# A Preliminary Work on Predicting Travel Times and Optimal Routes Using Istanbul's Real Traffic Data

Yasin Yılmaz[1], Süleyman Uludağ[2], Esma Dilek[3], Yunus Emre Ayözen[4]

**Abstract**

Increasing urbanization worldwide looks likely to be a trend to continue in the foreseeable future. It brings in many complexities in energy consumption, transportation, housing, and, in more general, living experience of the urban dwellers in terms of pollution, waste, noise, congestion, etc. At the same time, these challenges push new opportunities to the forefront in new innovation ecosystems. One such promising framework is the paradigm of Smart Cities, where the power of computational techniques is utilized for more optimized use and management of city's assets to improve urban living. An important subsystem for the success of Smart Cities is its transportation infrastructure by means of an intelligent transportation system (ITS). The defining characteristic of ITS is its data-driven methodology. While there are studies available in the literature about traffic management, real-world traffic analytics has received little attention. In this work, we report a preliminary study of the real traffic data collected by Istanbul Metropolitan Municipality to predict travel times and optimal routes.

***Keywords*** — Traffic Management, Estimating Smart City Traffic, Smart City Traffic Data, IMM, Travel Time Prediction

## 1 Introduction

As a result of demographic shifts worldwide, the density of urban areas is swiftly increasing. With some estimates, including recent UN reports [1], close to 70% of the world populations is expected to live in urban areas and big cities by 2050. This trend for greater urbanization brings about many problems in housing, transportation, energy, and manufacturing in terms of pollution, waste, noise, congestion, etc. At the same time, these issues are giving rise to new innovation ecosystems as facilitators [2]. One such framework in these developments is the paradigm of the *Smart City* (SC). The defining and distinguishing characteristic of SCs is the ubiquitous integration and exploitation of information and communications technologies (ICT) in the urban areas and by heavily resorting to computational, control, and optimization methodologies. Major goals of SCs include raising standards of living, increasing more efficient use of resources, providing convenience, improving environmental conditions, reducing crime, etc.

It is generally agreed that one of the most important subsystems of SCs is in smart mobility and transportation which is on the verge of a profound and large-scale transformation [3]. A general term used to refer to efforts in this regard is Intelligent Transportation System (ITS) [4]. The interaction and integration of ITS and SC is depicted in Figure 1. ITS makes use of Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) applications, such traffic management techniques as adaptive cruise control, cooperative navigation, cooperative adaptive cruise control (platooning), etc., to enhance road and vehicle safety, improve traffic efficiency, and provide entertainment and

[1]University of South Florida, Electrical Engineering Department, Tampa, Florida, USA, yasiny@usf.edu

[2]University of Michigan – Flint, Computer Science Department, Michigan, USA, uludag@umich.edu

[3]Istanbul Metropolitan Municipality, Department of Transportation, Directorate of Traffic, Istanbul, Turkey, esma.dilek@ibb.gov.tr

[4]Istanbul Metropolitan Municipality, Department of Transportation, Istanbul, Turkey, yunusemre.ayozen@ibb.gov.tr

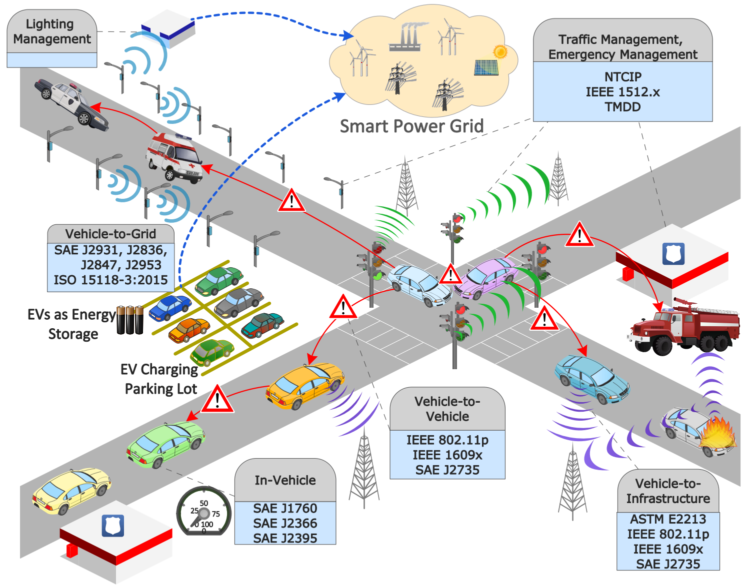



Figure 1: Intelligent Transportation System and Smart City are closely coupled.

