Towards Improved Vehicle Arrival Time Prediction in Public Transportation: Integrating SUMO and Kalman Filter Models

Ahmad Faisal Abidin
Computing Science and Mathematics
University of Stirling
Stirling, Scotland
faa@cs.stir.ac.uk

Mario Kolberg
Computing Science and Mathematics
University of Stirling
Stirling, Scotland
mko@cs.stir.ac.uk

Abstract— **Accurate bus arrival time prediction is a key component for improving the attractiveness of public transport. In this research, a model of bus arrival time prediction, which aims to improve arrival time accuracy, is proposed. The arrival time will be predicted using a Kalman Filter (KF) model, by utilising information acquired from social networks. Social Networks feed road traffic information into the model, based on information provided by people who have witnessed events and then updated their social media accordingly. In order to accurately assess the efficiency of KF model, we simulate realistic road scenarios using the traffic simulator Simulation in Urban Mobility (SUMO). SUMO is capable of simulating real world road traffic using digital maps and realistic traffic models. This paper discusses modelling a road journey using Kalman Filters and verifying the results with a corresponding SUMO simulation. As a second step, SUMO based measures are used to inform the KF model. Integrating the SUMO measures with the KF model can be seen as an initial step to verifying our premise that realtime data from social networks can eventually be used to improve the accuracy of the KF prediction. Furthermore, it demonstrates an integrated experimental environment.**

*Keywords***:** *SUMO, Kalman Filters, Bus Arrival Prediction*

I. INTRODUCTION

Literature reviews [1][2] have shown that most arrival time prediction models are based on historical arrival patterns and/or other explanatory variables correlated with the arrival time. The explanatory variables used in these previous studies include historical arrival time (or travel time), schedule adherence, weather condition, time-of-day, day-of-week, dwell time, number of stops, distance between stops and road–network condition [1][2]. The collection and transmission of such variables has been largely made possible using emerging technologies, including wireless communication, automatic vehicle location (e.g. Global Positioning System(GPS)), and other sensing technologies. The effect of congestion was treated differently in most models. For example, some have used traffic properties like volume and speed from simulation results [3], while others have clustered their data into different time periods [4]. Historical data based models are used in geographical areas with less traffic congestion, because the models assumed cyclical traffic patterns. Kalman Filter (FK) techniques and Artificial Neural Network (ANN) approaches were used mostly in urban areas [1]. KF models can be applied while a

bus trip is in progress due to its simplicity of calculation. While other models are dependent on cyclical traffic data patterns or require independence between variables, ANNs do not require that variables be uncorrelated or have a cyclic pattern.

There are many algorithms based on mathematical theory and/or statistical models that have been proposed for bus travel time prediction. However, there is a gap amongst those algorithms. One particular issue is the question of how the algorithm will receive and incorporate live real-time traffic event information. Without receiving such information, those algorithms could not produce an accurate result [5,6,7]. One example of such information is detail on road incidents. This kind of information is vital to be passed to KF Model for further processing. Social network communication is a novel way to collect and include current road condition information, rather than using GPS or other (road) sensors to detect the numbers of cars on the road and speed of travel. In addition, this approach allows for the identification of unexpected traffic events, and the subsequent inclusion of this new, real-time information, as part of potential route calculations and updates. This provides updated information during journeys that may not have been available when travel was initially planned or started. In this situation, social networks can play a pivotal role as a source of information. Simulation in Urban Mobility (SUMO) can run realistic simulations including traffic patterns and road conditions. In this paper we model a particular journey with a Kalman model and then try and verify the prediction with the SUMO simulation. Further, we feed location information from SUMO into a KF model to improve its prediction accuracy.

This demonstrates: verification of KF prediction results, an integrated experimentation environment for arrival time predictions, and that the use of live data improves the KF prediction accuracy. This can be exploited by using road condition information from social networks together with KF prediction models.

The remainder of this paper is divided as follows. Section II discusses the KF Models used in this paper, as well as briefly reviewing other work in the literature. Section III then provides a description of SUMO and the modifications made to successfully utilise it as part of this work. Experimental results are presented in section IV, with section V concluding the paper.

II. KALMAN FILTER(KF) MODELS

Much research has been conducted into forecasting bus travel/arrival time using a variety of techniques and models. This includes techniques such as time series analysis and ANNs, and a summary is shown in table 1. However, only two models, Statistical[8] and Kalman Filter models [9], take dynamic information into consideration. KFs are a powerful tool when it comes to controlling noisy systems [10]. The basic idea of KFs is that they process (filter) incoming data, and produce an improved quality of output data.

