Exploring Relationship Between Taxi Volume and Flue Gases' Concentrations

Marco Veloso
Centro de Informática e Sistemas
Universidade de Coimbra
Portugal
and
Escola Superior de Tecnologia e
Gestão de Oliveira do Hospital
Portugal
mveloso@dei.uc.pt

Santi Phithakkitnukoon
Computing Department
The Open University
United Kingdom
santi.phi@open.ac.uk

Carlos Bento
Centro de Informática e Sistemas
Universidade de Coimbra
Portugal
bento@dei.uc.pt

UbiComp'13 Adjunct, Sept 8-12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2139-6/13/09...$15.00.

Abstract
With the rapid increase in size and population of urban areas, it becomes important to understand urban environmental influencers so that better informed decisions can be made for more sustainable urban environments. Taxis represent one of the urban dynamics from which city planners can gain a better understanding of urban mobility as well as its relationship with other environmental elements. In this work, an analysis of the relationship between flue gases' concentrations (represented by nitrogen dioxide) and taxi volume in Lisbon, Portugal was carried out from which a strong correlation between the two was observed. Based on four months of data, we found that the flue gases' concentrations varied with taxi volume and in particular, taxi volume can be used to estimate the change in flue gases' concentrations of the next hour.

Author Keywords
Urban mobility, taxi-GPS traces, flue gases' concentrations, spatiotemporal analysis, time series analysis and linear regression.

ACM Classification Keywords
I.5.2. Patter Recognition: Pattern analysis.

General Terms

Algorithms.

Introduction

Automobile is one of the major sources of toxic compounds that are present in combustion gases that negatively impact the health of urban inhabitants. There is a need to address this issue today while low-carbon transport systems (which is a promising solution) are still being developed. Understanding of gas emission patterns and ability to estimate their concentrations in urban areas are thus essential in order to mitigate the problem.

Today, taxis in various cities are equipped with GPS to improve their services with a better dispatching system. By taking the opportunistic sensing approach, we used GPS traces of taxis in the city of Lisbon, Portugal to explore the relationship between taxi mobility patterns and level of concentration of flue gases.

Related Work

Mining taxi trajectories has recently attracted much attention. Taxi-GSP traces have been used in a number of studies to develop better solutions and services in urban areas such as estimating optimal driving paths [1-3], predicting next taxi pick-up locations [4-7], modeling driving strategies to improve taxi's profit [7-8], identifying flaws and possible improvements in urban planning [9], and developing models for urban mobility, social functions, and dynamics between the different city's areas [10-11].

Yuan et al. [1] present the T-Drive system that identifies optimal route for a given destination and departure time. Zheng et al. [2] describe a three-layer architecture using the landmark graph to model knowledge of taxi drivers. Ziebart et al. [3] present a decision-modeling framework for probabilistic reasoning from observed context-sensitive actions. The model is able to make decisions regarding intersections, route, and destination prediction given partially traveled routes.

Yuan et al. [4] develop a recommender system for both taxi drivers and passengers that takes into account the passengers' mobility patterns and taxi drivers' pick-up traces. Phithakkitnukoon et al. [5] present a model for predicting the number of vacant taxis for a given area of the city based on the naïve Bayesian classier with their developed error-based learning algorithm and a mechanism for detecting adequacy of historical data. Liu et al. [6] classify taxi drivers according to their income. They observe that top drivers operate in a number of different zones while maintaining exceptional balance between taxi demand and traffic conditions.

Ge et al. [7] present an approach for extracting energy-efficient transportation patterns from taxi traces and use it to develop a recommender system for pick-up locations and a sequence of waiting locations for a taxi driver. Zheng et al. [9] identify flawed urban planning in region pairs with traffic problems and the linking structure among these regions through their analysis of taxi traces. Qi et al. [10] investigate the relationship between regional pick-up and drop-off characteristics of taxis and social function of city regions. They develop a simple classification method to recognize regions' social areas. Veloso et al. [11] explore the relationship between taxi volume and mobile phone activity. They observe a strong relation between them i.e., the amount of mobile phone calls is strongly correlated with the taxi volume of the previous two hours. Moreover, the level of inter-predictability varied across different time of the day.

