# Motivating car-sharing services open-data mandatory APIs

Andrea Trentini, Federico Losacco
Dipartimento di Informatica
Via Comelico, 39 - 20135 Milano (ITALY)
andrea.trentini@unimi.it, federico.losacco@studenti.unimi.it

Interoperabilità dei dati e modalità di autenticazione e accesso ai dati

**Abstract**

Modern countries urban traffic is nowadays fought by introducing regulations to reduce private traffic: tolls, dedicated lanes, narrow lanes, low speed limits, reduced parking availability, etc. Some help can come from car-sharing, i.e., pools of shared vehicles to be rented for short periods of time. The authors scraped car-sharing vendors' websites for a couple of years, made data uniform and then queried and graphed the dataset, a summary of results is presented.

Car-sharing vendors should publish - in open data, mandatorily, through standardized API - data about the state of their vehicle pool because this data can be used to analyse the overall traffic behaviour in town and to study impacts, effectiveness and costs.

**Keywords: Application Programming Interface, urban congestion, open data, public accountancy**

# 1 Introduction

Air pollution became a problem for human beings with the industrial revolution [SP12, LS13, SRS13]. Afterwards many governments begun legislating [UE13] to reduce industrial emissions. From then on, air pollution slowly began to decrease as new generations of technologies replaced older ones (Figure 1).

While traffic pollution is being reduced, many governments and administrations are now addressing the "congestion factor" by introducing regulations [Van09] to lower private traffic.

A few years ago vendors started providing "car-sharing". The rationale behind car-sharing is that: 1) parking space is saved since a single parked vehicle serves a lot of users, not a single one (as in the case of a private car); 2) because of the higher (more than a private car, but less than a taxicab) costs *per use*, many users, even if subscribed to the service, would use it wisely (maybe mixing it with standard public transport), thus reducing the overall "car mileage load" on the city; 3) it may help in further pollution lowering [FM11]. A fairly complete review on car-sharing studies can be read in [JC13].

GPSs and smartphones track down available cars and lead users to precise parking locations. Detailed location data are, of course, stored and secured into vendors' systems but some data are available on the web. Data can be used to analyse the overall traffic behaviour in town. Vendors do publish status data, but only the real-time situation and often masking the data behind "webstacles" [Tre14]. Given the usefulness of this data, the authors propose the initiation of a standardization and "open-data-ization" process. Governments should force vendors to publish data (both real-time and historical) through public, documented and well-defined APIs (Application Programming Interfaces).

# 2 The quest for data (*web-scraping*)

In Milan (ITALY) there are five car-sharing vendors: 1) Share'NGo (`http://www.sharengo.it`); 2) Twist-Car (`http://twistcar.it`, discontinued 17-Nov-2015); 3) Enjoy (`http://enjoy.eni.com`); 4) Car2Go (`http://www.car2go.com`); 5) GuidaMi (`http://www.guidami.net`).

GuidaMi (now GirACI) does not publish data about cars. The other vendors do publish data about car positions and availability, some of them (Car2go) even offer APIs to query data, with limitations (low query frequency, need for a security key, real-time data only). The authors scraped website data, using `wget` in shell scripts or small `python` programs periodically run on a server and stored data in a database.

# 3 Data analysis

The following analyses are based on data between September 2015 and March 2016, they show an evident daily and weekly patterns, i.e.: 1) every day is "cyclical", e.g., night differs from day and office hours are peaks; 2) there are "seasonal" (weekly) effects: a) Mon-Fri days are similar to one another; b) Saturday and Sunday are similar; c) Saturday and Sunday differs from Mon-Fri workdays 3) there are, of course, random fluctuations. To better analyse the in-week patterns data were divided into weeks and for every week Q-Q plots (Quantile-Quantile plots check if two samples are statistically similar) were created. A Q-Q plot for every day pair (Mon-Tue, Mon-Wed, Mon-Thu, etc.) in every week was generated, originating $\binom{7}{2} = 21$ Q-Q plots per week (examples in Figures 3 and 5). This procedure statistically confirmed the evident (and expected) pattern difference between Mon-Fri and Sat-Sun. Abbreviations: `H` = Holiday; `NH` = Non Holiday;

## 3.1 Available cars

The total number of free cars at a given time $t$ is the sum of parked cars. When this value is low it means that many cars are "taken", i.e., are in use. Figure 2 shows the usage for two **typical** days (from 00:00 to 23:59), one weekday and one weekend day. The "late Saturday" (Sunday early morning) usage is evident in the lower graph. Other notable remarks in Figure 2 and in general, on the whole dataset, are: 1) night-time (between 2AM and 7AM) is a peak of free cars; 2) the usage peak (least number of free cars) is between 6PM and 9PM; 3) morning (between 8AM and noon) usage is lesser than afternoon (noon to 7PM) usage; 4) night-time peak sports a shorter timespan (between 6AM and 8AM) w.r.t. the Monday-to-Friday night-time peaks; 5) there is a usage peak between midnight and 2AM, often more substantial than the afternoon peak.