information to vehicle occupants. Rather than developing new roads or augmenting road capacities, which might not be possible due to financial, geographic, social, and historical constraints in some cities, ITS utilizes ICT to improve transportation experience of the citizens. It makes heavy use of static, vehicular, and/or mobile app-based sensor data. Major mechanisms ITS uses are congestion control systems, car-pooling, smart scheduling of traffic, the use of connected and autonomous vehicles (CAV), augmented mass transportation, and traffic estimation. The expected benefits of ITS are reduced traffic time waste[5], reduced energy consumption, improved productivity, decreased stress levels, lower pollution, diminished traffic accident fatalities. Massachusetts Institute of Technology (MIT) researchers predict that [6] a fleet of shared CAVs in Singapore can almost eliminate parking problems and provide mobility to everyone in the city of Singapore [3].

The most critical feature of an ITS in an SC is its data-driven methodology. With the increasing collection, access, and sharing of traffic data, it is essential to analyze the real-world big data and develop techniques to facilitate the aforementioned benefits. While there are many studies about traffic management, especially in traffic light coordination [7, 8, 9], to the best of our knowledge, real-world data analytics in developing these techniques has received little attention. It is within this line that the Istanbul Metropolitan Municipality (IMM) Directorate of Traffic has deployed over 700 sensors to collect relevant real-time traffic data [10]. In this study, we present a preliminary study of the initial data analysis we have conducted.

The rest of the paper is organized as follows: Section 2 summarizes the data collection infrastructure together with the major sensors deployed. We present our data-analysis methodology in predicting accurate travel times in metropolitan Istanbul and a model to pick optimal route based on the former analytics in Section 3 followed by a discussion of the results and the future work in Section 4. Concluding remarks are provided in Section 5.

[5]According to a TomTom report in 2013 as reported in [5], more than 75% extra time is lost due to traffic congestion in cities like Moscow, Istanbul, Rio de Janeiro, Mexico City and Beijing.

## 2 Data Collection Methodology

Istanbul Metropolitan Municipality (IMM) takes advantage of various systems to collect real-time traffic information in Istanbul. Systems used for traffic data collection are summarized below:

1. **Radar-Based Sensors:** These sensors detect vehicles using radar signals and are installed road-side to collect traffic measurement data over a period of time. Traffic measurement data, such as average speed, volume, occupancy and vehicle classification information, are collected from radar-based sensors every two or five minutes. Primary traffic data source used in Istanbul depends on data collected from radar-based sensors. Collected data are transmitted to IMM's Traffic Control Center (TCC) via GPRS at set intervals (two minutes or five minutes). After authentication carried out by IMM's server, raw sensor data is processed and stored into a database to be used by traffic applications. Three types of radar sensors deployed to collect data are shown in Figure 2.


(a) RMTS G4. (b) RMTS SX-300. (c) Wavetronix Smartsensor HD.

Figure 2: Remote Traffic Microwave Sensor (RMTS) G4, SX300, and Wavetronix Smartsensor HD.

2. **Image Processing Sensors:** These sensors, as shown in Figure 3a, are used both on highways, expressways and within tunnels to collect traffic measurement data. Live videos of traffic observation cameras are analyzed and processed by IMM's image processing sensors. Generated traffic measurement data are transmitted to TCC via wired or wireless connections and stored into IMM's database.

3. **Bluetooth Sensors:** These sensors, as shown in Figures 3b and 3c, are utilized by IMM to obtain point-to-point journey times. Bluetooth signals emitted by devices such as headsets, navigation tools, mobile phones are detected by bluetooth sensors. The MAC addresses of the detected devices are transmitted to IMM's server with no other identifying information (to protect their privacy) via wired or wireless connections. Point-to-point journey times and average speeds of the roads are calculated by detecting the same MAC addresses in successive bluetooth sensors within the city road network.

4. **Floating Car Data and Mobile Application Users' Data (Probe Data)**: In addition to physical traffic measurement sensors installed on highways, IMM also takes advantage of floating car data from private companies and probe data collected from IMM's mobile traffic application users, who allow (i.e opt-in) IMM to use their GPS position to be transmitted to TCC to generate instantaneous traffic data.

This study is carried out using traffic measurement data (i.e. average speeds and number of detected vehicles) collected from radar-based sensors of IMM which are installed on highways in Istanbul.