In addition to the dynamic in vehicular network, there are work focusing the study of flue gases' fluxes, and the development of environment data sensing methods [12-16].

Velasco et al. [12] use an eddy covariance (EC) flux system to obtain direct measurements of CO2 emissions in Mexico City. The analysis shows a clear diurnal pattern with the highest emissions during the morning and the lowest emissions during nighttime. The measured CO2 fluxes are closely correlated to traffic patterns. Liu et al. [13] apply a similar methodology to the city of Beijing, China, collecting data during a four-year period, with similar results. Daily and weekly cycles are observed, with strong dependency with road traffic. Zavala et al. [14] use a mobile laboratory to measure on-road vehicle emission ratios in Mexico City. The authors show that flue gases' emissions are strongly related with driving behaviors.

Mao et al. [15] present *CitySee*, a real-time CO2-monitoring system using wireless sensor networks for an urban area, in Wuxi, China, proposing a low-cost sensor deployment strategy. Hu et al. [16] propose a vehicular sensing system to collect CO2 concentration in urban areas, based on GSM short messages and GPS information of vehicles. Vehicles are used as carriers of sensing devices to monitor CO2 concentrations while driving through the city. The concept is tested using the ZigBee-based.

Datasets

Taxi volume

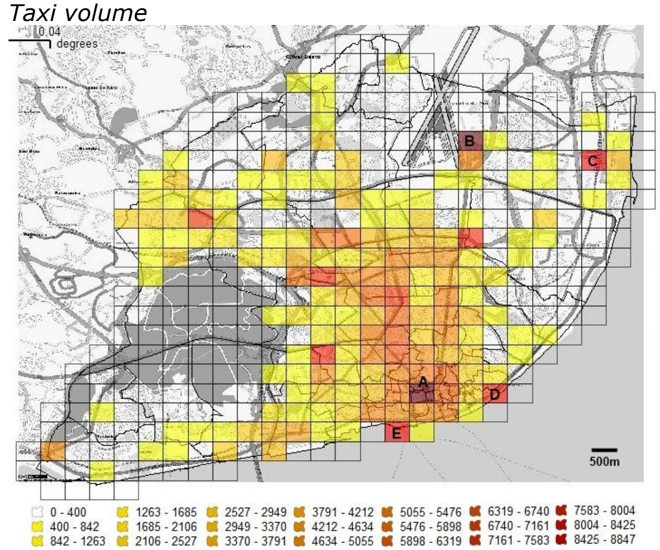

Figure 1. Spatial distribution of taxi volume (number of pick-ups).

Our taxi dataset was provided by GeoTaxi[1], a company that focuses on software development for fleet management, and holds about 20% of the taxi market share in Portugal. The dataset was composed of around 10 million taxi-GPS location points and collected from 230 taxis. Along with the GPS location (latitude, longitude) information, it reported speed, bearing, and occupancy status of the taxi. The amount of pick-ups and drop-offs were inferred, which accounted for 177,169 distinct trips. The number of pick-ups was termed taxi volume. A data cleaning process was applied to remove trips with less than 200m and more

[1] Geotaxi. http://www.geotaxi.com/ .

than 30km (the realistic longest trips from one side of the city to the other could be around 22km), and less than a minute and longer than three hours.

The overall taxi volume's spatial distribution in Lisbon is shown in Fig. 1 (on 500x500m2-grid cells), where the number of pick-ups on each cell during the period under study is represented by a color scale (red corresponds to cells with a higher number of pick-ups). Some major locations are identified, such as city downtown (A), airport (B), train stations (C, D) and ferry dock (E). Different public transportation modalities (e.g., airport, train, ferry, bus) are well connected through taxi services.

