
EXPLORING THE POTENTIAL OF WEB BASED INFORMATION OF BUSINESS POPULARITY FOR SUPPORTING SUSTAINABLE TRAFFIC MANAGEMENT

Jorge M. Bandeira¹, Pavlos Tafidis² Eloísa Macedo¹, João Teixeira¹, Behnam Bahmankhah¹, Cláudio Guarnaccia³, Margarida C. Coelho¹,

¹ *University of Aveiro, Centre for Mechanical Technology and Automation (TEMA), Department of Mechanical Engineering, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal, (+351) 234 370 830, jorgebandeira@ua.pt, macedo@ua.pt, jpteixeira@ua.pt, behnam.bahmankhah@ua.pt, margarida.coelho@ua.pt*

² *Hasselt University - Campus Diepenbeek Agoralaan Gebouw H - B-3590 Diepenbeek Kantoort H-B103b, Belgium (+32)11 29 2134 pavlos.tafidis@uhasselt.be*

³ *University of Salerno, Department of Civil Engineering, via Giovanni Paolo II 132, Fisciano, Italy, cguarnaccia@unisa.it*

This paper explores the potential of using crowdsourcing tools, namely Google "Popular times" (GPT) as an alternative source of information to predict traffic-related impacts. Using linear regression models, we examined the relationships between GPT and traffic volumes, travel times, pollutant emissions and noise of different areas in different periods. Different data sets were collected: i) crowdsourcing information from Google Maps; ii) traffic dynamics with the use of a probe car equipped with a Global Navigation Satellite System data logger; and iii) traffic volumes. The emissions estimation was based on the Vehicle Specific Power methodology, while noise estimations were conducted with the use of "The Common Noise Assessment Methods in Europe" (CNOSOS-EU) model. This study shows encouraging results, as it was possible to establish clear relationships between GPT and traffic and environmental performance.

Keywords: ICT, Google Maps, Crowdsourcing, Noise, missions, Sustainable Transport

1. Introduction

Road transportation is responsible for many external effects, such as air pollution, accidents, traffic congestion and noise (Santos, 2017). The reduction of the aforementioned negative externalities is one of the key objectives of the European Commission, namely by increasing the efficiency of the transportation system, taking full advantage of digital technologies (EC 2019). The global financial crisis that started a decade ago has also affected the transportation sector. In recent years, funding for services and infrastructure has been significantly reduced, which resulted in the decline of the reliability and effectiveness of transportation systems. Under the current circumstances, the collection of high quality data to support their operation has become more complicated taking also into consideration that traditional methods are usually costly, lengthy, limited to specific areas (Tafidis et al., 2017) and the data have poor quality (Serna et al., 2017). However, the possibilities that arise from the recent advancements in communication technologies can provide alternative sources of information that will overcome the current barriers, offering real-time data that captures the patterns, needs and experiences of road users.

Social media and web services can be considered both as cost efficient and effective data input, having valuable information to be harvested, although their use in transportation planning and management is still sporadic (Majumdar, 2017). This type of information has the feature to complement or even replace in certain cases traditional data after distinguishing the useful from the useless data and examining its utility and reliability. In the last years, many studies have explored the potential of using web-based data sources for transportation planning, management or operation (Rashidi et al., 2017). The real-time information that they provide allows commuters to improve their travel experience and transportation authorities to enhance their services quality. More specifically, it can allow city and transportation planners to gain a better understanding of mobility patterns and needs, while for individuals to move freely and reducing travel time (Tasse & Hong, 2014).

Human mobility is possible to be explained by 10 to 30% of social relationships and 50 to 70% by periodic behaviour (Cho et al., 2011). Many initiatives also tried to estimate traffic-related impacts using alternative sources of information. Tostes et al. (2013) tried to estimate traffic jams using information acquired from Bing Maps in the city of Chicago. Ni et al. (2014) developed a prediction model of short-term traffic flow based on Twitter features and focused on traffic conditions prior to sport events. Social media data has also been examined as a new data source to estimate travel demand. Location-based social networking (LBSN) data was used to estimate origin-destination (OD) matrix and when compared to traditional methods it can provide much higher temporal resolution at a lower cost (Yang, 2017; Jin, 2014). In another study, Lee et al. (2013), estimated an OD matrix based on social-media travel data acquired from Twitter and results were compared with those of a traditional travel demand model in the Greater Los Angeles metropolitan area and the preliminary findings were especially encouraging. Under the same context, Chaniotakis et al. (2016) examined data from different social media and compared it with conventional travel-diary surveys from the city of Thessaloniki, Greece, with the aim of identifying alternative sources of information to improve Intelligent Transportation Systems applications.

A growing body of research also explores potential correlations between traffic-impacts and data from social media. For instance, Ribeiro et al. (2012) found a mentionable correlation between real traffic conditions and data from Twitter regarding traffic conditions in Belo Horizonte, Brazil, while Tian et al. (2016) validated traffic incidents mentioned by social media users by comparing them with field cameras data in Austin, Texas. Pereira et al. (2015) developed a probabilistic data analysis model aimed to specify regular and nonregular overcrowding hotspots in public transportation systems using also data from social networks. The results showed the potential applicability of the proposed model in different cases. Finally, Teixeira et al. (2017) and Tafidis et al. (2017) explored the correlations between traffic congestion, emissions, speeds and traffic volumes with Google Maps traffic data. The preliminary findings showed encouraging results in the prediction of emission impacts in urban arterials. Nair et al (2019) used Google APIs for managing traffic speed data. The API calculates a representative speed value from the available crowdsourced data on a road link at any time of the day and estimated a congestion index for 29 major cities. Google speed data reflects loop detector speed both at the corridor level and at smaller of road sections within a corridor (Nair et. al, 2019).

Popular times is a feature in Google Maps that was launched in July 2015 and allows users to have a better insight on a place's busy time periods. With such information, people are assisted in their decision-making process regarding the best time to visit a specific place or area.

According to the best knowledge of the authors, no research has focused on analysing the potential of crowdsourcing information systems such as Google Popular times (GPT) to predict traffic conditions, greenhouse gases, critical air pollutants and noise emissions in different contexts yet. This article intends to fill this research gap. This information can be an add value to existing traffic information provided (e.g. google traffic) given that:

- In links with low traffic volumes (but with circumstantially high demand levels) there is still no traffic information,
- Existing information reflects current values or historical trend data, not allowing to anticipate circumstantial short-term congestion and pollution.

2. Methodology

2.1. Methodological Approach

In this paper, we empirically analyse whether GPT may be a reliable source to predict the demand of several attraction poles of mobility in cities, and at the same time if it can explain the variability in traffic activity and related environmental impacts in the surrounding areas. Specifically, this study focused on exploring relationships between GPT data with traffic volumes, travel times, emissions and traffic noise in different areas and over different periods. The methodology is based on three fundamental stages: 1) empirical work, 2) modelling of traffic-related impacts and, 3) statistical analysis and model development (see Figure 1).