(a) Image Processing Sensor (RackVision-Terra). (b) Bluetooth DC Model. (c) Bluetooth SC Model.

Figure 3: DeepBlue DC-model and SC-model bluetooth sensors.

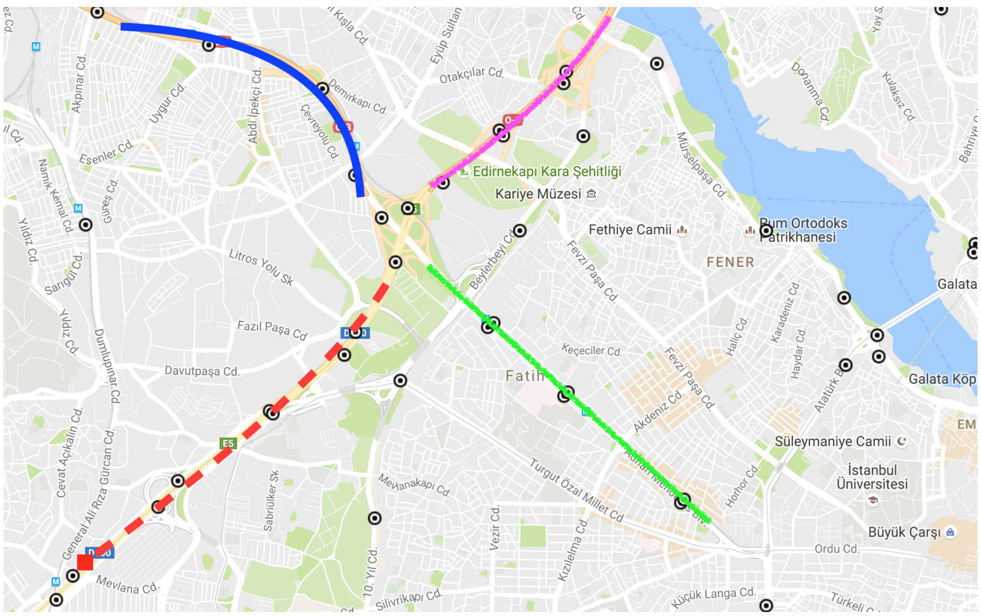



Figure 4: Road segment examples. They are typically separated by intersections. Each line consists of two segments running in opposite directions. Sensors are marked with circles. Map is obtained from Google Maps.

## 3 Data Analysis

We follow an object-oriented methodology to analyze the IMM traffic data. Our objects are the *road segments* that constitute the road network of Istanbul. They can be considered as the pieces of a puzzle. We define a road segment as a connected part of a major road (e.g., D100) that throughout exhibits a homogeneous traffic pattern (i.e., the same number of lanes, similar number of vehicles and vehicle speeds). In that sense, the road segments that we consider are typically a few kilometers, but not necessarily the same length (e.g., they range from 1.5 km to 6.4 km in the example shown in Fig. 6). A road segment has a direction. Hence, in major roads, there are usually two parallel segments for both traffic directions. Examples of road segments are shown in Fig. 4. We monitor each segment using a number of sensors, which is typically different for each segment (see Fig. 4). As described in Section 2, sensors provide different types of measurements. In this paper, we consider the average speed and volume (i.e., number of detected vehicles) measurements.

We aim to predict the future travel times based on the statistics of the road segments estimated from the available data. To that end, we initially organize the available data to estimate the statistical features of the road segments.

| | 0 km/h | $\cdots$ | 80 km/h | 81 km/h | $\cdots$ | 199 km/h |
|---|---|---|---|---|---|---|
| Mon 01:00 | 0 | $\cdots$ | 2425 | 116 | $\cdots$ | 0 |
| $\vdots$ | | | | | | |
| Thu 14:00 | 0 | $\cdots$ | 4029 | 295 | $\cdots$ | 0 |
| $\vdots$ | | | | | | |
| Sun 00:00 | 0 | $\cdots$ | 2444 | 144 | $\cdots$ | 0 |

Table 1: Example of a table showing, for a certain road segment, the numbers of vehicles traveling at speeds in the $0-199$ km/h range for each hour in the week. Each number in the table represents the total number reported by all sensors monitoring the same road segment. Such a table can be represented by a $168 \times 200$ matrix.