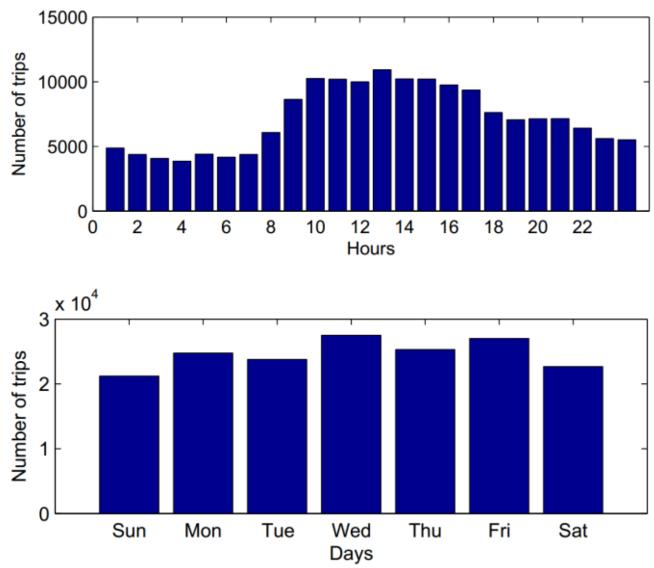

Figure 2. Taxi volume variation according to hours of day (top) and days of week (bottom).

Taxi volume varies in time and space. Fig. 2 presents temporal variation of the taxi services. As expected, the taxi service variation follows the business hours. It gradually increases in from 5am, reaches the maximum between 11am and 1pm, and slowly drops down in the afternoon. By the same token, there are more taxi services in working days than in weekends. On average, we observed a reduction of taxi volume of about 46.7% at night (from 10pm to 7am) and 13.6% on weekends.

Flue gases

The flue gases' dataset was provided by both the '*Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo*' (CCDR-LVT)[2], and the '*Agência Portuguesa do Ambiente*'[3], which are governmental institutions responsible for monitoring atmospheric pollutants. The dataset was composed of hourly readings of different gases concentrations on seven monitoring stations (shown in Fig. 3). Every station monitors nitrogen oxide (NOx), nitrogen monoxide (NO), nitrogen dioxide (NO2), and carbon monoxide (CO), measured in μg/m^3, which are exhaust combustion gases, also called flue gases. In our preliminary analysis, only nitrogen dioxide was considered in this paper.

Although the current work focuses on a common window of observation from September to December 2009, the flue gases' database contains data from 2008 to 2011, which is explored in this section.

[2] CCDR-LVT. http://www.ccdr-lvt.pt/pt/ .

[3] Agência Portuguesa do Ambiente. http://www.qualar.org .

The monitoring stations were classified into two groups: traffic stations (D and E, Fig. 3) and background stations (A, B, C, F, G, Fig 3). The traffic stations are located near traffic roads while the background stations are located away from main roads. On average, traffic station perceives higher concentrations of flue gases (65.3 μg/m^3 for NO2) than background station (36.5 μg/m^3 for NO2), which is in line with Ndoke and Jimoh [18] who observed that concentrations of flue gases decreased as when moving away from the roads.

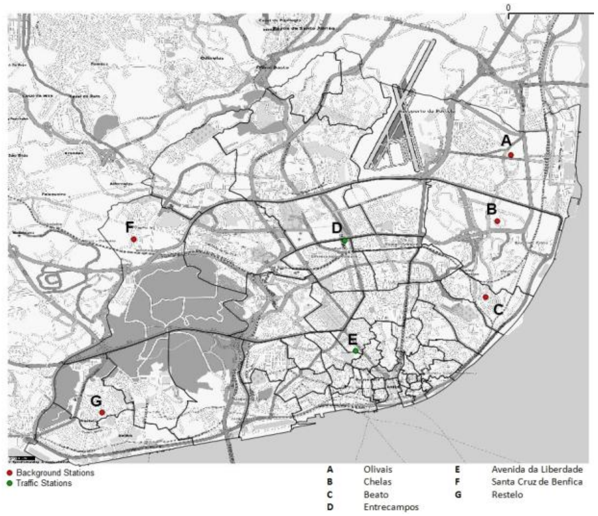

Figure 3. Locations of monitoring stations.