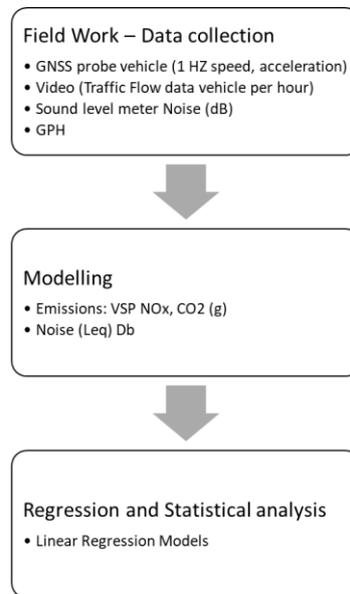


Figure 1 Methodology overview

2.2. Data collection

2.2.1 Activity Data

Google Maps presents popular times with the use of a bar chart. Popularity for any given hour is shown relative to the typical peak popularity for the business for the week, so no absolute values are provided. This information describes how busy a certain place is during different times of the day and based on average popularity over the last several weeks. Live visit data is updated in real-time and overlaid on the popular times graph. In order to obtain a discrete variable that can be correlated with the other parameters of the study, we developed a methodology to transform the qualitative information provided in GPT into a quantifiable variable. We assume that the minimum value of the bar is zero and the maximum is one. Then, we divided it in ten equal parts and we assigned a value between 1 (maximum occupancy) and 0 (minimum occupancy) for real-time bar in pink (Figure 2).

During the experiment, a print screen of this information was stored all 15 minutes to enable a subsequent graphical analysis of the information.



Figure 2 Example of information provided by GPT and decimal grid developed to assign a discrete variable to real-time popularity information (0.4 in the example provided)

We explored two methodological approaches to find the aforementioned relationships based on a standard interval of 15 minutes. In the first approach, values of each studied variable were compared to GPT value in the end of the respective interval, while in the second the comparison was made with the value in the end of the next 15 minutes period. We only present the results regarding the first approach since the correlations were slightly higher.

Furthermore, we also examined the reliability of this tool to estimate demand in situ, by assessing the relationship between the parking occupancy of a specific commercial area and the popularity of that location provided by GPT.

2.2.2. Traffic data

Different traffic data were gathered in order to support the assessment of demand, and assessment of traffic impacts.

i) Traffic dynamics - A light-duty vehicle equipped with a GNSS data logger was used to gather travel time, instantaneous speed, and acceleration. Each car performed 10 runs per hour for each link. To increase the heterogeneity of the driving behaviour different drivers were used (Downling et al., 2014), which drove according the perception of traffic flow (Turner et al., 1998). These data will serve as input on emission models.

- ii) Traffic volumes - with the use of cameras and manual counts
- iii) Number of vehicles in the parking lot

2.3. Modelling traffic-related impacts

2.3.1 Emission Estimation

For the estimation of the emissions, we used the methodology based on the vehicle specific power (VSP), since it allows the estimation of instantaneous emissions from second-by-second vehicle dynamics. VSP accounts for the effect of different driving modes (acceleration, deceleration, cruise, idling), and also includes a wide range of engine displacement values and can be applied to the Portuguese and Spanish car fleet (US EPA, 2014). VSP is a function of acceleration and deceleration, instantaneous speed and slope expressed as:

$$VSP = v[1.1a + 9.81(a \tan(\sin(\text{grade}))) + 0.123] + 0.000302v^3 \quad (1)$$

where: v = is the vehicle speed (m/s), a is the= vehicle acceleration/deceleration rate (m/s²), and grade = represents the vehicle vertical rise divided by the horizontal run (%).

VSP bins are categorized into 14 modes and each mode is defined by a range of values associated to an emission rate (Coelho et al., 2009). We focus on the estimation of CO₂ and NO_x emissions (respectively, a greenhouse gas and a critical local pollutant - precursor to troposphere ozone and with demonstrated effects in human health).

For the purpose of this study, we considered the following distribution fleet composition:

- 38% of light duty gasoline vehicles and 62% of light duty diesel vehicles for the Portuguese studied areas;

• 44% of light duty gasoline vehicles and 56% of light duty diesel vehicles for the Spanish studied areas based on the respective national vehicle classification (EMISIA, 2017).

Although some deviations may occur regarding the estimation of total emissions, the authors assume this approach as suitable to reflect the relative emissions variation associated to different driving behaviour and congestion levels in the studied road links.

2.3.2 Noise Estimation

Though the team performed noise measurements using a sound level meter the following noise results are based on numerical simulation for two main reasons:

- 1) In several periods, the weather conditions (high wind speed) did not allow to collect reliable data;
- 2) The background noise varies significantly and inconsistently in the study areas.

Therefore, for the traffic noise estimation we used “The Common Noise Assessment Methods in Europe” (CNOSSOS-EU). This model is a uniform approach to noise assessment in European countries. CNOSSOS is based on the assessment of the noise produced by a single vehicle, summing the rolling and the propulsion noise per each frequency octave (in the range 125 Hz - 4000 Hz). The former noise has a log dependence by the mean speed of the traffic flow, while the latter has a linear dependence. Once the power of the source "vehicle" is evaluated for each category of vehicles (passenger cars, medium heavy vehicles, heavy-duty vehicles), the presence of a given number of vehicles per hour and the propagation model are implemented. A line source is assumed, and the formula can include favourable and homogeneous conditions. Several attenuation factors, such as reflections, diffractions, atmospheric effects, ground absorption, etc., can be included in the formula, in order to better simulate the phenomena that affect the propagation (Kephelopoulos et al., 2012). The noise emission of a traffic flow per each link considers the average speed recorded with the GNSS data logger, while the vehicle flow considers the video data and manual counts. Finally, with the use of Equation (2), we estimated noise levels for each street

$$L_{eq} = 10 \text{ Log}(10^{L_i/10} + 10^{L_j/10})$$

where: "Leq = is the sum of the estimated equivalent sound levels of each direction, respectively, Li and Lj."

2.4. Regression analysis

For each 15-minute period, the decimal value of the GPT was matched with observed traffic volume, and estimated system CO₂, NO_x and noise emissions. The relationships between the predictor (GPT) and the response variables were examined using linear models. Specifically, we focused on linear models to describe data, since they are simple and easily interpretable, and application of higher order polynomials (or even more complex models) may result in overfitting (James et al., 2013).

