categories and different equations were used for calculating the fuel consumption for mobility activity, stationary activity with the engine on while stationary activity with the engine off has no fuel consumption. In this model, the fuel consumption is estimated using the distance traveled by the vehicles between two control points with average velocity. The fuel consumption of the mobility activity of gasoline passenger car with 1.4 L to 2.0 L is estimated using the COPERT Model. According to Equation (1), *FC* is the fuel consumption factor (g/km) while *V* is the average speed for each activity type. All the consumption maps are generated using this technique.

$$FCma = (217 + 0.253V + 0.00965V^2)/(1 + 0.096V - 0.000421V^2) \quad (1)$$

4. Results and Discussion

As the datasets were huge, the vehicles are classified city-wise and accordingly spatial queries are written in Apache Spark. Figures 3 and 4 show the vehicles which are in the proximity of Islamabad, Lahore, Karachi and Motorway, respectively. The visualization shows vehicles movement for a given time duration with a temporal resolution of 5 s. The vehicles are traveling all over the city and depict the traffic conditions on the road such as traffic congestion, fuel consumption, vehicles on the road, vehicles around the traffic signals. All these parameters can be extracted from the data using distance covered in the control points along with the velocity. If there is a frequent change in velocity, this means it can be due to a populated commercial place, which causes small traffic congestion. These are the areas on the roads where vehicles are wasting fuel and energy and causing the environmental problems.

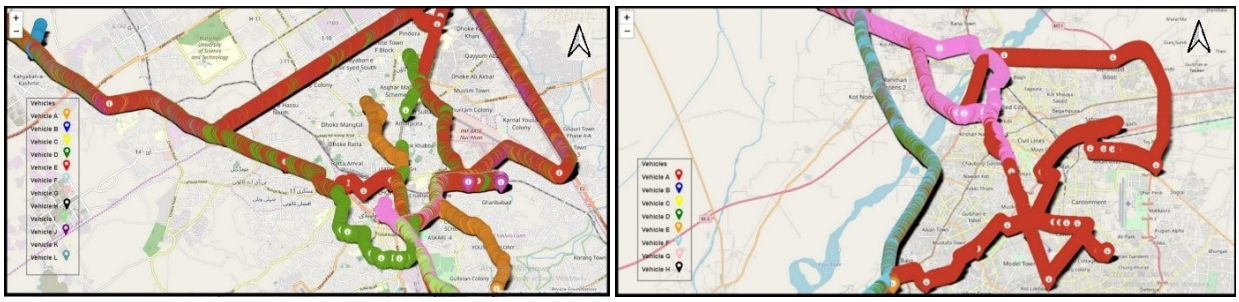


Figure 3. Representation of vehicle traveling in Islamabad (**left**) and Lahore (**right**).

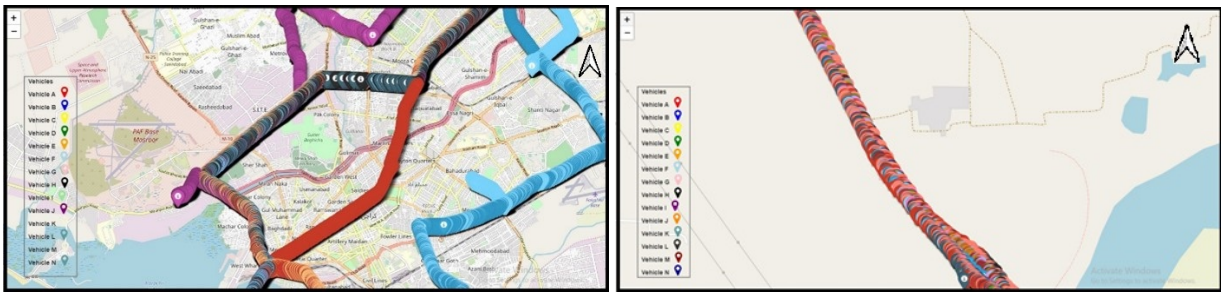


Figure 4. Representation of vehicles traveling in Karachi (**left**) and Motorway (**right**).