In Figure 3 the Q-Q plots of two day pairs taken as example: 2015-10-05 against 2015-09-29 (Tue against Mon), near the $y = x$ line → statistically similar; 2015-11-22 against 2015-11-17 (Sun against Tue), far from the $y = x$ line → statistically different. I.e., (as expected!) users' behaviour on Sunday is different from a weekday.

## 3.2 Overall (average) distance from city centre

Notable remarks in Figure 4 (top is "Mon to Fri", bottom is "Weekend", they represent typical days from the dataset, X-axis is from 00:00 to 23:59) are: 1) "lung" effect': average distance decreases during the day (8AM to 6PM) and increases during the rest of the day, i.e., people move into the city during daytime and move out of the city otherwise; 2) there is a concentration peak between 11AM and 2PM, i.e., during lunchtime users are closer to the centre; 3) between 8AM and 11AM there is a sudden out-movement of people closely followed by an in-movement, i.e., many users move in&out in those hours, but not exactly at the same time;

In Figure 5 the distances Q-Q plots of two day pairs taken as example: 2015-10-22 against 2015-10-21 (Wed against Thu), near the $y = x$ line → statistically similar; 2015-11-15 against 2015-11-16 (Mon against Sun), far from the $y = x$ line → statistically different. I.e., again, users' behaviour on Sunday is different from a normal weekday.

It is also interesting to show the average distance of vehicles "per vendor" (Figure 6): 1) Enjoy is always more distant than the others (Car2go and Sharengo); 2) Sharengo (all electrical cars) is always the closest; 3) there has been a "getting near" trend in Car2go cars between 2015 and 2016.

## 3.3 Movements

A "movement" is simply a vector (that can be plotted!) between a car "catch" and a "release". The database of movements contains these records: *current_fuel, delta_fuel, delta_km, delta_time, end_date, end_pos_lat, end_pos_lon, id_vehicle, kmh, start_date, start_pos_lat, start_pos_lon, type_vehicle*. Figure 7 shows a very small time-frame (10 minutes) of car movements, this kind of graph may be used to find movement patterns and/or visually identify outliers.

# 4 Mandatory APIs proposal

The *rationale* of this proposal is based on these: 1) a resource (a large pool of cars) is converted from being private to "public" (private firms actually) control; 2) the private form of the resource is artificially limited (by regulations and policies) to push users towards the shared pool; 3) the cost burden is loaded on the public (because firms need to get revenues); 4) costs for users can be higher or lower than pre-sharing, based on usage profile; 5) workflow is heavily changed (from private to shared use); 6) environment impact can be lower if the whole system is well adapted to the actual needs of users; 7) at present, car-sharing management is in the hands of private firms which do not publish enough data to let third parties implement independent analyses.

Thus, in the authors' point of view, some form of mandatory public access to status (both real-time and historical) data should be put into place by law, to give citizens the right to public accountability also in the field of car-sharing. The ideal reification of this "third party" access would be the implementation of standardized APIs to be mandatorily supplied by car-sharing vendors. This constraint should be introduced in every public tendering procedure.

Far from being a detailed and complete proposal (a standardization process - such as [HB96] - should be activated) a suggestion for such an API could be defined using a FSA (Finite State Automaton) representation of a single car in the car-sharing system, using state transitions and dynamic and static properties:

**Dynamic:** **position** (lat - long); **fuel gauge**; **parked/taken** anonymized **user ID** (if taken).

**Static:** car **ID**; **type** (electric, diesel, gasoline, LPG, ...); **odometer**; usage **fare**; **owner**.

A "start of discussion" function list proposal is listed below. The functions could be implemented via HTTP+REST [Fie00]

- getStaticVehicles() => list of all "parked" vehicles;
- getMovingVehicles() => lista of all "non parked" vehicles;
- getInfo(id_vehicle) => status of a single vehicle;
- getInfoService(service) => combined info on service
- getPath(id_vehicle,interval) => returns the list of all paths followed by a vehicle in the time interval;
- getUserPath(id_user,interval) => returns the list of all paths followed by a single user during the interval.