2.5. Case Studies

The case studies are located over 2 representative medium-sized European cities, (Aveiro, Portugal and Badajoz, Spain). Medium-sizes urban areas are particularly attractive test beds to explore the potential of web-based solutions for supporting sustainable traffic management due to three main reasons:

- 1) Providing public transport in relatively low-density areas is usually cost-inefficient and the mobility tends to be dominated by private and individual transport (Bandeira et al., 2018),
- 2) Notwithstanding the lower absolute traffic volumes when compared to large cities, road traffic remains a major contributor to harmful air pollution (Bandeira et al., 2011).
- 3) These cities have limited financial resources to invest in advanced traffic monitoring and management systems.

In a second phase, we selected cities with different characteristics related to demographics, land use and shopping patterns. The city of Aveiro and surrounding area has a population of 77.436 (INE, 2017). The region is characterized by a dispersed distribution of population, for presenting an irregular urban form with high urban sprawl (Meneses, 2010). Some important commercial areas are located outside the urban core. The city of Badajoz has 150 530 (INE, 2018) inhabitants but there is a clear delimitation of the

compact urban area. Moreover, it contains an important commercial area located in the city Centre. Shopping centres are closed at Sundays unlike what happens in Aveiro.

Lastly, we examined three important commercial hotspots with different characteristics regarding its integration in the mentioned urban context (see Figure 3). Two of them are in the city of Aveiro in Portugal, while the last one is in Badajoz, Spain. The study area (a) is Aveiro Shopping Center, which is located in the industrial zone of the city and surrounded by four road links. Link 11 is the main entrance to the commercial zone, while 12 is the main exit of it. Links 13 and 14 connect the city of Aveiro with the industrial zone. The studied links are located between roundabouts and there are various unsignalized intersections with minor roads from both sides. One crosswalk interrupt both links 11 and 12.

The second study area (b) is Glicínias Plaza Shopping Center, which is within the urban area of the city. The main links that lead to (link 15) and out (link 16) of the shopping centre are examined. The links are between roundabouts, one crosswalk interrupts the links and in link 16 there is the only entrance and exit to a gas station. The third studied area (c) is a Hypermarket in Badajoz, Spain. Links 17 and 18 represent, respectively, the main entrance and exit to the shopping centre located near the outer limits of the city. The land use of this area is mainly residential. The studied road links 17 and 18 connect two roundabouts and contain three crosswalks.



Figure 3 Aerial view of study areas and experimental set up scheme (Google Earth) a) Aveiro Shopping Center, Aveiro, Portugal, b) Glicínias Plaza Shopping Center, Aveiro, Portugal, c) Hypermarket, Badajoz, Spain.

3. RESULTS AND DISCUSSION

In this section, we summarize and discuss the main results. First, we analyse the reliability of the GPT to predict total demand in a commercial area. Then, we present the various relationships between GPT and traffic volumes, CO₂ and NOX emissions. Finally, we provide a seasonal comparison of GPT's ability to predict traffic impacts in different periods.

3.1. Reliability of Popular times

As a first step of the analysis, we assess the reliability of GPT to predict overall demand in the case study of Aveiro Shopping Center. As already, mentioned, the commercial area is located in the industrial zone of the city and the low frequency level of public transport services encourage the use of private vehicles. To test the reliability of GPT as a predictor variable for monitoring the area's popularity, we monitored the number of vehicles in the parking lot in 15 minutes intervals during the study period. The analysis (Figure 4) showed that there is a high correlation between the parking occupancy and GPT ($R^2=0.89$). The vehicle occupancy rate has been also recorded (1.81 persons per vehicle), which allowed to estimate that an increase of 0.1 in GPT scale represents an increase in demand of approximately, 142 vehicles and 248 visitors. Understandably, this response would change in other case studies according to the overall attractiveness of each commercial area.

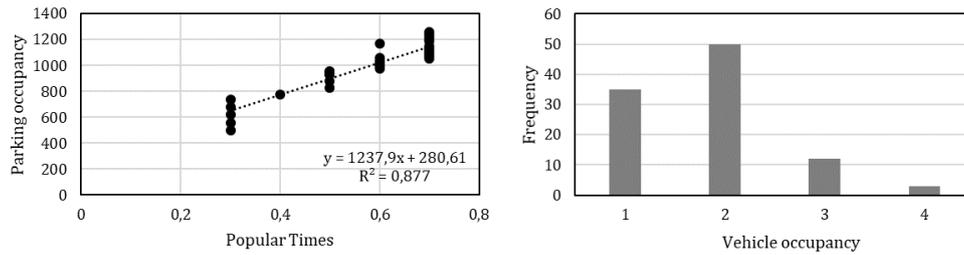


Figure 4 Correlation between Popular times and parking lot occupancy; b) relative distribution of vehicle occupancy.

3.2. Potential of GPT to predict traffic demand and environmental performance

As it can be seen in Table 1, the majority of the examined linear correlations presented good results with statistically significant values ($p\text{-value} < 0.05$), which means that the chosen models significantly predict the response variable. In general, higher coefficient of determination values were obtained for traffic volumes, CO_2 and noise estimations, meaning GPT can explain their variability. Regarding NO_x emissions, the coefficient of determination tend to be lower as NO_x emissions is strongly affected by drivers' behaviour. The linear relationships are not so evident for links l3 and l4 during the weekend maybe because there are no significant traffic flows from those links to the shopping area.

Table 1 Regression parameters of Relationship between model response variables and GPT

| Links | Study Period | Variables | Coefficients of the Model | | R^2 (Coefficient of determination) | p-value |
|---------|--------------|----------------------|---------------------------|--------|---|---------|
| | | | a | b | | |
| ll1 | weekday | Traffic volumes | 154.11 | 35.161 | 0.88 | < 0,05 |
| | | System CO_2 | 8450.9 | 256.63 | 0.93 | < 0,05 |
| | | System NO_x | 29.76 | -3.271 | 0.76 | < 0,05 |
| ll2 | weekday | Traffic volumes | 186.39 | 15.599 | 0.85 | < 0,05 |
| | | System CO_2 | 7517.6 | 168.12 | 0.82 | < 0,05 |
| | | System NO_x | 28.076 | -2.743 | 0.67 | < 0,05 |
| ll1, 12 | | Noise | 4.982 | 54.503 | 0.85 | < 0,05 |
| ll3 | weekday | Traffic volumes | 454.17 | -85.09 | 0.84 | < 0,05 |
| | | System CO_2 | 36971 | 13607 | 0.6 | < 0,05 |
| | | System NO_x | 77.188 | -27.36 | 0.59 | < 0,05 |
| ll4 | weekday | Traffic volumes | 82.264 | 98.755 | 0.53 | < 0,05 |
| | | System CO_2 | 12396 | 10989 | 0.53 | < 0,05 |
| | | System NO_x | 36.757 | 24.407 | 0.55 | < 0,05 |
| ll3, 14 | | Noise | 9.9992 | 54.917 | 0.55 | < 0,05 |
| ll5 | weekday | Traffic volumes | 323.33 | -18.5 | 0.55 | < 0,05 |
| | | System CO_2 | 656.9 | -249 | 0.48 | < 0,05 |
| | | System NO_x | 12.661 | 0.6283 | 0.44 | < 0,05 |
| ll6 | weekday | Traffic volumes | 401.89 | -58.21 | 0.81 | < 0,05 |
| | | System CO_2 | 23681 | -6430 | 0.52 | < 0,05 |
| | | System NO_x | 47.497 | -11.53 | 0.47 | < 0,05 |
| ll5, 16 | | Noise | 7.8001 | 54.922 | 0.69 | < 0,05 |