Moreover, the activities of all control points were extracted as mobility activity and stationary activity (MAs and SAs) by calculating the average speed of vehicle and distance with each control point. The process was adopted for all the cities including vehicle movements in Islamabad, Lahore, Karachi, and motorway. Mobility Activity and Stationary Activity fuel consumption is calculated using average speed in the COPERT model. Results were estimated for different times such as peak and off-peak hours. The approach used in this study has relatively high accuracy due to dividing GPS control points into three activity types and estimated fuel consumption using the different equations in COPERT model for different activities according to average speed.

Furthermore, the estimates take into account all the scenarios during vehicle movements. For example, mobility activity depicts that on the road, the vehicle is moving freely and the road has no congestion. Similarly, in stationary activity with an engine on the road, the vehicle is unable to move freely due to congestion. When scenarios are run in different cities, vehicle fuel efficiency is calculated and accordingly maps are generated. The visualization of fuel consumption for different cities can be visually analysed in Figure 5 and 6. The volume of fuel consumption of vehicles depends not only on the vehicle engine but also on the road, city, and traffic conditions. As fuel consumption is estimated for all cities in the study, in Islamabad during peak hours, the vehicles covered 12 km distance in 25 min. The fuel consumed during this ride is almost 55 g/km. This means that roads have a lot of traffic in Islamabad during these peak hours; thus vehicles travel slowly in areas

such as Kashmir Highway. In the off-peak hours, the vehicles travelled 17 km distance in 20 min and consumed 60 g/km. Therefore, the traffic on the roads during off-peak hours is apparently less and vehicles travel smoothly, covers more distance in less time and does not consume extra fuel.

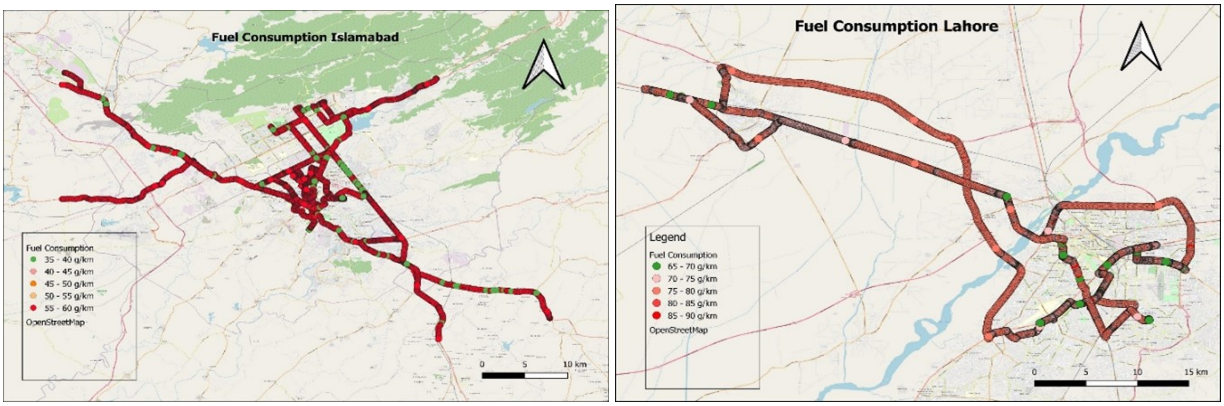


Figure 5. Fuel Consumption in Islamabad (**left**) and Lahore (**right**).