All these functions should offer both the real-time and the historical version, i.e., they should also accept a "time-frame parameter such as in the last two examples.

# References


[Fie00] Roy Fielding. Representational state transfer. *Architectural Styles and the Design of Netowork-based Software Architecture*, pages 76–85, 2000.

[FM11] Jörg Firnkorn and Martin Müller. What will be the environmental effects of new free-floating car-sharing systems? the case of car2go in Ulm. *Ecological Economics*, 70(8):1519–1528, 2011.

[HB96] Richard Hovey and Scott Bradner. The organizations involved in the ietf standards process. 1996.

[JC13] Diana Jorge and Gonçalo Correia. Carsharing systems demand estimation and defined operations: a literature review. *European Journal of Transport and Infrastructure Research*, 13(3):201–220, 2013.

[LS13] Lester B Lave and Eugene P Seskin. *Air pollution and human health*, volume 6. Routledge, 2013.

[SP12] John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2012.

[SRS13] Susanne Steinle, Stefan Reis, and Clive Eric Sabel. Quantifying human exposure to air pollution—moving from static monitoring to spatio-temporally resolved personal exposure assessment. *Science of the Total Environment*, 443:184–193, 2013.

[Tre14] Andrea Trentini. Lombardy epa obtorto collo data and anti-pollution policies fallacies. *Journal of e-Learning and Knowledge Society*, 10(2), 2014.

[UE13] US-EPA. History of the clean air act. `http://www.epa.gov/air/caa/amendments.html`, 2013.

[Van09] Tom Vanderbilt. *Traffic: Why We Drive the Way We Do (and What It Says About Us)*. New York Times, 2009.

# List of Figures

Figure 1: Historical trend of $SO_2$ (source: ARPA Lombardia)