Table 1 Model Variables between Response Variables and GPT (cont.)

| Links | Study Period | Variables | Coefficients of the Model | R^2 | p-value |
|-------|--------------|-----------|---------------------------|-------|---------|
|-------|--------------|-----------|---------------------------|-------|---------|

| | | | a | b | (Coefficient of determination) | |
|---------|---------|------------------------|--------|--------|--------------------------------|-------|
| ll1 | weekend | Traffic volumes | 222.55 | -9.915 | 0.86 | <0.05 |
| | | System CO ₂ | 9271.9 | 21.937 | 0.77 | <0.05 |
| | | System NOx | 22.67 | 6.1765 | 0.72 | <0.05 |
| ll2 | weekend | Traffic volumes | 166.16 | 21.963 | 0.89 | <0.05 |
| | | System CO ₂ | 5402.8 | 1479.8 | 0.78 | <0.05 |
| | | System NOx | 14.025 | 6.2858 | 0.31 | <0.05 |
| ll1, 12 | | Noise | 7.1113 | 57.653 | 0.75 | <0.05 |
| ll3 | weekend | Traffic volumes | 26.175 | 81.263 | 0.09 | <0.05 |
| | | System CO ₂ | 2480.2 | 10958 | 0.06 | <0.05 |
| | | System NOx | 2.4859 | 30.514 | 0.01 | <0.05 |
| ll4 | weekend | Traffic volumes | 14.398 | 92.259 | 0.01 | <0.05 |
| | | System CO ₂ | -1816 | 14053 | 0.01 | <0.05 |
| | | System NOx | -10.07 | 39.676 | 0.02 | <0.05 |
| ll3, 14 | | Noise | -2.539 | 61.884 | 0.14 | <0.05 |
| ll7 | weekday | Traffic volumes | 253.44 | 33.219 | 0.5 | <0.05 |
| | | System CO ₂ | 45123 | 1314.9 | 0.62 | <0.05 |
| | | System NOx | 103.14 | 5.82 | 0.56 | <0.05 |
| ll8 | weekday | Traffic volumes | 235.97 | 56.292 | 0.48 | <0.05 |
| | | System CO ₂ | 37136 | 8180.8 | 0.46 | <0.05 |
| | | System NOx | 83.132 | 31.149 | 0.3 | <0.05 |
| 17, 18 | | Noise | 5.9122 | 60.244 | 0.39 | <0.05 |

In the next sections, we will look in more detail at the GPT's ability to predict traffic volume, emissions and noise.

3.2.1 Traffic Analysis

The analysis regarding the correlations between GPT suggests that in most cases GPT can explain a significant part of the variation in traffic volumes. All models are adequate and statistically significant to fit the data (p-value<0.05). More specifically, in links 11 and 12 there were high correlations between the number of vehicles and the values of GPT ($R^2 > 0.85$) both during the weekday and the weekend as the two links consist of the main street that leads to the shopping area. Figure 5 presents the linear regression models with higher coefficient of determination. The lowest correlations were achieved in links 17 and 18, since a significant number of drivers used a secondary entrance of the hypermarket zone.

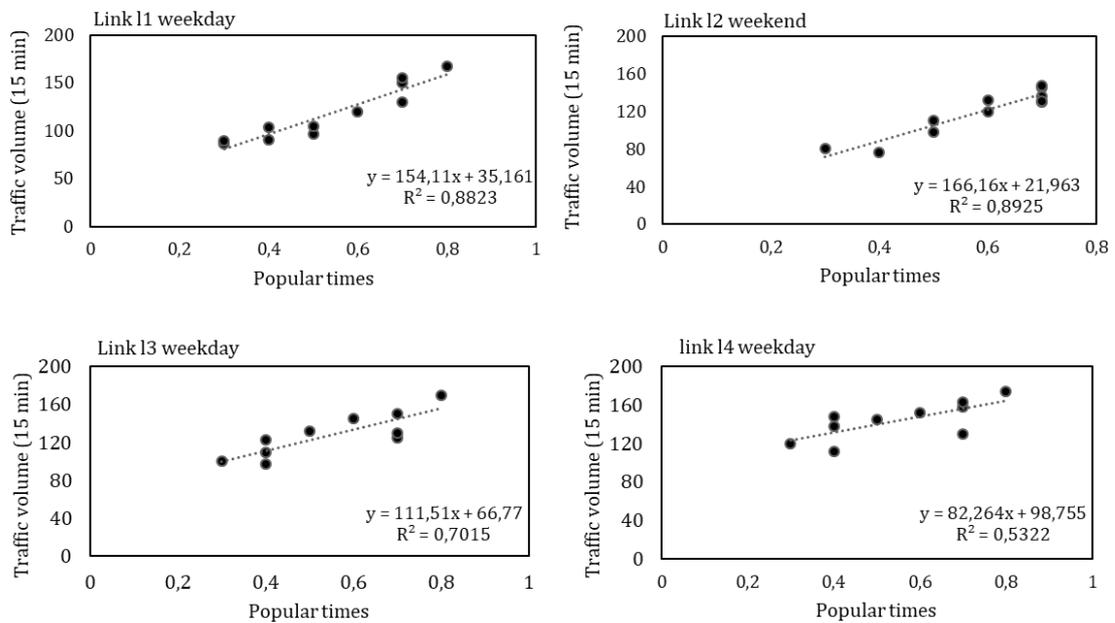


FIGURE 5 Linear regressions between traffic volumes and Popular times

3.2.2. Emission Analysis

Figures 6 to 8 present the linear regressions between GPT and CO₂ and NO_x emissions for all the studied links. Links 11 and 12 presented the strongest correlations regarding emissions for both CO₂ and NO_x emissions. GPT can explain 94% and 76% of CO₂ and NO_x variability, respectively, in link 11, and 82% and 67%, respectively, in 12. Regarding the links 13 and 14, we can observe weaker relationships. However, GPT is still able to explain around 59% of their variability in link 13, and 53% and 55% of CO₂ and NO_x variability, respectively, in 14. For links 15 and 16, GPT describes up to 52% of CO₂ and 47% of NO_x emissions variability. Regarding the study area in Badajoz, which comprises links 17 and 18, GPT can explain up to 62% and 56% of CO₂ and NO_x variability, respectively. Results on correlations with NO_x emissions are rather weaker, when compared with CO₂, because the emissions of NO_x are strongly affected by drivers' behaviour. Overall, these results reinforce that GPT can be used for estimating emissions or at least for minimizing the error in its estimation.