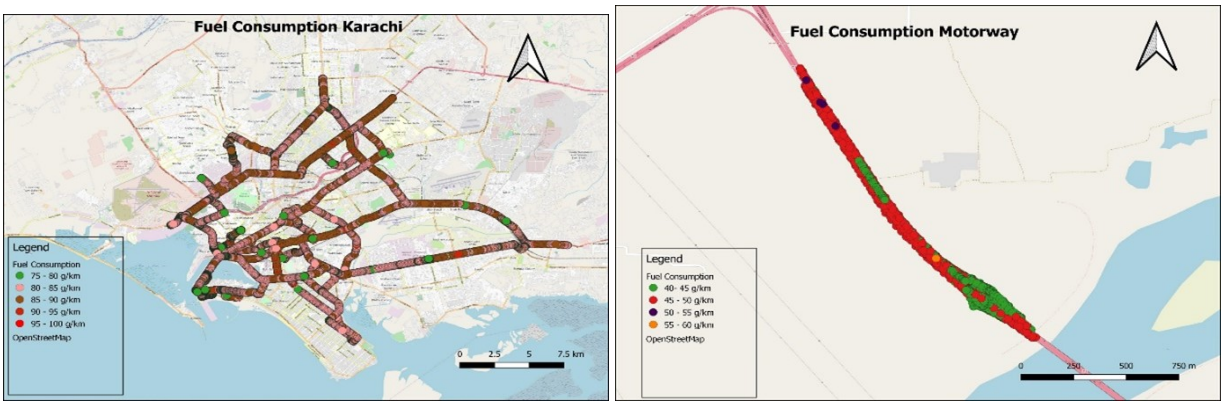


Figure 6. Fuel Consumption in Karachi (**left**) and Motorway (**right**).

Similarly, the fuel consumption is estimated for Lahore during peak hours (see Figure 7). The vehicles covered 7.5 km distance in 19 min and fuel consumed is 80 g/km. The traffic on the roads is higher during peak hours because Lahore is the capital of Punjab and it is the most populated city of Punjab. During peak hours, the vehicles are unable to cover much distance, and faces hurdles in movement and spends extra time on roads. In off-peak hours in Lahore, the vehicle travelled 12 km distance in 21 min and consumed 85 g/km of fuel (see Figure 8). Likewise, fuel consumption is estimated for Karachi during peak hours. The vehicles covered 21 km distance in 59 min and fuel consumed is 95 g/km. Karachi is one of the most populated city in the world. When the data were analyzed, it was evident that there was no significant difference between peak and off-peak hours because Karachi roads remains crowded most of the time as shown in (see Figures 7 and 8). The vehicles experience hurdles in covering distance quickly. Accordingly, Karachi's fuel consumption is not great in comparison to Islamabad and Lahore. The fuel consumption on the motorway is different from the internal cities of Karachi, Lahore, and Islamabad. The results show that on the motorway vehicle moves freely. The vehicle movement all over the day remains the same as either peak or off-peak hours (see Figure 9).

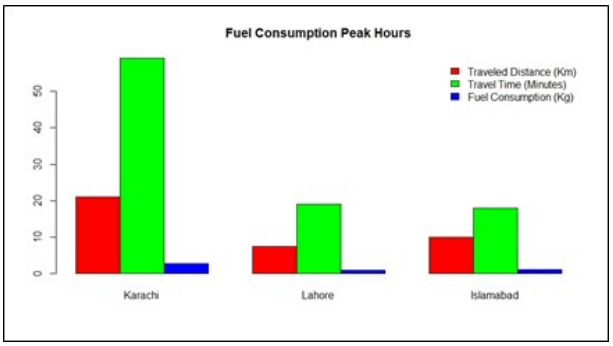


Figure 7. Fuel Consumption of cities in peak hours.

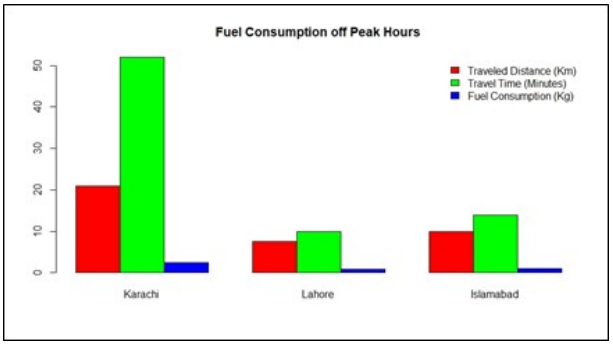


Figure 8. Fuel Consumption of cities in off peak hours.