### 3.1 Data Organization

We organize sensor data in terms of the road segments. Specifically, the average speed and number of vehicles data from the sensors which monitor the same segment are grouped together to form the measurement dataset for that segment. Some sensors monitor vehicles traveling in both directions, so they contribute to the datasets of two segments. However, some sensors monitor traffic flowing only in one direction.

For each road segment, we form a table which shows the number of vehicles traveling at a certain speed at a particular time. More specifically, our table holds the number of vehicles for each hour of the week (i.e., $7 \times 24 = 168$ hours from Monday 00:00-01:00 to Sunday 23:00-00:00) for a reasonable range of speed, $0 - 199$ km/h. Table 1 illustrates the table for a road segment.

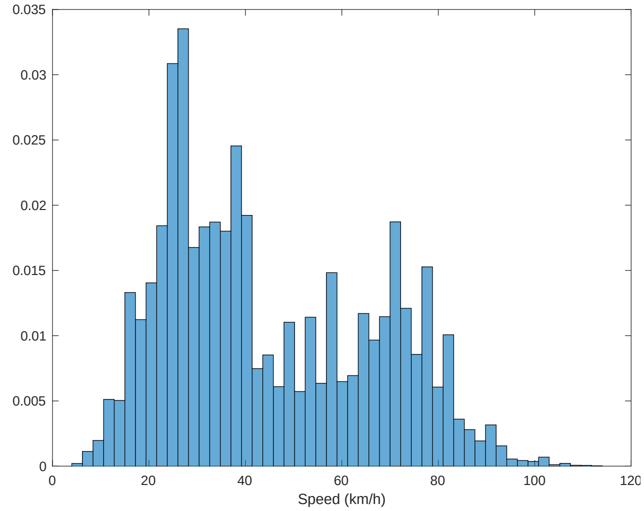
### 3.2 Prediction

Using the data tables we can estimate the probability density function (pdf) of vehicle speeds in a road segment for each hour of the week, i.e., 168 pdfs for each road segment. As shown in Fig. 5, we can estimate a pdf using histogram- or a kernel-based method, e.g., using Gaussian kernel function. Both histogram- and kernel-based density estimation are nonparametric methods. Alternatively, parametric estimation can be used by fitting a probability distribution to the data samples. Specifically, the parameters of the assumed distribution (e.g., mean and variance of the Gaussian distribution) are estimated using the samples. If the assumed distribution fits well to the data, parametric estimation works well with a relatively small number of samples. However, it performs poorly unless the assumed model fits the data well. We here resort to nonparametric methods since we have abundance of samples in our dataset that suffice to estimate the statistics well.

We compute the sample mean for each road segment $i$ at each hour $j$ as follows:

$$\bar{x}_{ij} = \frac{\sum_{k=0}^{199} n_{ijk}\, k}{\sum_{k=0}^{199} n_{ijk}}, \tag{1}$$

where k denotes the speed, and $n_{ijk}$ is the number of vehicles recorded as traveling with speed $k$



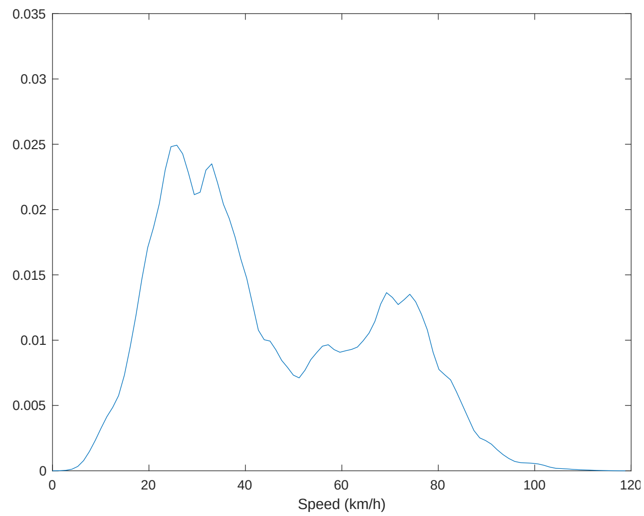




Figure 5: Histogram-based (top) and Gaussian kernel-based (bottom) pdf estimation of vehicle speeds for the segment of the D100 highway between Zeytinburnu and Topkapi (shown by the red dashed line in Fig. 4) on Monday between 07:00 and 08:00.

km/h in road segment $i$ at hour $j$. The confidence interval for the sample mean $\bar{x}_{ij}$ is given by

$$\left( \bar{x}_{ij} - z^*_{ij} \frac{s_{ij}}{\sqrt{\sum_{k=0}^{199} n_{ijk}}}, \bar{x}_{ij} + z^*_{ij} \frac{s_{ij}}{\sqrt{\sum_{k=0}^{199} n_{ijk}}} \right), \tag{2}$$

where $s_{ij}$ is the sample standard deviation, and $z^*_{ij}$ is found from the t-distribution table using a confidence level (e.g., 0.95) and the degree of freedom is $\sum_{k=0}^{199} n_{ijk} - 1$ [11].