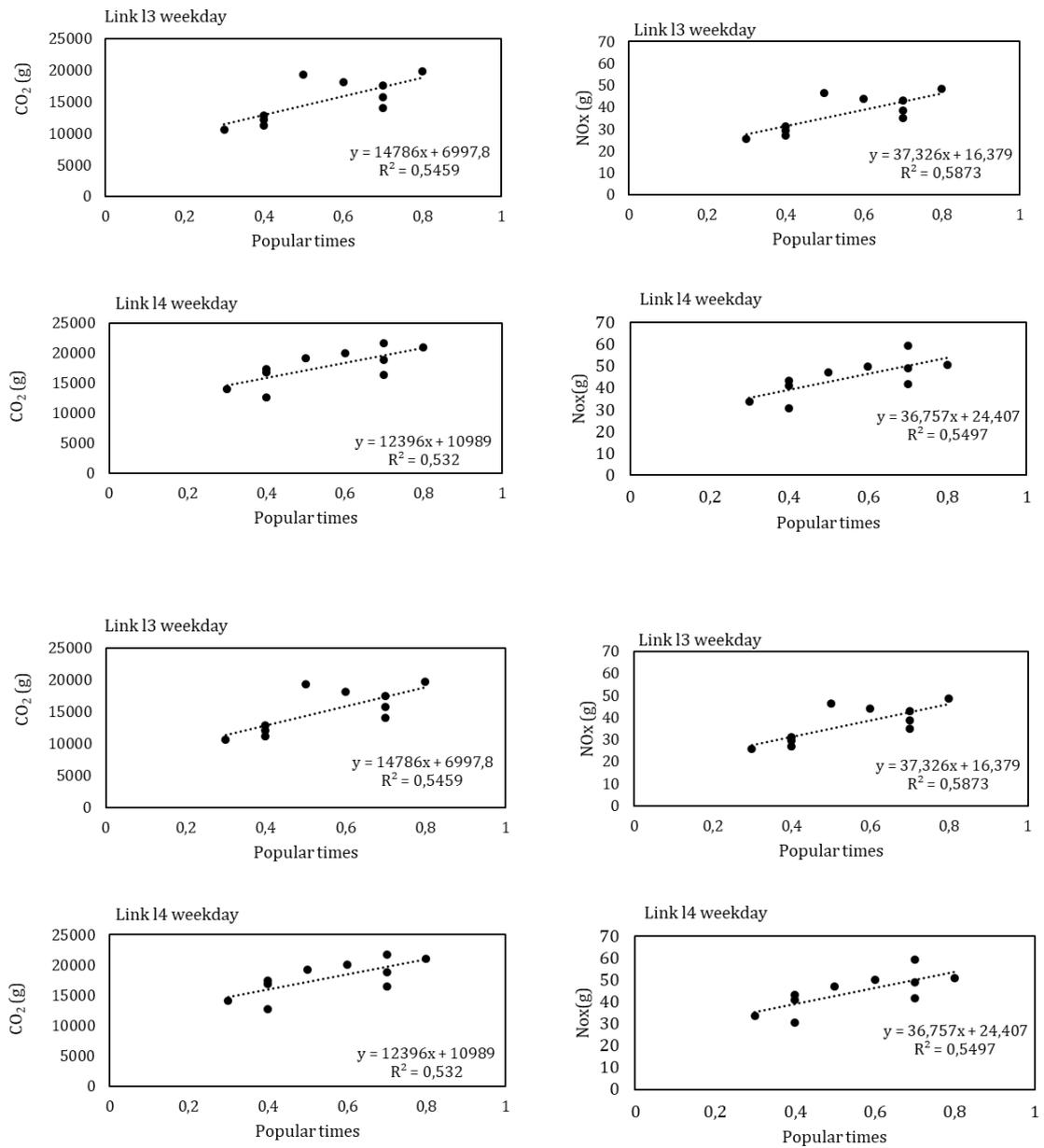


FIGURE 6 Linear regressions between CO₂ (left) and NO_x (right) emissions and Popular times for Aveiro Shopping Center.

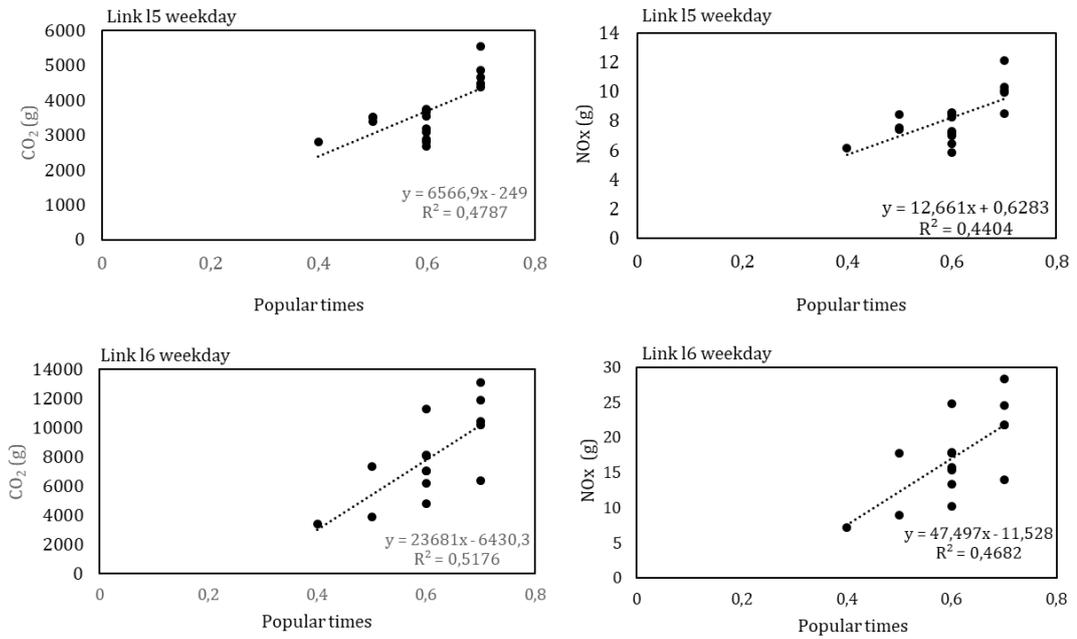


Figure 7 Linear regressions between system, O₂ (left) and NO_x (right) emissions and Popular times for Glicínias Plaza Shopping Center (Aveiro, PT)