TABLE I.   MODELS OF BUS TRAVEL TIME PREDICTIONS

Technique	Remarks	Dynamic Info Considered
Historical [11]	Predict travel time at particular time as the average travel time for the same period historically	No
Time series analysis[12]	Predict assuming historical patterns will remain the same in the future	No
Artificial Neural Networks [13]	Predict based on example data using large database for accurate prediction	No
Real-time approach[14]	Assume the future travel time to be the same as the present one	No
Statistical [8]	Predict travel time based on dependent variable based on a function with independent variables	Yes
Kalman Filters(KF) [8,10]	Establish relationships between variables, and corroborates using field observation	Yes

Kalman Filters are an estimation approach. That is they infer values of parameters from observations which may be noisy, inaccurate, and uncertain. Importantly, unlike many other approaches, Kalman Filters are recursive and hence can take new observations into account as they arrive. With this, Kalman Filters can be executed at runtime of the system under observation. The algorithm is referred to as a 'filter' as calculating the estimate from noisy input data essentially 'filters' out the noise[15]. Kalman filters estimate a process by estimating the process state at a given time and then obtaining feedback in the form of noisy measurements. Generally, the equations for the Kalman filter fall into two groups: time update equations and measurement update equations. Time update equations are responsible for predicting forward (in time), using the current state and error covariance estimates to obtain a priori estimates for the next time step. The measurement update equations are responsible for the feedback, i.e. for incorporating a new measurement into the a priori estimate to obtain an improved a posteriori estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. Indeed, the final estimation algorithm resembles a predictor/corrector algorithm for numeric problems [15].

The addition of state constraints (physical information) to a KF can significantly improve the estimation accuracy of the filter [5,7]. In this sense, the addition of linear information may be useful for that goal. KF models theoretically deliver the best and most up to date results when they have continuous access to dynamic information. Many existing models therefore make use of dynamic information. However, the performance of these models can often suffer due to issues with taking account of scenarios that utilise rapidly updating real world information. Many road users utilise GPS to navigate and estimate the duration of their journeys. However, it is not always possible to handle some events solely by GPS information. For instance, if there is an accident and we know it will take 30 minutes or more to clear; this information has to be made available to and be included in the prediction algorithm. Social networks can distribute real time information on road traffic, which could potentially improve the accuracy of arrival time prediction.

III. SIMULATION OF URBAN MOBILITY (SUMO)

SUMO [16,17] is a road traffic simulation engine which is designed to handle large road networks. Every vehicle is modelled explicitly. Every vehicle has their own routes, and moves individually through the network. This means that for every vehicle in the simulated network its current location and speed are known. At every time step which has a duration of 1s, these values are updated. The simulation of the network is time discrete and space continuous. When simulating traffic, the street attributes, such as maximum velocity and right of way rules, are taken into account. Conceived as multi-modal environment, SUMO not only can simulate a model of car movements, but also a variety of public transportation models including trains and buses. Public transportation is described by a departure time and the route it takes, which again is made up of subroutes that describe a single traffic modality.

A. Traffic Modelling in SUMO

After having generated a network, one could inspect it using SUMO graphical user interface (SUMO-GUI), but no cars would be driving around. One still needs some kind of description about the vehicles. This is called the traffic demand [18]. A trip is a vehicle movement from one place to another defined by the starting edge (street), the destination edge, and the departure time. A route is an expanded trip, that means, that a route definition contains not only the first and the last edge, but all edges the vehicle will pass. SUMO and SUMO-GUI need routes as input for vehicle movements. There are several ways to generate routes for SUMO. The choice depends on the available input data. In this work, the vehicular demand modelling was derived from the project Travel and Activity Patterns Simulation (TAPAS) in Cologne City, Germany [19]. The TAPAS project generates trips of drivers by exploiting information on the population, demographic characteristics, the points of interests in the urban area, places where work and free-time activities take place. Within the context of the TAPASCologne project, data collected in the City of

Cologne was recorded, including 30,700 daily activity reports from more than 7000 households. TAPAS published this open data and permits users to integrate it with other vehicular models.