Two daily peaks of gas concentration, which is related to traffic congestion were also observed in [17]. The morning peak quickly increases in from 5am, reaches the maximum around 8am and quickly drops down, corresponding to the inbound traffic to the city. In the afternoon, gas concentration gradually rises around 3pm and reaches the maximum around 7pm and slowly drops down, corresponding to the outbound traffic from the city. The rate of dispersion of gases is affected by temperature. Gases react to heat by expanding their volume as higher temperature increases molecules' speed, and hence disperses more quickly. When facing cold, gases respond by contracting and by dispersing slowly [19]. On average, we observed a reduction of flue gases' concentrations of about 19.1% at night (from 10pm to 7am) and 23.1% on weekends.

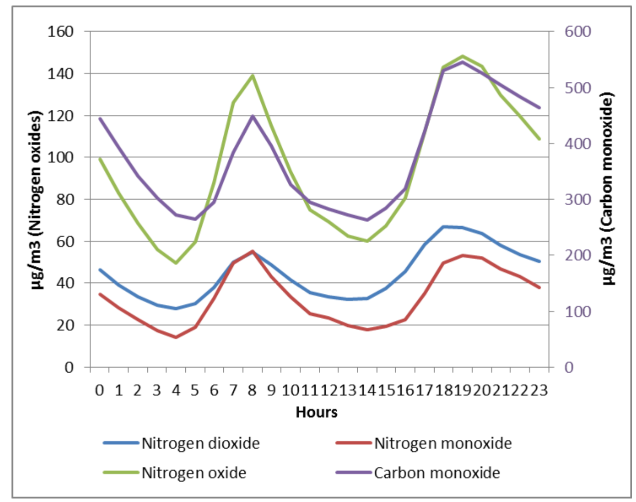

Figure 4. shows the average variation of flue gases over the course of a day.

Likewise, warmer months (June, July and August) have in average lower gases concentrations (25.7 μg/m^3 for NO2) than colder months (44.8 μg/m^3 for NO2 on October, November and December), which can be observed in Fig. 5.

In warmer months, the morning peak reaches higher values of gas concentrations than the afternoon peak (37.8 μg/m^3 against 29.8 μg/m^3 for NO2), while in colder months the afternoon peak attains higher values than the morning peak (67.8 μg/m^3 against 55.9 μg/m^3 for NO2). Moreover, there is a narrower gap between the maximum and minimum average concentrations of flue gases in warmer months (23.3 μg/m^3 for NO2) when compared with colder months (40.2 μg/m^3 for NO2). Similar patterns were observed when exploring data from different years (from 2008 to 2011).

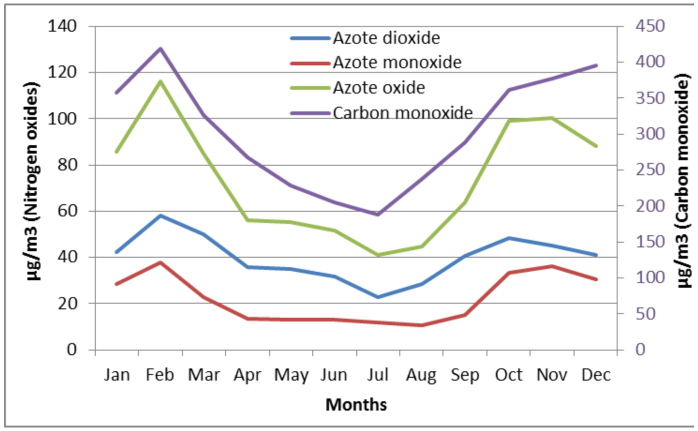

Figure 5. Average variation of flue gases across every month in 2009.

Analysis and results

To explore the relationship between taxi volume and flue gases' concentrations (represented by nitrogen dioxide in this study) we extracted data as a hourly aggregated time series, normalized to [0, 1]. We overlaid both time series on the same plot as shown in Fig. 6 and observed similar temporal patterns. Both exhibited daily cycles, although taxi volume shows a more regular pattern.