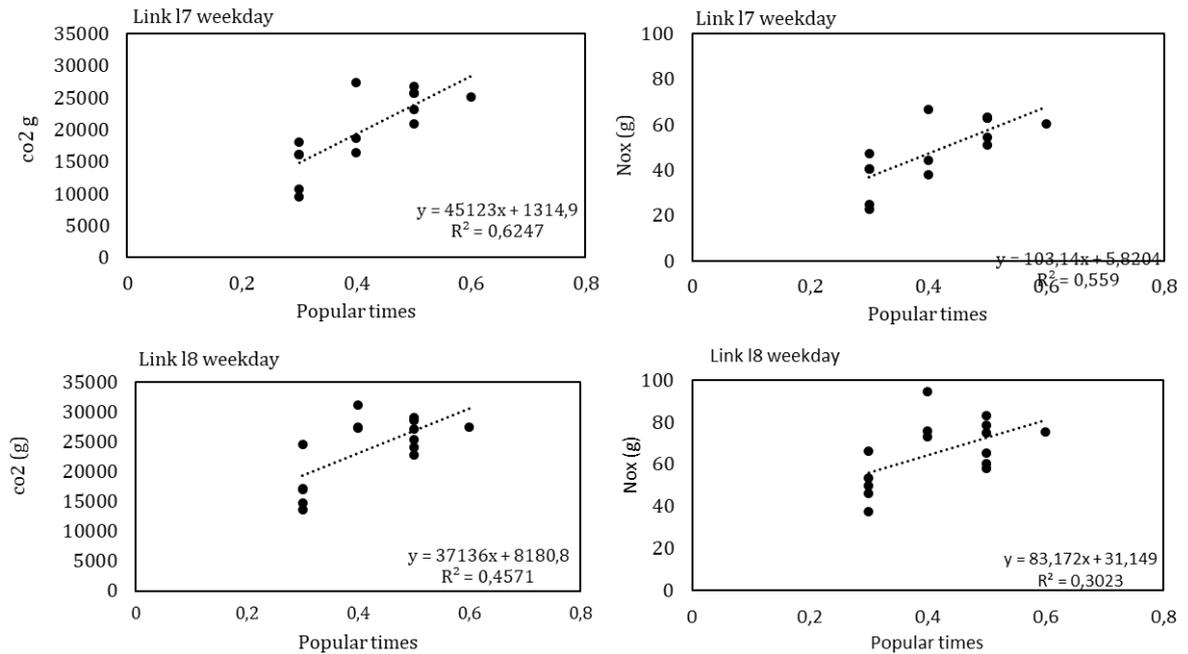


Figure 8 Linear regressions between CO₂ and NO_x emissions and Popular times for Hypermarket (Badajoz, SP).

3.2.3 Noise Analysis

We examined the relationships between noise and GPT by combining the line noise sources from both directions of each street (Figure 9). As in the previous analyses, linear regression models are only valid for the range of GPT values that were observed during the fieldwork. However, for these intervals the GPT

can reasonably explain the variability in the traffic noise levels in the studied links. In links 11 and 12, GPT can explain on weekends and on weekdays 85% and 75% of noise levels variability, respectively. Due to high demand on weekends (also observed in nearby commercial areas), an increase of 0.49 dBA was observed on weekdays and 0.71 dBA on weekends per a decimal increase GPT. In links 15 and 16 GPT explains 68% of noise variability. In links 17 and 18, GPT can justify only 38% of noise emissions variation. This lower value may be related to a weaker relation between noise and traffic volume combined with a high dispersion of average speed.

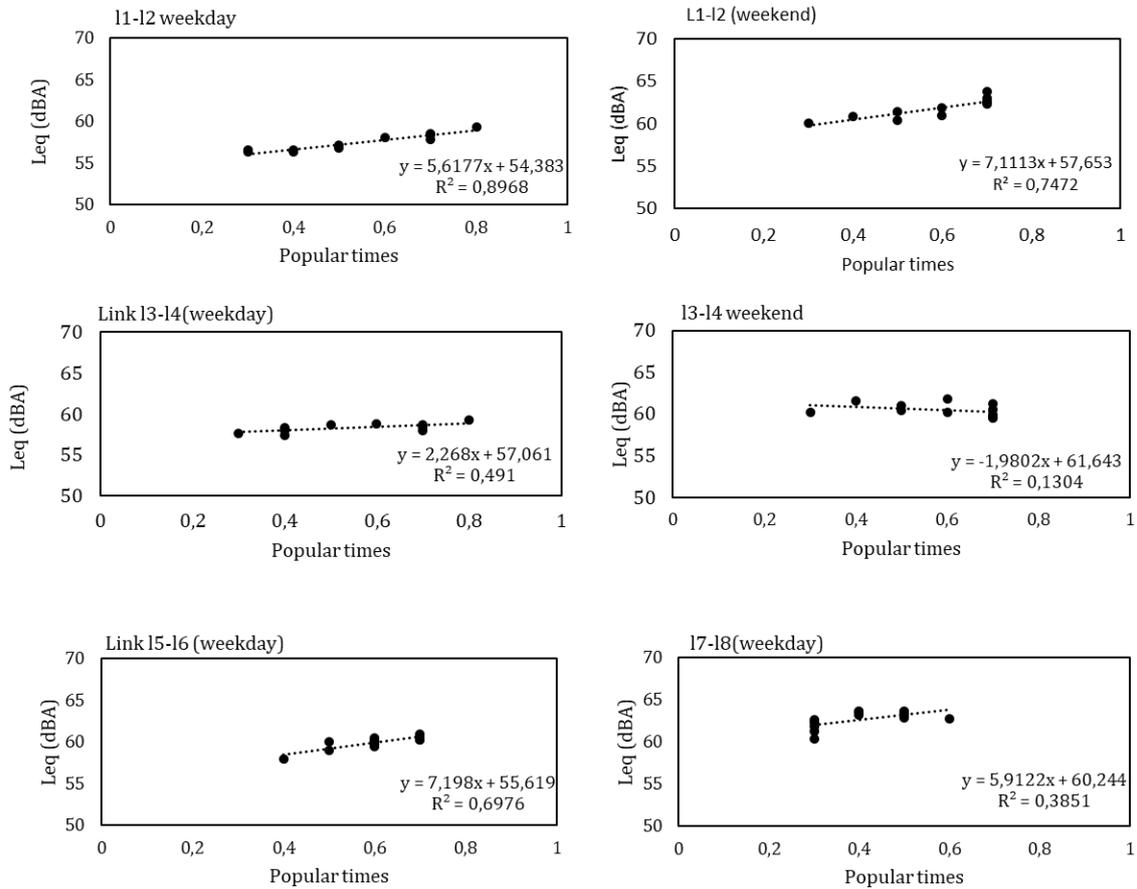


Figure 9 Linear regressions between noise estimation and Popular times.