Figure1. Cologne City network imported into SUMO

The research presented in this paper leverages the openness of TAPAS by constructing a model of bus movement based on the TAPAS data. Road accidents were constructed using TAPAS data on how the traffic flow reacts over particular traffic incidents. This resulted in a realistic traffic model for our experimentation. Figure 1 and 2 show two screenshots of SUMO. Figure 1 presents a large scale overview of the road network, whereas Figure 2 depicts a small part of the network with some vehicles present.
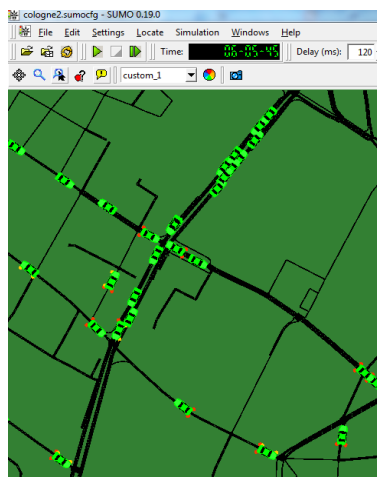
Figure 2. Screenshot of vehicle movement in Cologne City

B. SUMO Output

SUMO allows for the generation and output of a large number of different measures to files or socket connection. Outputs such as a raw vehicle positions, trip information, and vehicle route information are triggered using the command line. For this research, we employ all of the outputs above. Raw vehicle positions contain edges (streets) and lanes along with the vehicles driving on them for every time step, where vehicles are described by their name, position and speed.

This output has been utilised as input to our KF model to update the predictions based on real time trip information, as will be discussed in the following section. The trip information contains the vehicle departure time, the journey start time, and the arrival time; all stored as XML data (Figure 3 and Figure 4).

Figure 3. Screenshot of Raw Dump output in SUMO

Figure 4. Screenshot of network generate files in SUMO

Values in Figure 3 are defined in Table II:

TABLE II. RAW DUMP DESCRIPTIONS

Name	Type	Description
Time	(Simulation)seconds	The time step described by the values within this timestep-element
Id	Id	The id of the edge/lane/vehicle
Pos	Metres	The position of the vehicle at the lane within the described time step
Speed	Metres/seconds	The speed of the vehicle within the described time step

From Figure 3, it can be seen that at time step 22630 the bus is at position 506.26 metres of edge #23450698 and lane #23450698_0. As can be seen from the static road network data in Figure 4, the total length of lane 23450698_0 is 671.38 metres. Thus the total distance travelled by the bus up to this point can be calculated by the lengths of the edges and lanes travelled so far plus the current position in the current edge/lane. This data is then used together with the Kalman model.

C. Linking SUMO and Kalman Filter Model

To link SUMO and the KF model, we integrate the components discussed in previous sections in Matlab. The outcome from the visualisation can then be plotted to produce the end result. A bus journey which has been modelled in SUMO is utilised to feed data into and also verify the results from the KF Model.
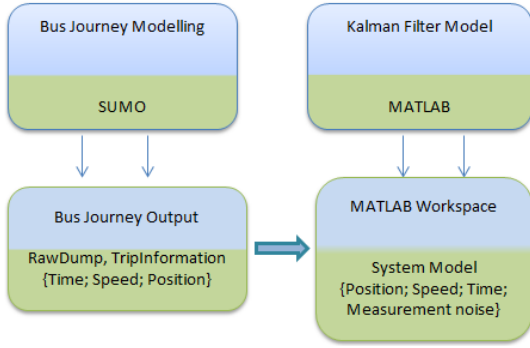
Figure 5. Integration between SUMO and Kalman Filter Model

Journey information from SUMO will be used with the Kalman Filter. Our KF Model has been designed to correspond to the journey programmed in SUMO.

The Kalman model is designed to produce an estimate of position based on the vehicle speed. Bus position and speed are the data items from the SUMO simulation. We configure the Kalman Model as follows:

$$x = \begin{pmatrix} Position \\ Speed \end{pmatrix} \quad (1)$$

We then set the system model as follows:

$$x_{k+1} = \alpha x_k \quad (2)$$

$$y_k = \mu x_k + z_k \quad (3)$$

$$\alpha = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, \quad \mu = (0|1) \quad (4)$$

x_k represents the state variable, y_k the measurement variable, α the state transition matrix, μ the state measurement matrix, and z_k is the measurement noise.

α describes how the system changes over time. Δt is the interval of the time measuring position. Equation 2 predicts the state at the next timestep for system model α. Subscript k clarifies KF model is executed in a recursive manner. The state measurement matrix μ is used in Equation 3 predicting the position based on measured velocity.