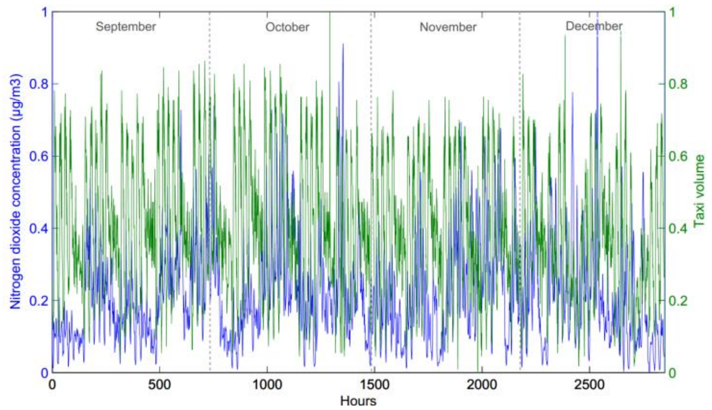

Figure 6. Normalized time series of taxi volume (green) and nitrogen dioxide (blue) over four months of observation.

To quantify the difference between these two time series, we computed the Euclidean distance (ED) as follows:

$$ED_i = \sqrt{(g_i - t_i)^2} = |g_i - t_i|$$

where g_i represents the nitrogen dioxide, concentrations at hour *i* and t_i denotes taxi volume at hour *i*. Hence, $G = \{g_1, g_2, \ldots, g_n\}$ and $T = \{t1, t_2, \ldots, t_n\}$ represent the normalized time series of nitrogen dioxide concentrations and taxi volume of length *n*, respectively.

Euclidean distance of these time series was 0.27928, and hourly distances are shown in Fig. 7. Higher values of ED were observed mostly between 9am and 3pm, a period where the concentration of nitrogen dioxide decreased while the taxi activity stayed high.

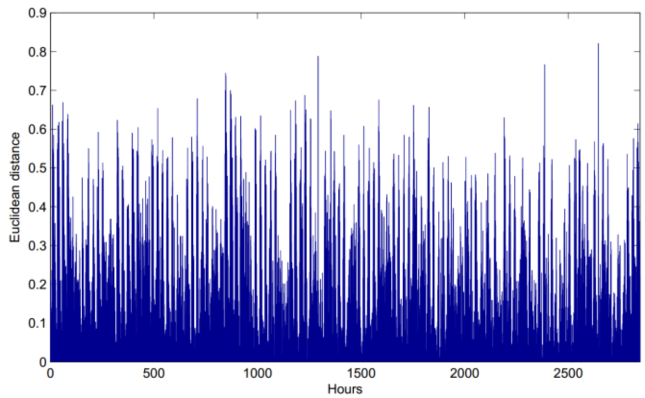

Figure 7. Hourly Euclidean distance of the normalized time series of nitrogen dioxide concentrations and taxi volume.

To further explore in terms of predictability between the two data sources, we employed the coefficient of determination or R^2 (that is widely used for regression analysis) to measure the interdependency between them. The coefficient of determination, or R^2, can be calculated as:

$$R^2 = \frac{\sum_i (y_i - \bar{y})^2 - \sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean and $\hat{y}$ denotes the predicted value of y (i.e., $\hat{y}_i = a + bx_i + \varepsilon_i$). The R^2 value between the two time series was found to be 0.86833, which represents a significant interdependency.

Furthermore, we observed daily and weekly cycles. We observed highest similarities between these time series was during weekdays (R^2 = 0.870014) and active hours (8am to 10pm, R^2 = 0.80723). However, low R^2-value observed between taxi speed and nitrogen dioxide concentrations.

To further investigate the predictability that one data source had on the other, was used a time shifting. For example, one-hour lag of *X* yields a high R^2 value with *Y* implies that *X* is likely a one-hour predictor of *Y*, i.e., the variation in values of *X* suggest a similar variation in values of *Y* of the next hour. By fixing nitrogen dioxide time series and shifting taxi time series between -5 hours to +5 hours (e.g., -5 hours of time shift means considering nitrogen dioxide data at time *t* against taxi data at time *t*-5 hours), the highest R^2-value was found at the time shift of -1 hour (R^2 = 0.871251.)