3.3. Comparative Analysis

In this section, we perform a comparative analysis based on data collection during different periods of the year. More specifically for links 11 to 16, we repeated the data collection procedure during winter and summer (Figure 10).

In links 11 and 12, the developed linear models present a similar pattern. However, the slope of the regression line is higher in winter suggesting that the commercial area has a more relevant impact in this period in urban traffic. Once more, in such links, GPT has greater ability to predict variability in traffic conditions than in links 13 and 14, which may be due to a higher percentage of drivers with distinct origin and destinations than the commercial areas. Regarding links 15 and 16 (study area b), the results may be explained by the fact that both links are used for different trip destinations and there is a significant seasonal demand variation implying that the same scale (GPT) should not be used for different seasons.

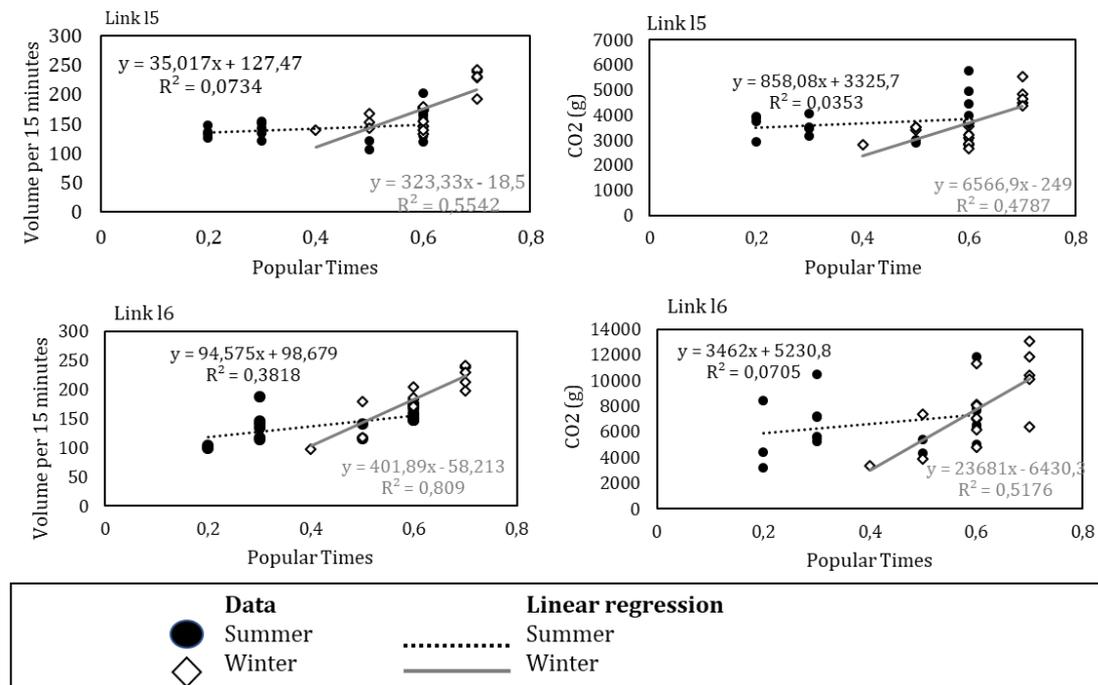


Figure 10 Seasonal comparison of the studied areas.

4. Conclusions

This study assessed the potential of using GPT to predict traffic volumes and traffic-related externalities. Three different study areas and 8 links near to important shopping areas in Portugal and Spain have been assessed during weekdays and weekends in different seasons. Linear regression models were developed to fit empirical and model-based derived data related to road traffic, NO_x and CO₂ emissions, and noise levels with GPT. The results suggest that for specific time periods it is possible to establish clear relationships between GPT and traffic volumes (up to 90%), CO₂ emissions (up to 98%) and noise levels (up to 85%). Results also show that GPT demonstrated lower capability to explain NO_x emissions due to high variability in driving behaviour and high dependence of individual acceleration and deceleration patterns under free flow regime. However, in shorter links such as 11, GPT can justify up to 76% of NO_x emissions variability.

Regarding the main question of the paper “Can web based information of business popularity be a reliable source of Information for traffic management?” the answer is affirmative but under a considerable set of restrictions. While in some links we found that the models can be applied in different seasons of the year with minor deviations, in others the regressions coefficients depend on the season or the day. This fact suggests that the relative scale of the GPT is not a uniform approach for different periods of the year and there is a considerable seasonal variability in the OD matrix of road users of the assessed road links. Before the implementation of a monitoring system based on this type of information, it will be necessary to collect data over an extended period or to have access to similar data sets e.g., based on absolute values for urban zones on open data platforms. This type of information would be especially useful in cities where there is limited funding to monitoring traffic conditions through traditional traffic monitoring systems and environmental sensors. Furthermore, leisure trips are harder to predict and this information can contribute to adjust spatio-temporal information of urban OD matrices with high accuracy.

This type of information and crowdsourcing information may be considered an asset to anticipate in advance potential congestion solutions due to high levels of popularity. Moreover, it may allow the optimization of intelligent transport systems such as partial-metering strategy or dynamic traffic lights. Finally, the collected environmental information can be included in environmental information systems and real-time link-based information to be included in eco-routing platforms.

Future work will contain the development of a global model that could be used to estimate traffic volumes, CO₂ and NO_x emissions, and noise in links near commercial areas using GPT as predictive

variable. In addition, integrated analysis is going to be conducted based on street networks and not on separated links.

Acknowledgments

The authors acknowledge the support of the following projects: UID/EMS/00481/2019-FCT - Fundação para a Ciência e a Tecnologia (FCT); CENTRO-01-0145-FEDER-022083 - Centro2020 Regional Operational Programme, under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund; MobiWise (P2020 SAICTPAC/0011/2015); DICA-VE (POCI-01-0145-FEDER-029463), Driving2Driverless (POCI-01-0145-FEDER-031923) and InFLOWence (POCI-01-0145-FEDER-029679) which were approved by the Executive Committee of the Competitiveness and Internationalization Operational Program in its ERDF component, by the Portuguese Foundation for Science and Technology, I.P. The cooperation of Toyota Caetano Auto, S.A. (that allowed the use of Toyota vehicles) is also appreciated.