IV. KALMAN FILTER RESULTS AND DISCUSSION

In this section we present the results from a comparison of estimation accuracy of Linear KF, Extended KF (EKF), and Unscented KF (UKF). Linear and non-linear KF models each have their own advantages [20]. No general statement can be made as to which is best, because it depends on the particular situation at hand. Comparison of different types of KF models, including conventional linear, and various types of non-linear KFs (which are essentially variants of extended KFs linearizing nonlinear assumptions), is vital to identify the optimal model for estimation of arrival time.

This section presents our initial experimentation with our integrated KF-SUMO approach to estimate the distance travelled of a bus. We implemented the bus journey in SUMO, which then provided both the final time estimates for comparison, and also the step by step journey data that is used with our KF approaches.

There are 2 sets of figures presented in this section, with each corresponding to a programmed journey. Fig. 5 to 8 depict data on a journey which completes as planned. No unforeseen events such as an accident occur. Fig. 9-12 correspond to the same journey, but with an incident occurring in the 10th minute. The length of the journey is 3075 metres. Under normal conditions, the journey takes around 13 minutes, with the bus travelling on average 22 metres/minute. Process noise during the journey (delays caused by traffic lights, roadwork etc.) is set at 10 metres/minutes.
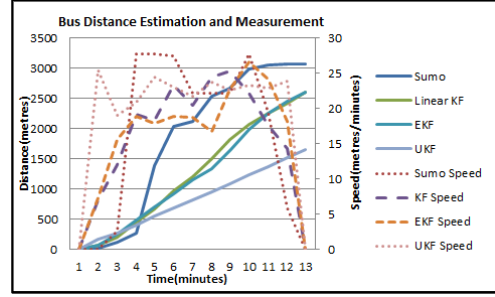

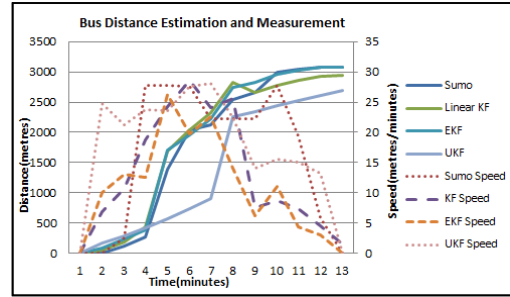Figure 6. Results without data exchange between KF and SUMO

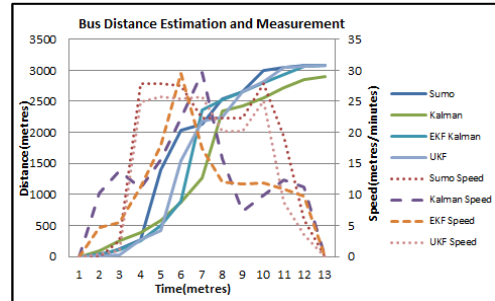Figure 7. Data Exchanges between KF and Sumo every 6 minutes

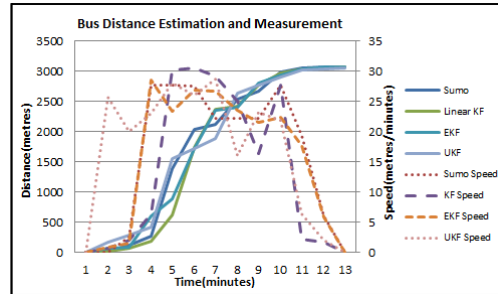Figure 8. Data exchanges between KF and Sumo every 3 minutes

Figure 9. Data exchanges between KF and Sumo every minute

A. Incident Free Journey

Fig. 6 - 9 show results from the incident free journey. Solid lines represent the estimation of the distance travelled using the left Y axis. The dashed line indicate the vehicle speed using the right Y axis.

In Fig 6, we can observe considerable differences between the predictions by all the KF algorithms and the SUMO measures. At the end of the SUMO measures, the closest KF estimate the distance travelled at just over 2500m. Clearly, this needs improvement.

Fig. 7 and 8, show results when SUMO measures are used to update the KF predictions at regular intervals during the journey. Fig. 7 shows results with the KF models loading SUMO journey data every 6 minutes. Although there is still considerable difference between the KF models results and the Sumo measurements (e.g. min 4 to min 9), by the end of the journey the KF and SUMO results are much closer than in Fig. 6. Results shown in Fig 8 for a data exchange interval of 3 minutes, indicate a further improvement of the estimation results of all KF models.

Finally, results in Fig. 9 show that data input to the KF models every minute, the KF results are very similar to the actual SUMO measurement. The Kalman predicated velocity is quite close to the curve of SUMO velocity. Thus additional data input into KF improve predictions considerably. More frequent information feeds will produce better results.