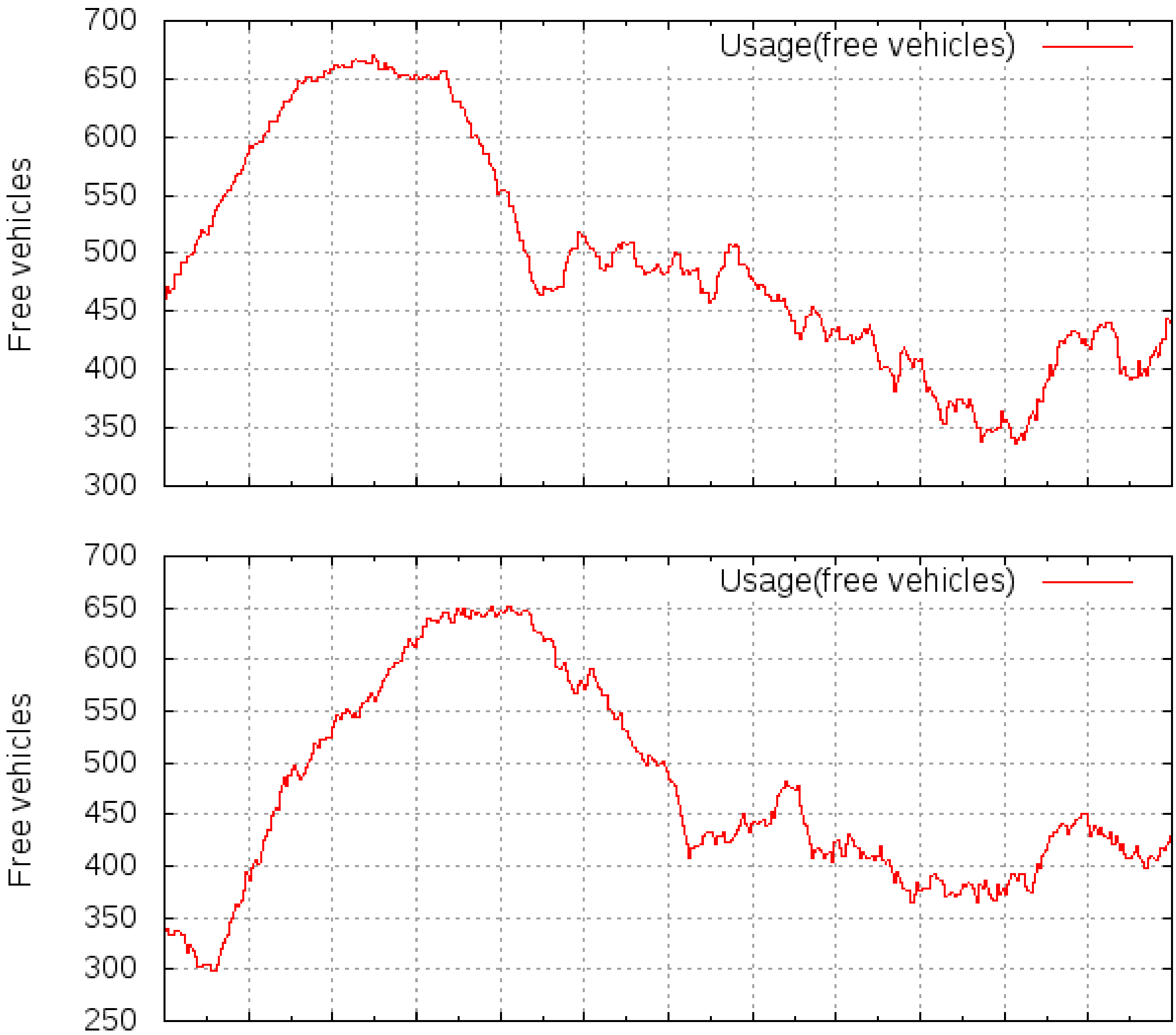
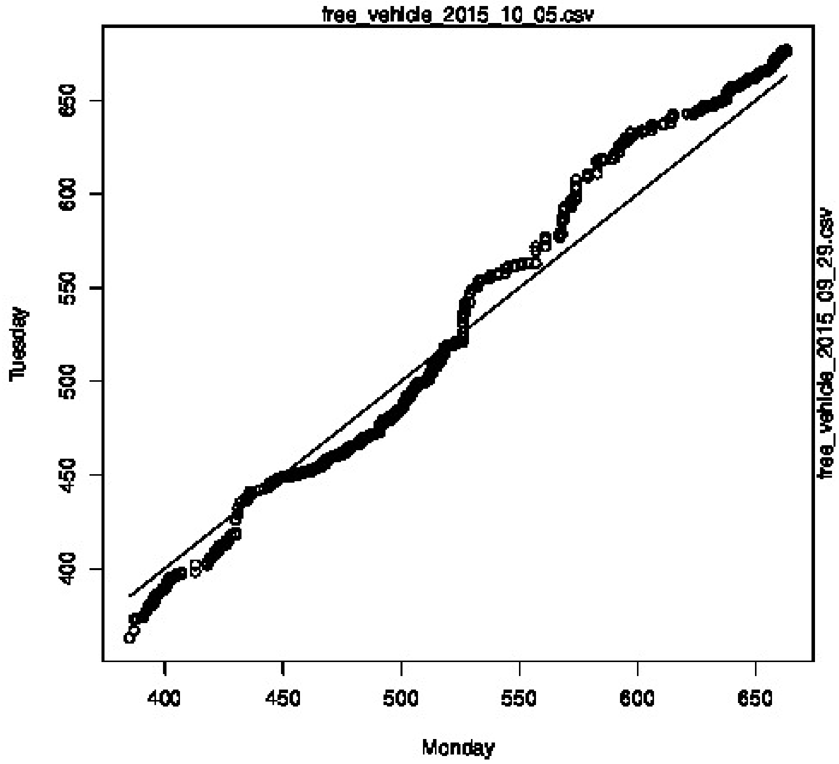



Figure 2: Usage (# of free cars) average: a typical weekday (Wed) vs. weekend day (Sun)