References

1. Bandeira, J. M., Coelho, Sá, M. E., Tavares, R., & Borrego, C. (2011). Impact of land use on urban mobility patterns, emissions and air quality in a Portuguese medium-sized city. *Science of The Total Environment*, 409(6), 1154–1163. <https://doi.org/10.1016/j.scitotenv.2010.12.008>
2. Bandeira, J. M., Fernandes, P., Fontes, T., Pereira, S. R., Khattak, A. J., & Coelho, M. C. (2018). Exploring multiple eco-routing guidance strategies in a commuting corridor. *International Journal of Sustainable Transportation*, 12(1), 53–65. <https://doi.org/10.1080/15568318.2017.1328545>
3. Chaniotakis, E., Antoniou, C., Grau, J. M. S., & Dimitriou L. (Eds.). (2016, November). Can Social Media data augment travel demand survey data?. *Proceedings of the 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil (pp. 1642–1647). IEEE. doi:10.1109/ITSC.2016.7795778.
4. Cho, E., Myers, S. A., & Leskovec, J. (Eds.). (2011, August). Friendship and mobility. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1082-1090). New York, USA. doi:10.1145/2020408.2020579.
5. Coelho, M. C., Frey, H. C., Roupail, N. M., Zhai, H., & Pelkmans, L. (2009). Assessing methods for comparing emissions from gasoline and diesel light-duty vehicles based on microscale measurements. *Transportation Research Part D: Transport and Environment*, 14(2), 91-99.
6. Downing, R., Skabardonis, A., & Alexiadis, V. (2004). *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. (Report No. FHWA-HRT-04-040). FHWA, U.S. Department of Transportation.
7. EC Commission - *Strategy for low-emission mobility from* <https://ec.europa.eu/transport/themes/strategies/news/>
8. Emisia SA (2014). COPERT DATA (Version 5) [Computer Software], Thessaloniki, Emisia SA.
9. INE, 2017 Statistes Portugal web portal <https://ine.pt>
10. INE, 2018 Spanish Statistics National Institute <https://www.ine.es/>
11. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer. doi:10.1007/978-1-4614-7138-7_1.
12. Jin, P., Cebelak, M., Yang, F., Zhang, J., Walton, C., & Ran, B. (n.d.). Location-Based Social Networking Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2430, 72–82. doi:10.3141/2430-08.
13. Kephelopoulos, S., Paviotti, M., & Anfosso-Lédée, F.(2012). *Common Noise Assessment Methods in Europe (CNOSSOS-EU)*. Publications Office of the European Union, Luxembourg, 201e.
14. Lee, J. H., Gao, S., Janowicz, K., & Goulias, K. G. (2015, July). Can Twitter data be used to validate travel demand models? Paper presented at the *14th International Conference on Travel Behaviour Research*, Windsor, UK.
15. Majumdar, S. R. (2017). *The case of public involvement in transportation planning using social media. Case Studies on Transport Policy*, 5(1), 121–133. doi:10.1016/j.cstp.2016.11.002.
16. Nair DJ, Gilles F, Chand S, Saxena N, Dixit V. (2019) Characterizing multicity urban traffic conditions using crowdsourced data. Wang Y, editor. *PLoS One*. Available from: <http://dx.plos.org/10.1371/journal.pone.0212845>

-
17. Ni, M., He, Q., & Gao, J. (Eds.). (2014, January). Using Social Media to Predict Traffic Flow under Special Event Conditions. *Proceedings of the 93th Annual Meeting on Transportation Research Board (TRB)*, Washington D.C., USA. Retrieved from <https://trid.trb.org/view.aspx?id=1288769>.
 18. Pereira, F.C., Rodrigues, F., Polisciuc, E., & Ben-Akiva, M. (2015). Why so many people? Explaining Nonhabitual Transport Overcrowding with Internet Data. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1370–1379. doi:10.1109/TITS.2014.2368119..
 19. Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, 197–211. doi:10.1016/j.trc.2016.12.008.
 20. Ribeiro, S. S., Davis, C. A., Oliveira, D. R. R., Meira, W., Gonçalves, T. S., & Pappa, G. L. (2012, September). Traffic observatory. In *Proceedings of the 5th International Workshop on Location-Based Social Networks - LBSN '12*, New York. doi:10.1145/2442796.2442800.
 21. Santos, G. (2017). Road fuel taxes in Europe: *Do they internalize road transport externalities?*. *Transport Policy*, 53, 120–134. doi:10.1016/j.tranpol.2016.09.009.
 22. Serna, A., Gerrikagoitia, J.K., Bernabé, U., & Ruiz, T. (2017). Sustainability analysis on Urban Mobility based on Social Media content. *Transportation Research Procedia*, 24, 1-8. doi:10.1016/j.trpro.2017.05.059.
 23. Tafidis, P., Teixeira, J., Bahmankhah, B., Macedo, E., Coelho, M. C., & Bandeira, J. (2017, June). Exploring crowdsourcing information to predict traffic-related impacts. Paper presented at the *17th International Conference on Environment and Electrical Engineering (IEEE)*, Milan, Italy. doi: 10.1109/EEEIC.2017.7977595.
 24. Tasse, D., & Hong J. I. (Eds.). (2014, August). Using Social Media Data to Understand Cities. *Proceedings of NSF Workshop on Big Data and Urban Informatics*, Chicago. Retrieved from <http://repository.cmu.edu/hcii>.
 25. Teixeira, J., Fernandes, P., Bandeira, J., & Coelho, M. C. (2017, January). Information Management for Smart and Sustainable Mobility. Poster session presented at *96th Annual Meeting Transportation Research Board (TRB)*, Washington D.C., USA.
 26. Tian, Y., Zmud, M., Chiu, Y. C., Carey, D., Dale, J., Smarda, D., Lehr, R., & James, R. (2016, January). Quality assessment of social media traffic reports – a field study in Austin, Texas. Paper presented at the *95th Annual Meeting on Transportation Research Board (TRB)*, Washington D.C., USA.
 27. Tostes, A. I. J., Duarte-Figueiredo, F., Assunção, R., Salles, J., & Loureiro A. A. F. (Eds). (2013, August). From data to knowledge: city wide traffic flows analysis and prediction using Bing Maps. *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, IL, USA*.
 28. Turner, S. M., Eisele, W. L., and Benz, R. J. (1998). Travel Time Data Collection handbook. (Report No. FHWA-PL-98-035). FHWA, U.S. Department of Transportation.
 29. United States Environmental Protection Agency. (2014). Evaporative Emissions from On-road Vehicles in MOVES2014 (Report No. EPA-420-R-14014). Ann Arbor, MI, United States, Environmental Protection Agency (US).
 30. Yang, F., Jin, P. J., Wan, X., & Ran, B. (Eds.). (2014, January). Dynamic Origin-Destination Travel Demand Estimation using Location Based Social Networking Data. *Proceedings of the 93th Annual Meeting on Transportation Research Board (TRB)*, Washington D.C., USA. Retrieved from <https://pdfs.semanticscholar.org/af46/0f908bbd51388892ee7b45fa6a33ef343a4c.pdf>.