Since the number of samples in our datasets is quite high, we are able to accurately estimate the average speed in each road segment for each hour. For the instance illustrated in Fig. 5, the number of samples is 693191, and the 99% confidence interval for the sample mean 44.9 km/h is $(44.83, 44.97)$ km/h.

Therefore, we predict the travel time in road segment $i$ at hour $j$ by

$$\hat{t}_{ij} = \frac{\ell_i}{\bar{x}_{ij}}, \tag{3}$$

where $\ell_i$ is the length of the segment $i$ in km. Then, the total travel time for route $I$ is predicted as the sum of the predicted travel times in the segments that constitute $I$, i.e.,

$$\hat{t}_{Ij} = \sum_{i \in I} \hat{t}_{ij}. \tag{4}$$

When there are multiple possible routes $I = \{I_1, \ldots, I_m\}$, the one with the smallest predicted travel time is chosen as the optimal route, i.e.,

$$I^* = \arg \min_{r=1,\ldots,m} \hat{t}_{I_r j}. \tag{5}$$

In short, given the departure and arrival locations, and the departure/arrival time, our proposed procedure finds the potential routes $I = \{I_1, \ldots, I_m\}$, and provides the user with the optimal route $I^*$, and the predicted travel time

$$t^* = \min_{r=1,\ldots,m} \hat{t}_{I_r j}. \tag{6}$$

Note that only (4)–(6) are computed online. The predicted travel times $\{\hat{t}_{ij}\}$ for all segments and hours of the week are computed (see (3)) and stored offline. Nevertheless, with the new coming sensor data, (1) and accordingly (3) can be easily updated. Hence, the proposed prediction scheme, having a very low computational complexity, is capable of performing online updates and predictions.

### 3.3 Example

To demonstrate our procedure we provide experimental results for a sample trip from Zeytinburnu to Taksim (see Fig. 6). From the sensor data from January 1, 2016 to August 31, 2016 we predict, using (3), the travel duration for the three routes shown in Fig. 6.

In Fig. 7, we show the weekly traffic pattern. Morning and evening rush hours are clearly noticed on weekdays. Saturday and Sunday have different traffic patterns than weekdays, as expected. For instance, there is no morning traffic jam over the weekends. The predicted travel times are the smallest on Sunday due to the common "rest day" notion.

| | Mon 08:00 | Tue 12:00 | Thu 17:00 | Sat 02:00 |
|---|---|---|---|---|
| Route 1 (15.3 km) | 17.4 min | 15 min | 17.9 min | 11.1 min |
| Route 2 (14.2 km) | 17.9 min | 16.4 min | 19.1 min | 12.2 min |
| Route 3 (14.1 km) | 17.8 min | 16.1 min | 18.3 min | 12.1 min |

Table 2: Predicted travel times for different times of the week, i.e., rush hour for Monday morning and Thursday evening, and less busy hours for Tuesday noon and Saturday night. Route lengths are also shown next to the route names. Note that the travel times are underestimated due to the coarse modeling based solely on average speeds. For finer modeling and more accurate predictions, see the discussions in Section 4.

Route 1, which has the highest highway percentage among all routes, seems to be the fastest route. Actually, this is an expected result since our predictions are solely based on average speeds,


Figure 6: Three alternative routes for the Zeytinburnu-Taksim trip. Routes are determined by Google Maps. Road segments and their lengths are annotated in red.

which are obviously higher on highways. However, the advantage of Route 1 disappears during rush hours, which, in Fig. 7, correspond to the double peaks on weekdays.