B. Journey with Traffic Incident

Similarly to the previous section, Fig. 10 - 13 show the prediction results from the KF models together with the measures from SUMO. However, now a traffic incident occurs. The vehicle comes to a complete stop for a few minutes causing a delay in the arrival time. The distance travelled overall is identical to the experiments in the previous section. The incident represents a challenge to arrival time prediction using KF models. The drop in speed and lack of movement can be seen in Fig. 10-13, starting at min 11, with the incident cleared in min 17.

In Fig. 10, where the KF models alone is being used with no updated journey information, there is a large discrepancy between the Kalman predictions and SUMO results. At the end of the journey, the KF models estimates that the bus travels 2049 metres, which is significantly lower than the actual SUMO results.

Fig. 11 then shows the KFs results compared to SUMO measures with data fed into the KF models every 6 minutes. These results show that although the discrepancy between KFs and SUMO is smaller, the velocity estimation is still rather noisy. These results are improved in Figure 12, where the data exchange interval is increased to 3 minutes. Now the velocity prediction in the Kalman models increases during minutes 1 to 6, and then reduces from just under 30 metres/minutes to close to zero due to the incident. The distance prediction is much more accurate due to receiving information from SUMO during the journey.

Finally, in Fig. 13 the KF models receive information every minute. The KF velocity prediction is very similar in terms of behaviour to the SUMO results. The KF predictions

of distance travelled closely follows the SUMO measurements throughout the journey. Clearly the more frequent the information can be made available to the KF model, the more accurate are its predictions. This is especially important when unforeseen events occur during a journey.
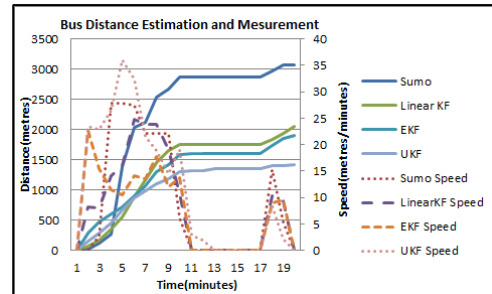

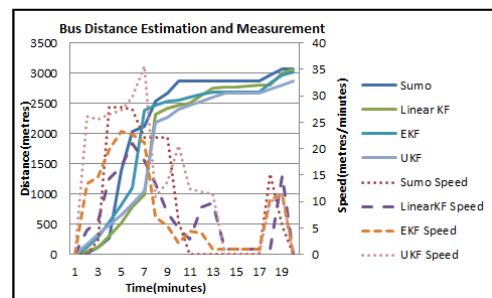Figure 10. Results without data exchange between KF and SUMO

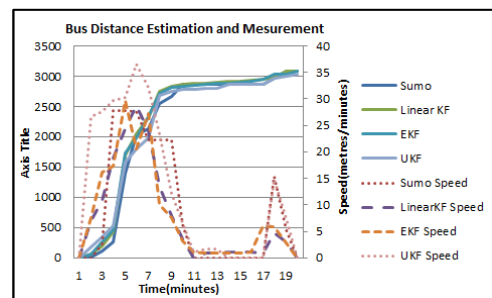Figure 11. Data exchanges between KF and Sumo every 6 minutes

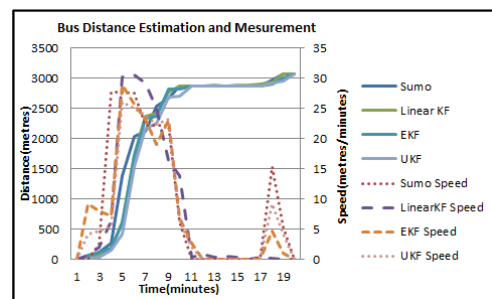Figure 12. Data exchanges between KF and Sumo every 3 minutes

Figure 13. Data exchanges between KF and Sumo every minute

From all the results presented in this paper, no particular Kalman Filter algorithm stands out in terms of its accuracy. Perhaps the linear and EKF estimations are closest for the cases where no SUMO measures assist the Kalman prediction. The EKF model is quite similar to the linear KF model However, EKF models uses nonlinear system model equations $f(x_k)$ and $h(x_k)$ in place for αx_k and $\mu\ x_k$ in the linear model [20]. Secondly, the matrices α and μ are Jacobians of the nonlinear model [20].

While the UKF deviates slightly more than the linear and EKF models, all three Kalman models can be utilized to estimate bus arrival time.