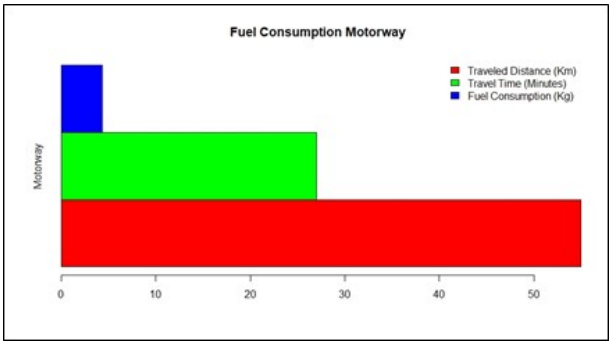


Figure 9. Fuel Consumption of motorway in peak vs. off peak hours.

5. Conclusions

In this study, fuel consumption, a classic case of urban design using disruptive technologies, is investigated in a developing country. High resolution GPS data obtained from vehicles is used to build trajectories. The stationary and mobility activity of vehicles from these trajectories are calculated and used in COPERT model to estimate fuel consumption. Fuel consumption in Karachi is found to be more in comparison to Islamabad and Lahore.

Furthermore, in Karachi, there is not much difference in fuel consumption in peak and off-peak hours, whereas in Lahore and Islamabad, there is a significant difference between peak and off-peak hours. Geographical Information Systems (GIS) maps and graphs show precisely those areas where fuel consumption is high or low in a given time window. These results can be vital for town planners and city administrators to plan city efficiently. Currently, the major cities are undergoing master planning and our findings can act as a guideline for them in order to better design the road infrastructure and thus avoid traffic congestion which will save both energy and time. Similarly, Apache Spark significantly reduced the query response time and is very efficient in GPS big data processing. It has a high accuracy rate as compared to other models.

As part of our future work, we will explore other models and vehicles types where data are available in batches and streams. GPS mobility data analysis is an active area

for research. This study can be used in future to identify the exact spots where vehicles are not able to proceed further or take more time (these are the location where fuel is actually wasted). Thus, actual wastage of fuel is more important for urban sustainability planners who can take proper actions. The fuel consumption and emission impact the overall environment where humans live. Therefore steps taken in this domain will be beneficial to our society.