For example, if the user wants to go to Taksim from Zeytinburnu on Monday at 8:00, then our predicted travel times for the three routes are all around 18 minutes, as shown in Table 2. Considering the route lengths, which are given in Table 2, in this case, Route 3 seems to be the optimum route. The tradeoff between driving an extra km and driving a little shorter is also seen in the evening rush hour on Thursday (at 17:00). In this case, the user may still choose Route 3 considering the small difference (only 0.4 min) in the predicted travel times for Route 1 and Route 3. The highway advantage of Route 1 is more obvious during less busy hours (e.g., Tue 12:00 and Sat 02:00 in Table 2). According to the results shown in Table 2 (and others that could not be presented here), Route 3 is more advantageous than Route 2. Specifically, it is a little shorter both in length and duration. Hence, to decide the optimum route, the comparison is usually between Route 1 and Route 3. As already discussed, Route 1 is a little shorter in duration, but at the same time a little longer in length compared to Route 3, manifesting a tradeoff between the time and fuel consumption. The same tradeoff in general exists in many trips with multiple potential routes. It is up to the user to strike a balance between the route length (or fuel consumption) and duration.

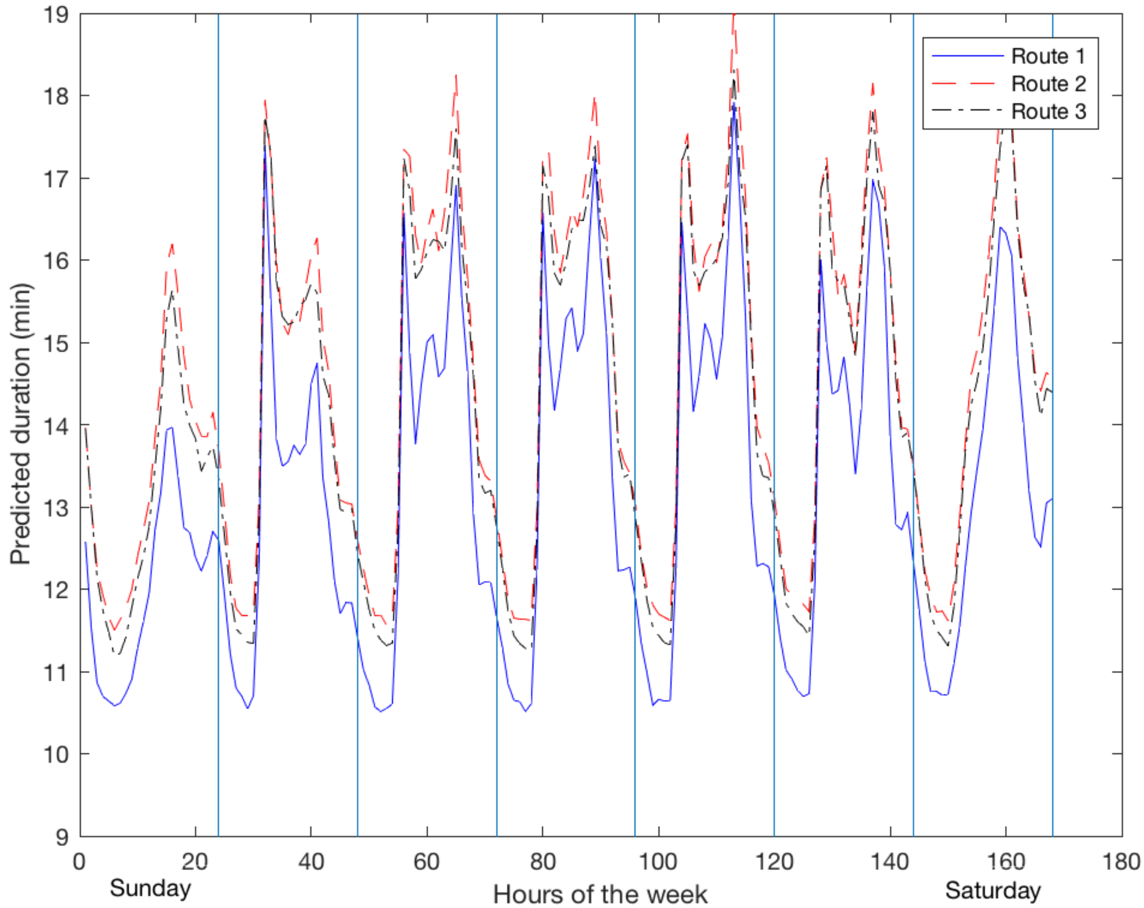



Figure 7: Graph showing the predicted travel durations in minutes for the three alternative routes from Zeytinburnu to Taksim against the hours of the week. Note that the week starts with Sunday.