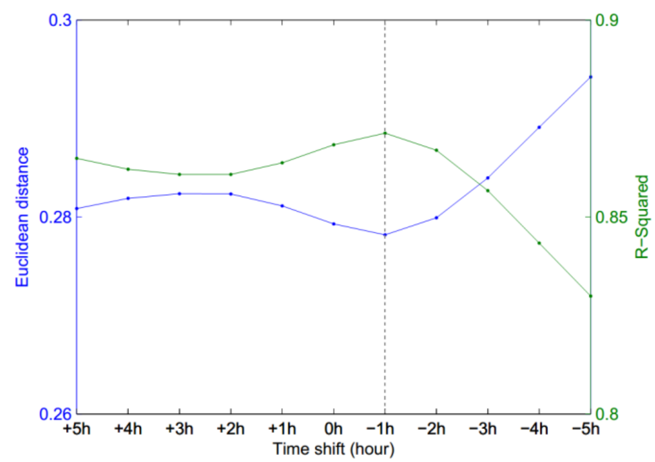

Figure 8. Euclidean distance and R^2-values from the sliding windows between azote dioxide and taxi data.

As shown in Fig. 8, at time shift of -1 hour the R^2 and Euclidean distance values were 0.871251 and 0.278177, respectively, which suggests that generally taxi volume is a 1-hour predictor of nitrogen dioxide concentration. In other words, the variation in the amount of taxis is an indicative variable for the nitrogen dioxide of the next hour. With 1-hour time shifting, Fig. 9 shows fitted linear equation, $y = p_1 x + p_2$, where p_1 = 0.12381, p_2 = 38.149, and R^2 = 0. 871251.

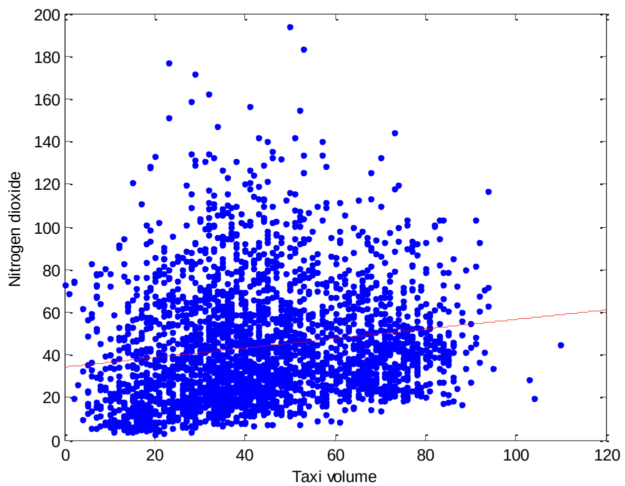

Figure 9. The fitted linear function of the taxi volume (at time t-1) against the Nitrogen dioxide concentrations (at time t) with R^2 = 0.871251.

The variation of nitrogen dioxide concentrations from warmer months to colder months suggests that the abovementioned relationship could vary throughout the year. To explore this, a time shift was used for each month individually. As shown in Table 1, the time shift decreases from warmer months to colder months. This is an indication that weather condition plays a part in the relationship between taxi volume and flue gases' concentrations and this among others will be further investigated in our future work.

Month	**Time Shift (h)**	**R^2**
September	-2	0,91432
October	-1	0,88321
November	0	0,91058
December	0	0,92350

Table 1. Predictability across the year (from September to December 2009).

Conclusions

In this work, we explored a relationship between the taxi volume and flue gases' concentrations in Lisbon, Portugal. Using four months of data, we observed that taxi volume can.be used to estimate the concentration of nitrogen dioxide in the next hour. As weather condition has shown some effect on gas concentration, our future work will explore this effect along with other influential factors.

References

[1] Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X. and Huang, Y. T-Drive: Driving Directions Based on Taxi Trajectories. In Proc. ACM SIGSPATIAL GIS 2010, Association for Computing Machinery, Inc. 1 (2010), 99-108.

[2] Zheng, Y., Yuan, J., Xie, W., Xie, X., Sun and G. Drive Smartly as a Taxi Driver. In 7th Int. Conference on Ubiquitous Intelligence & Computing and 7th Int. Conference on Autonomic & Trusted Computing (UIC/ATC) (2010), 484-486.