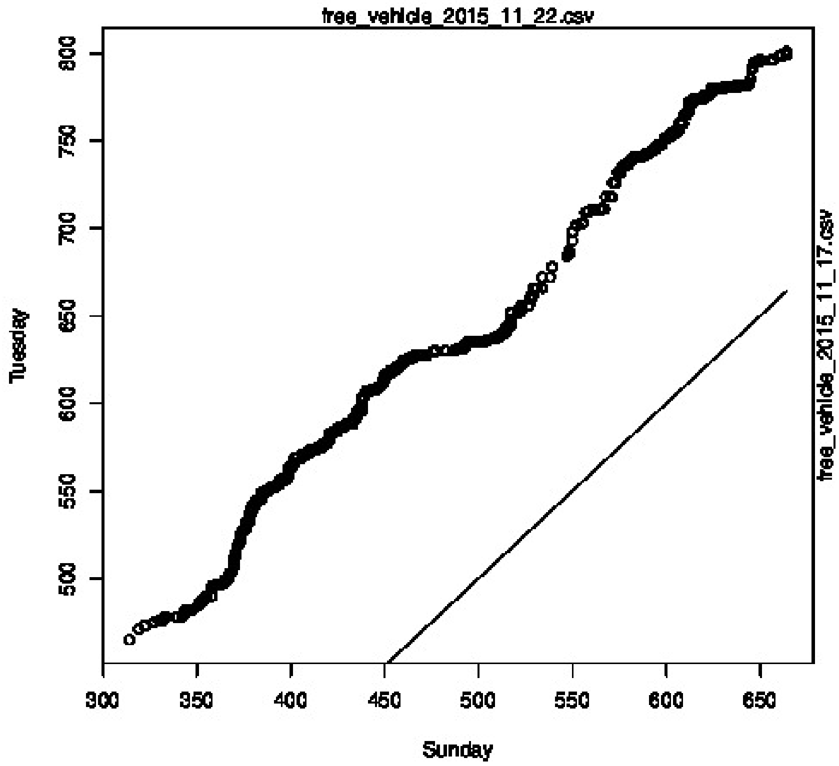




Figure 3: Usage Q-Q plots: typical NH-NH vs. H-NH

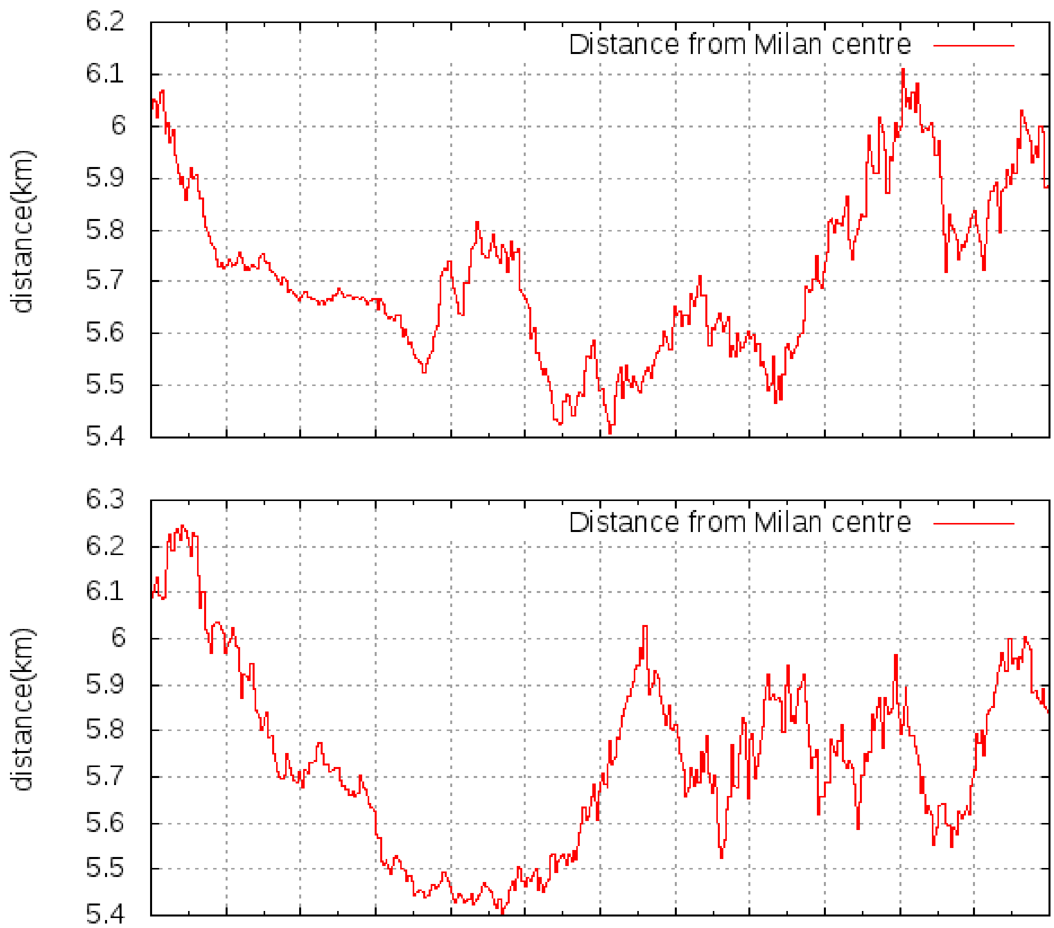



Figure 4: Average distances: a typical weekday (Wed) vs. weekend day (Sun)