# 4 Discussions

In this paper, we present a methodology for predicting travel times. To provide more accurate predictions, a more detailed modeling than (3) is needed. An immediate remedy is to consider shorter road segments (cf. Fig. 6), which would increase the number of segments, hence the computational complexity, but at the same time would enable finer modeling, hence more accurate predictions. We should also consider delays due to the traffic lights and intersections. In a future work, using the car tracking dataset, which is also available from IMM, we plan to regress the point-to-point travel times on the average speeds of the tracked cars in each road segment along the path. Then, using the predicted average speed of each road segment in the learned regression model we can predict the travel times more accurately.

Adding such correction factors into our model would ensure more accurate predictions. Seasonal effects, such as the difference between summer and winter traffic patterns, should also be considered for more realistic prediction. Moreover, city events, such as sports events, concerts, social/political gatherings, cultural events, and holidays, can be incorporated into the proposed predictor by analyzing the past data for such events. If the trip is planned for such a special day, statistics for such special days would be considered, regardless of the day of the week.

An important advantage of estimating the probability distribution of vehicle speeds is that, in addition to prediction, we can also detect the anomalies in the sensor data. Statistically significant deviations from the baseline can be detected using different methods, such as p-values, chi-squared test, Kolmogorov-Smirnov test, and online change detection. While anomalies in the sensor data may simply originate from sensor failures, they may also correspond to important reasons such as accidents, or congestion due to another reason.

As future research directions, we plan to improve our prediction scheme, and also consider detecting anomalies in the sensor data, as described above.

# 5 Conclusion

An indispensable subsystem to the success of the emerging Smart Cities concept is the transportation infrastructure. Intelligent Transportation System (ITS) is expected to make significant positive contributions to the dense urban areas, which is forecast to make about 70% of world population by 2050. While traffic management, mostly in terms of traffic light management, has been studied for a long time, the developments of information and communication technologies coupled with huge data collection deployments now provide unprecedented opportunities to develop new smart techniques. In this study, we report a preliminary analysis of the real-world traffic data as collected by Istanbul Metropolitan Municipality. We provide a traffic travel time prediction and optimal path selection methodology using this real traffic from Istanbul.

# References


[1] United Nations Department of Economic and Social Affairs, "World urbanization prospects, the 2014 revision," Accessed October 21, 2016. [Online]. Available: http://esa.un.org/unpd/wup/

[2] S. W. Turner and S. Uludag, "Intelligent transportation as the key enabler of smart cities," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, April 2016, pp. 1261–1264.

[3] President's Council of Advisors on Science and Technology (PCAST), "Technology and the future of cities," February 2016, Accessed October 21, 2016. [Online]. Available: https://goo.gl/3XLTWE

[4] S. W. Turner and S. Uludag, "Chapter 6 : Towards Smart Cities: Interaction and Synergy of the Smart Grid and Intelligent Transportation Systems," in *Smart Grid: Networking, Data Management and Business Models*, Hussein T. Mouftah and Melike Erol-Kantarci, Ed. CRC Press, April 2016. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/b19664-9

[5] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nature Communications*, vol. 7, pp. 10 793 EP –, Mar 2016, article. [Online]. Available: http://dx.doi.org/10.1038/ncomms10793

[6] K. Spieser, K. Treleaven, R. Zhang, E. Frazzoli, D. Morton, and M. Pavone, *Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore*. Cham: Springer International Publishing, 2014, pp. 229–245.

[7] E. R. Lira, E. Fynn, P. R. S. L. Coelho, L. F. Faina, L. Camargos, R. S. Villaca, and R. Pasquini, "An architecture for traffic sign management in smart cities," in *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, March 2016, pp. 580–587.

[8] M. Shahidehpour, Z. Li, S. Bahramirad, and A. Khodaei, "Optimizing traffic signal settings in smart cities," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[9] A. C, M. Kumar, and P. Kumar, "City traffic congestion control in indian scenario using wireless sensors network," in *Wireless Communication and Sensor Networks (WCSN), 2009 Fifth IEEE Conference on*, Dec 2009, pp. 1–6.

[10] E. Dilek and Y. E. Ayözen, "Smart mobility in istanbul with ibb ceptrafik," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, April 2016, pp. 1273–1278.

[11] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, ser. Springer Texts in Statistics. Springer, 2004. [Online]. Available: https://books.google.com/books?id=th3fbFI1DaMC