Author Contributions: Conceptualization, M.A., M.D.K. and A.T.; methodology, M.A. and M.D.K.; software, M.D.K.; validation, A.T., S.A. and M.D.K.; formal analysis, M.A. and M.D.K.; investigation, M.A. and A.T.; resources, M.D.K. and A.T.; data curation, A.T. and M.D.K.; writing—original draft preparation, M.A.; writing—review and editing, M.D.K. and A.T.; visualization, M.A. and S.A.; supervision, A.T. and S.A.; project administration, A.T. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The partial data can be made available upon special request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kan, Z.; Tang, L.; Kwan, M.P.; Zhang, X. Estimating vehicle fuel consumption and emissions using GPS big data. *Int. J. Environ. Res. Public Health* **2018**, *15*, 566. [CrossRef]
2. Wohlstadter, M.; Shoaib, L.; Posey, J.; Welsh, J.; Fishman, J. A Python toolkit for visualizing greenhouse gas emissions at sub-county scales. *Environ. Model. Softw.* **2016**, *83*, 237–244. [CrossRef]
3. Zhao, S.; Yu, Y. Effect of short-term regional traffic restriction on urban submicron particulate pollution. *J. Environ. Sci.* **2017**, *55*, 86–99. [CrossRef] [PubMed]
4. Du, Y.; Wu, J.; Yang, S.; Zhou, L. Predicting vehicle fuel consumption patterns using floating vehicle data. *J. Environ. Sci.* **2017**, *59*, 24–29. [CrossRef]
5. Vilaça, A.; Aguiar, A.; Soares, C. Estimating fuel consumption from GPS data. In *Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Berlin, Germany, 2015; pp. 672–682.
6. Wu, T.; Shen, Q.; Xu, M.; Peng, T.; Ou, X. Development and application of an energy use and CO_2 emissions reduction evaluation model for China's online car hailing services. *Energy* **2018**, *154*, 298–307. [CrossRef]
7. Wickramanayake, S.; Bandara, H.D. Fuel consumption prediction of fleet vehicles using machine learning: A comparative study. In Proceedings of the 2016 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 5–6 April 2016; pp. 90–95.
8. Ekström, M.; Sjödin, Å.; Andreasson, K. Evaluation of the COPERT III emission model with on-road optical remote sensing measurements. *Atmos. Environ.* **2004**, *38*, 6631–6641. [CrossRef]
9. Papadopoulos, G.; Keramydas, C.; Ntziachristos, L.; Lo, T.S.; Ng, K.L.; Wong, H.L.A.; Wong, C.K.L. Emission factors for a taxi fleet operating on liquefied petroleum gas (LPG) as a function of speed and road slope. *Front. Mech. Eng.* **2018**, *4*, 19. [CrossRef]
10. Peça, J.O.; Serrano, J.M.; Pinheiro, A.; Carvalho, M.; Nunes, M.; Ribeiro, L.; Santos, F. Speed advice for power efficient drawbar work. *J. Terramechanics* **2010**, *47*, 55–61. [CrossRef]
11. Park, S.H.; Kim, Y.J.; Im, D.H.; Kim, C.K.; Jung, S.C.; Kim, H.J.; Jang, Y.; Kim, S.S. Development of eco driving system for agricultural tractor. *J. Biosyst. Eng.* **2010**, *35*, 77–84. [CrossRef]
12. Wörz, S.; Bernhardt, H. A novel method for optimal fuel consumption estimation and planning for transportation systems. *Energy* **2017**, *120*, 565–572. [CrossRef]
13. Perrotta, F.; Parry, T.; Neves, L.C. Application of machine learning for fuel consumption modelling of trucks. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 6–10 November 2017; pp. 3810–3815.
14. Kan, Z.; Tang, L.; Kwan, M.P.; Ren, C.; Liu, D.; Pei, T.; Liu, Y.; Deng, M.; Li, Q. Fine-grained analysis on fuel-consumption and emission from vehicles trace. *J. Clean. Prod.* **2018**, *203*, 340–352. [CrossRef]
15. Liu, Y.; Hu, X.; Wu, H.; Zhang, A.; Feng, J.; Gong, J. Spatiotemporal Analysis of Carbon Emissions and Carbon Storage Using National Geography Census Data in Wuhan, China. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 7. [CrossRef]
16. Özener, O.; Özkan, M. Fuel consumption and emission evaluation of a rapid bus transport system at different operating conditions. *Fuel* **2020**, *265*, 117016. [CrossRef]
17. Zhang, Y.T.; Claudel, C.G.; Hu, M.B.; Yu, Y.H.; Shi, C.L. Develop of a fuel consumption model for hybrid vehicles. *Energy Convers. Manag.* **2020**, *207*, 112546. [CrossRef]