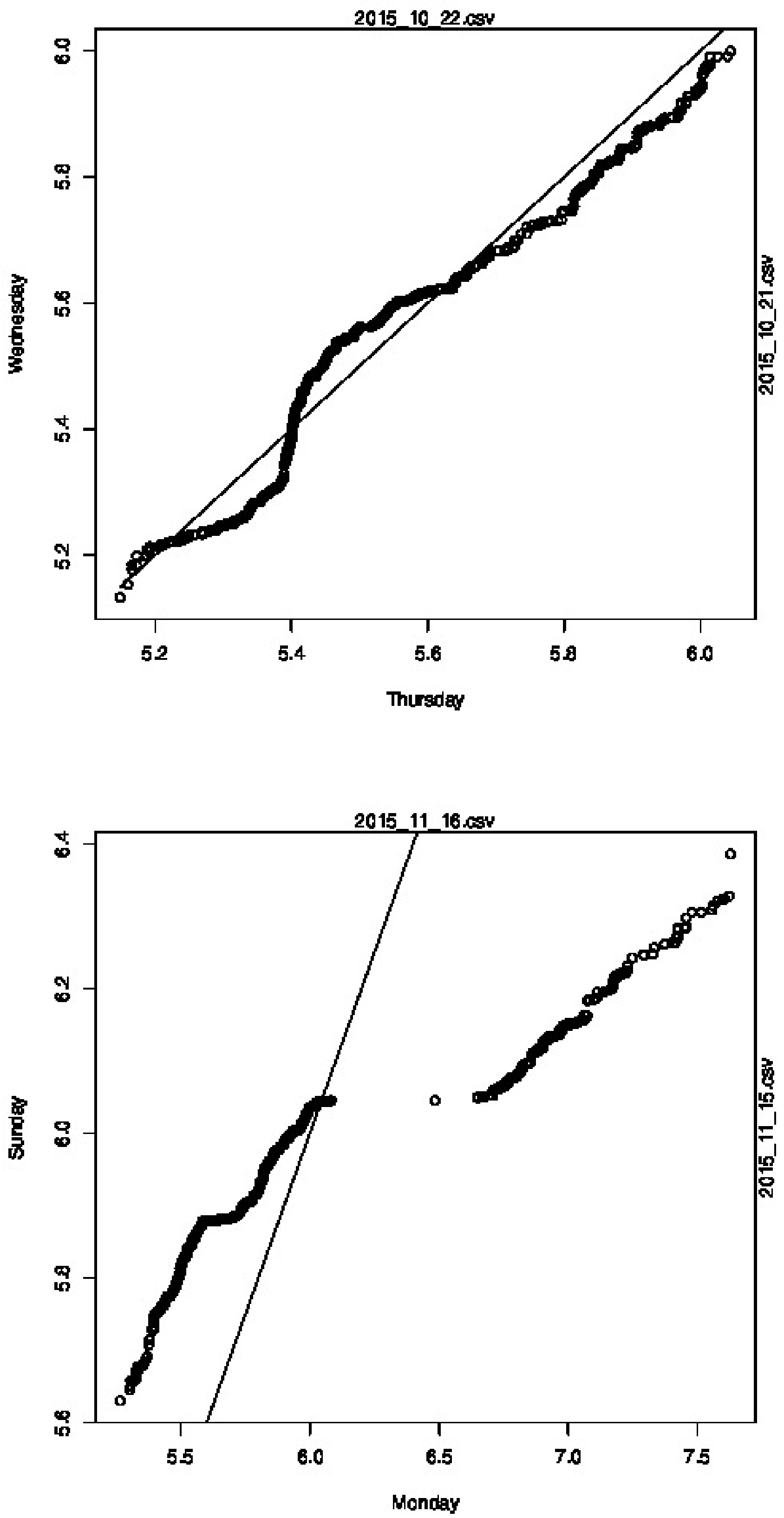



Figure 5: Distances Q-Q plots: typical NH-NH vs. H-NH



Figure 6: Difference in average distances "per vendor" (top: 2015/09/06, bottom: 2016/02/27)