V. Conclusions

In this work, we integrated journey information obtained from a SUMO simulation into a Kalman Filter models in order to improve the accuracy of bus journey arrival time predictions. We have successfully demonstrated the integration of the two techniques. This provided us with an experimental platform which allowed us to check the Kalman predictions against SUMO measures. Furthermore, we have shown that inserting additional journey data obtained from SUMO in the Kalman prediction can greatly increase its accuracy. Regular information updates can further improve the accuracy.

For the predictions we have employed different types of Kalman Filter models: Linear, Extended, and Unscented Kalman Filters. No single approach offers significantly better predictions than others with the Linear and EKF models having slight advantages of the UKF. As linear models are the most straightforward approaches, for scenarios as described in this paper, these models are sufficient and offer good performance.

Future work is therefore to consider the extension of this approach to make use of social network data in order to have access to up-to-date traffic and road condition information.

REFERENCES

[1] Zhu, Tongyu, Jian Dong, Jian Huang, Songsong Pang, and Bowen Du. "The bus arrival time service based on dynamic traffic information." 6th Intl Conf Application of Information and Communication Technologies (AICT), pp. 1-6. IEEE, 2012.

[2] Williams, Billy M., and Lester A. Hoel. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results." J of transportation engineering 129, no. 6 (2003): 664-672.

[3] Li, Shu-Bin, Bai-Bai Fu, and Wen-Xiu Dang. "Dynamic Analyses of Urban Expressway Network with Microscopic Traffic Flow Model Integrated Variable Speed Limits." J Nonlinear Dynamics 2014.

[4] Nath, Rudra Pratap Deb, et al. "Modified K-means clustering for travel time prediction based on historical traffic data." Knowledge-Based and Intelligent Information and Engineering Systems. Springer, 2010. 511-521.

[5] H. Liu, et al. "Predicting urban arterial travel time with state-space neural networks and Kalman filters." J Transp.Research Record: (2006): 99-108.

[6] Y. Wang, M. Papageorgiou. "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach."Transportation Research Part B: Methodological 39.2 (2005): 141-167.

[7] Li, Li, Xiqun Chen, and Lei Zhang. "Dept. of Automation, Tsinghua University, Beijing 100084, China." (2014): 1-14.

[8] Altinkaya, Mehmet, and Metin Zontul. "Urban Bus Arrival Time Prediction: A Review of Computational Models." Journal of Recent Technology and Engineering (IJRTE)(2013):2.4

[9] Huang, Yanguo, et al. "Urban Expressway Travel Time Prediction Method Based on Fuzzy Adaptive Kalman Filter." Appl. Math 7.2L (2013): 625-630.

[10] Simon, Dan. "Using nonlinear Kalman filtering to estimate signals." Embedded Systems Design 19.7 (2006): 38.

[11] Chien, Steven I-Jy, and Chandra Mouly Kuchipudi. "Dynamic travel time prediction with real-time and historic data." Journal of Transportation Engineering 129.6 (2003): 608-616.

[12] Chien, Steven I-Jy, Yuqing Ding, and Chienhung Wei. "Dynamic bus arrival time prediction with artificial neural networks." J Transportation Engineering128.5 (2002): 429-438.

[13] Hu, T.Y.; Chen, L.W. (2008). Real-time implementation of simulation-based Dynamic Traffic Assignment model. 11th IEEE Conf. on Intelligent Transportation Systems, 645-650.

[14] Ta-Yin Hu, and Li-Wen Chen. "Real-Time Implementation of Simulation-based Dynamic Traffic Assignment Model", 11th IEEE Conf Intelligent Transportation Systems, 2008.

[15] A.F. Abidin, M. Kolberg, and A. Hussain. "Improved Traffic Prediction Accuracy in Public Transport Using Trusted Information in Social Networks." 7th York Doctoral Symposium on Computer Science & Electronics, p. 19. 2014.

[16] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker. "Recent development and applications of SUMO simulation of urban mobility." J On Advances in Systems and Measurements 5, no. 3 and 4 (2012): 128-138.

[17] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz. "Sumo-simulation of urban mobility-an overview." 3rd Intl. Conf. on Advances in System Simulation, pp. 55-60. 2011.

[18] S. Uppoor and M. Fiore. "Large-scale urban vehicular mobility for networking research." IEEE Vehicular Networking Conference (VNC), pp. 62-69. IEEE, 2011.

[19] http://kolntrace.project.citi-lab.fr/#trace

[20] Simon, Dan. "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms." *IET Control Theory & Applications* 4.8 (2010): 1303-1318.