18. Li, H.; Ravey, A.; N'Diaye, A.; Djerdir, A. Online adaptive equivalent consumption minimization strategy for fuel cell hybrid electric vehicle considering power sources degradation. *Energy Convers. Manag.* **2019**, *192*, 133–149. [CrossRef]
19. Meseguer, J.E.; Toh, C.K.; Calafate, C.T.; Cano, J.C.; Manzoni, P. Drivingstyles: A mobile platform for driving styles and fuel consumption characterization. *J. Commun. Netw.* **2017**, *19*, 162–168. [CrossRef]
20. Lotfi, R.; Ghatee, M. Smartphone based driving style classification using features made by discrete wavelet transform. *arXiv* **2018**, arXiv:1803.06213.
21. Yao, Y.; Zhao, X.; Wu, Y.; Zhang, Y.; Rong, J. Clustering driver behavior using dynamic time warping and hidden Markov model. *J. Intell. Transp. Syst.* **2021**, *25*, 249–262. [CrossRef]
22. Kamal, M.D.; Tahir, A.; Kamal, M.B.; Naeem, M.A. Future location prediction for emergency vehicles using big data: A case study of healthcare engineering. *J. Healthc. Eng.* **2020**, *2020*, 6641571. [CrossRef]
23. Nyhan, M.; Sobolevsky, S.; Kang, C.; Robinson, P.; Corti, A.; Szell, M.; Streets, D.; Lu, Z.; Britter, R.; Barrett, S.R.; et al. Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmos. Environ.* **2016**, *140*, 352–363. [CrossRef]
24. Karimi, Y.; Haghi Kashani, M.; Akbari, M.; Mahdipour, E. Leveraging big data in smart cities: A systematic review. *Concurr. Comput. Pract. Exp.* **2021**, e6379. [CrossRef]
25. Phiboonbanakit, T.; Horanont, T. Analyzing Bangkok city taxi ride: Reforming fares for profit sustainability using big data driven model. *J. Big Data* **2021**, *8*, 1–27. [CrossRef]
26. Tao, S.; Zhang, M.; Wu, J. Big data applications in urban transport research in Chinese cities: An overview. In *Big Data Applications in Geography and Planning*; Edward Elgar Publishing: Cheltenham, UK, 2021 .
27. Yao, Q.; Shi, Y.; Li, H.; Wen, J.; Xi, J.; Wang, Q. Understanding the Tourists' Spatio-Temporal Behavior Using Open GPS Trajectory Data: A Case Study of Yuanmingyuan Park (Beijing, China). *Sustainability* **2021**, *13*, 94. [CrossRef]
28. Qin, J.; Mei, G.; Xiao, L. Building the traffic flow network with taxi GPS trajectories and its application to identify urban congestion areas for traffic planning. *Sustainability* **2021**, *13*, 266. [CrossRef]
29. Daud Kamal, M.; Tahir, A.; Babar Kamal, M.; Moeen, F.; Naeem, M.A. A Survey for the Ranking of Trajectory Prediction Algorithms on Ubiquitous Wireless Sensors. *Sensors* **2020**, *20*, 6495. [CrossRef] [PubMed]
30. Bibri, S.E.; Krogstie, J. A novel model for data-driven smart sustainable cities of the future: A strategic roadmap to transformational change in the era of big data. *Future Cities Environ.* **2021**, *7*, 3. [CrossRef]
31. Andrienko, G.; Andrienko, N.; Boldrini, C.; Caldarelli, G.; Cintia, P.; Cresci, S.; Facchini, A.; Giannotti, F.; Gionis, A.; Guidotti, R.; et al. (So) Big Data and the transformation of the city. *Int. J. Data Sci. Anal.* **2021**, *11*, 311–340. [CrossRef]
32. Liu, Y.; Singleton, A.; Arribas-Bel, D.; Chen, M. Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in New York City. *Comput. Environ. Urban Syst.* **2021**, *86*, 101592. [CrossRef]
33. Bibri, S.E. Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning. *Comput. Urban Sci.* **2021**, *1*, 1–29. [CrossRef]
34. Yu, Q.; Zhang, H.; Li, W.; Song, X.; Yang, D.; Shibasaki, R. Mobile phone GPS data in urban customized bus: Dynamic line design and emission reduction potentials analysis. *J. Clean. Prod.* **2020**, *272*, 122471. [CrossRef]
35. Xu, F.; Nash, N.; Whitmarsh, L. Big data or small data? A methodological review of sustainable tourism. *J. Sustain. Tour.* **2020**, *28*, 144–163. [CrossRef]
36. Yang, L.; Zhang, F.; Kwan, M.P.; Wang, K.; Zuo, Z.; Xia, S.; Zhang, Z.; Zhao, X. Space-time demand cube for spatial-temporal coverage optimization model of shared bicycle system: A study using big bike GPS data. *J. Transp. Geogr.* **2020**, *88*, 102861. [CrossRef]
37. Spark, A. Lightning-Fast Unified Analytics Engine. 2021. Available online: https://spark.apache.org (accessed on 27 July 2021).