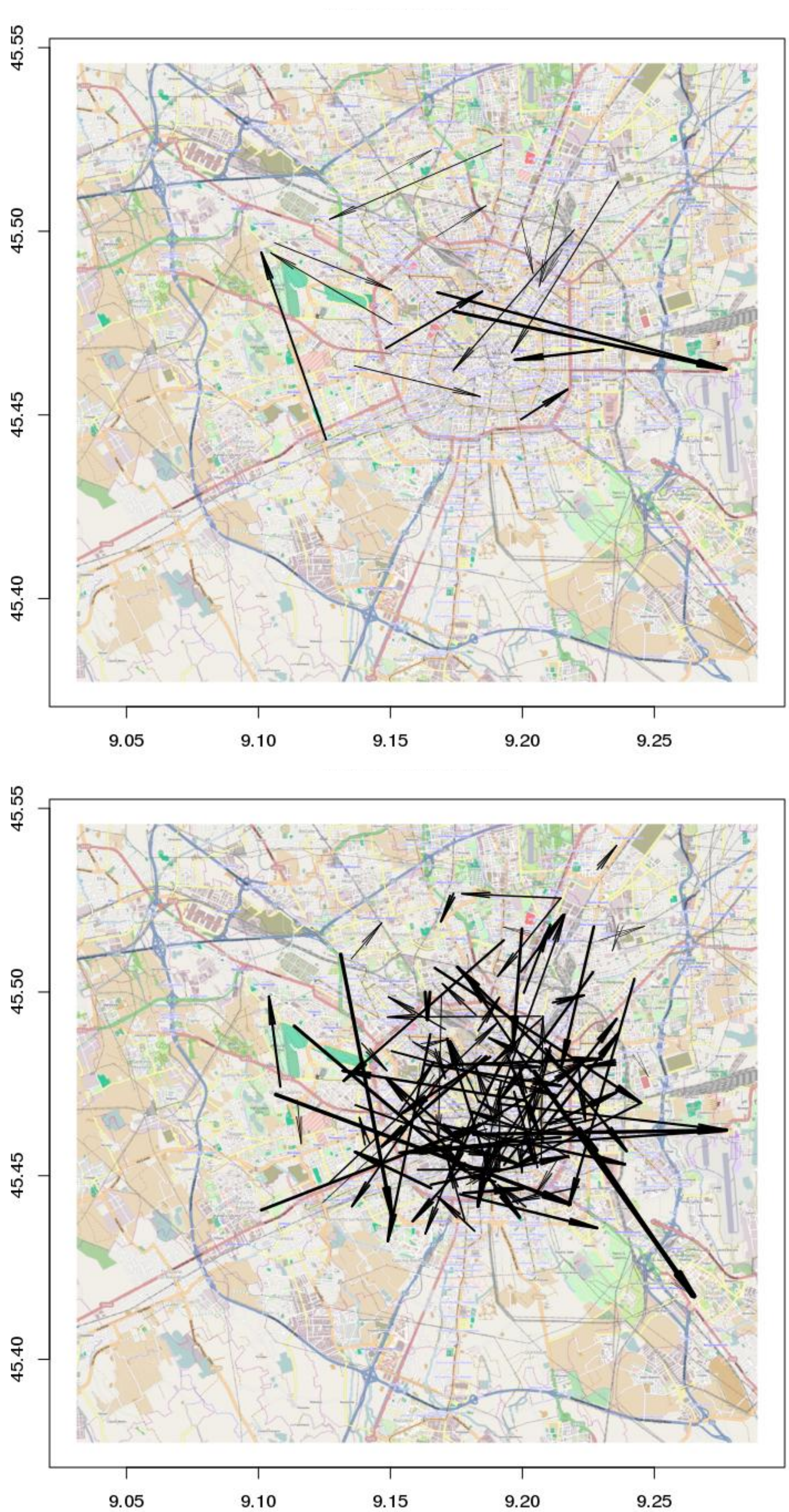



Figure 7: At 2015/09/02 08:40 (top), two cars moving from the city centre towards the airport (rightmost arrows) in the morning; at 2016/02/14 20:00 (bottom) many cars moving in the evening.