[3] Ziebart, B.D., Maas, A.L., Dey, A.K. and Bagnell, J.A. Navigate like a cabbie: probabilistic reasoning from

observed context-aware behavior. In UbiComp '08: Proc. of the 10th int. conf. on Ubiquitous computing, New York, NY, USA, ACM (2008), 322-331.

[4] Yuan, J., Zheng, Y., Zhang, L., Xie, X. and Sun, G. Where to Find My Next Passenger? In 13th ACM Int. Conf. on Ubiquitous Computing (UbiComp 2011), China (2011).

[5] Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A. and Ratti, C. Taxi-Aware Map: Identifying and predicting vacant taxis in the city. In Proc. AmI 2010, First International Joint Conference on Ambient Intelligence (2010), 86-95.

[6] Liu, L., Andris, C., Biderman, A., Ratti, C. Uncovering cabdrivers' behavior patterns from their digital traces.Iin Computers, Environment and Urban Systems, (2010).

[7] Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., Pazzani, M. J. An Energy-Efficient Mobile Recommender System. In Proc. KDD 2010, ACM Press (2010), 899-908.

[8] Liu, L., Andris, C., Bidderman, A. and Ratti, C. Revealing taxi drivers mobility intelligence through his trace. In Movement-Aware Applications for Sustainable Mobility: Technologies and Approaches, (2010), 105-120.

[9] Zheng, Y., Liu, Y., Yuan, J. and Xie, X. Urban Computing with Taxicabs. In 13th ACM Int. Conference on Ubiquitous Computing (UbiComp 2011), (2011).

[10] Qi, G., Li, X., Li, S., Pan, G., Wang, Z. and Zhang, D. Measuring Social Functions of City Regions from Large-scale Taxi Behaviors. In PerCom- Workshops, (2011), 21-25.

[11] Veloso, M., Phithakkitnukoon, S., Bento, C. Exploring the Relationship between Mobile Phone Call Intensity and Taxi Volume in Urban Area. In 15th IEEE Intelligent Transportation Systems Conference, (2012).

[12] Velasco, E., Pressly, S., Allwine, E., Westberg, H. and Lamb, B. Measurements of CO2 fluxes from the Mexico City urban landscape. In Atmospheric Environment 39, (2005), 7433-7446.

[13] Liu, H. Z., Feng, J. W., Jarvi, L. and Vesala, T. Eddy covariance measurements of CO2 and energy fluxes in the city of Beijing. In Atmospheric Chemistry and Physics 12, (2012), 7677-7704.

[14] Zavala, M., Herndon, S. C., Slott, R. S., Dunlea, E. J., Marr, L. C., Shorter, J. H., Zahniser, M., Knighton, W. B., Rogers, T. M., Kolb, C. E., Molina, L. T. and Molina, M. J. Characterization of on-road vehicle emissions in the Mexico City Metropolitan Area using a mobile laboratory in chase and fleet average measurement modes during the MCMA-2003 field campaign. in Journal of Atmospheric Chemistry and Physics 6, (2006), 5129-5142.

[15] Mao, X., Miao, X., He, Y., Zhu, T., Wang, J., Dong, W., Li, X. and Liu, Y. CitySee: Urban CO2 Monitoring with Sensors. In Annual IEEE Int. Conference on Computer Communications (IEEE INFOCOM), (2012).

[16] Hu, S., Wang, Y., Huang, C., Tseng, Y., Kuo, L. and Chen, C.Vehicular Sensing System for CO2 Monitoring Applications, (2009).

[17] CCDR-LVT. http://www.ccdr-lvt.pt/pt/ .

[18] Ndoke, P. N. and Jimoh, O. D. Impacts of Traffic Emission on Air Quality in a Developing City of Nigeria. In AU Journal of Technology, Bangkok, Thailand, Vol. 8, no. 4, (2005), 222-228.

[19] Beychok, M. R. Fundamentals of Stack Gas Dispersion. Publisher: Milton R. Beychok, 2005.