Analysis and Modeling of Transit Passenger Arrivals to a Bus Terminal using a Doubly Stochastic Model

Abdullah Kurkcu, Ph.D.
Research Associate, C2SMART Center (A Tier 1 USDOT UTC),
Department of Civil and Urban Engineering &
Center for Urban Science and Progress (CUSP),
Tandon School of Engineering, New York University (NYU)

Kaan Ozbay, Ph.D. (Corresponding author)
Professor & Director, C2SMART Center (A Tier 1 USDOT UTC),
Department of Civil and Urban Engineering &
Center for Urban Science and Progress (CUSP),
Tandon School of Engineering, New York University (NYU)

1 Introduction and Background

The frequency of transit services affects the behavior of users in deciding when and how to travel. The primary objective of creating a user-centric public transit service is to reduce passenger wait times and overcrowding at stations [1], which requires an in-depth understanding and accurate modeling of arrival patterns of passengers during both peak and off-peak periods. The introduction of smart card-based automated fare collection (AFC) systems has recently allowed many researchers to analyze spatio-temporal journey information at an individual level to understand passenger arrival behavior [1-8]. When such AFC systems are not readily available, emerging IoT devices can address the lack of detailed passenger demand data by collecting wireless traces of the riders' smart devices. Recent studies have shown that low-cost and ubiquitous sensors can be used as a viable alternative to traditional sensors to collect travel time, count, and origin-destination data when fixed infrastructure-sensors are not readily available and online transportation data sources lack the level of detail required for specific applications [9-11]. These IoT devices that can detect pedestrians by sniffing Wi-Fi and Bluetooth signals are coined as Human Movement Detection (HMD) sensors in this study.

In this paper, we used the data collected by these low-cost, ubiquitous, customized HMD sensors to understand passenger arrival behavior and then developed arrival models that can replicate stochastic variations of passenger demand. Several studies in the past have used time-dependent distributions to model time-varying transit passenger arrival behavior in transportation networks [12-15]. This study uses a specialized nonhomogeneous arrival process called modulated renewal process event model to capture sudden changes in arrival intensity with time. The modulated renewal model is enhanced by Lasko [16] to improve its efficiency and flexibility over other models by introducing a new inference method. The models used to infer arrival intensities in the literature are either intractable on long-duration observations of high-resolution events or fail to explain bursty event streams [16]. Intractable inference models require thinning to retrieve intensity functions. The direct inference method by Lasko [16] is two orders of magnitude more efficient than thinning. It also can efficiently model bursty, memoryless or regular events that thinning models fail to do.

The individual passenger data used in this paper is collected through low-cost HMD sensors. It is important to emphasize the fact that our sensors can capture 25 percent of overall transit arrivals because it detects transit riders who carry smart devices. From the initially collected data, we observed that arrivals can occur at intervals in short sudden episodes and show bursty behavior, especially for off-

peak periods [17]. The modulated renewal process is capable of efficiently modeling both bursty and regular events. The proposed model may be validated by installing these sensors to that the AFC or ground truth data is readily available. The validation of the market penetration of smart devices and the evaluation of the passenger counts using the sensors throughout the day were executed in a previous study [10].

2 Data

HMD sensors collected data for a 6 week period in the transit terminal. These sensors can also detect non-mobile devices, thus, filtering algorithms prescribed in [10] are used to clean the collected data. In the post-processing phase, only MAC addresses belonging to mobile devices are analyzed to retrieve a more reliable subgroup of the whole passenger population. The data is extrapolated to represent the total population using a fixed 25% market penetration rate of discoverable devices based on the previously collected manual data collection experiments. The market penetration rate is also confirmed by the following calculation. We assumed that all the passengers are cleared from the platform at the end of the peak period. Then, it is possible to estimate the number of passengers arriving at the transit terminal using the number of departed buses in the peak period. Dividing the number of sensor detections by this estimated total arrival number confirms the market penetration rate to be around 25%. Figure 1 is an illustration of temporal differences in arrival behavior. It shows the density plots of passenger arrival probability against the time that has passed since the last bus departure.
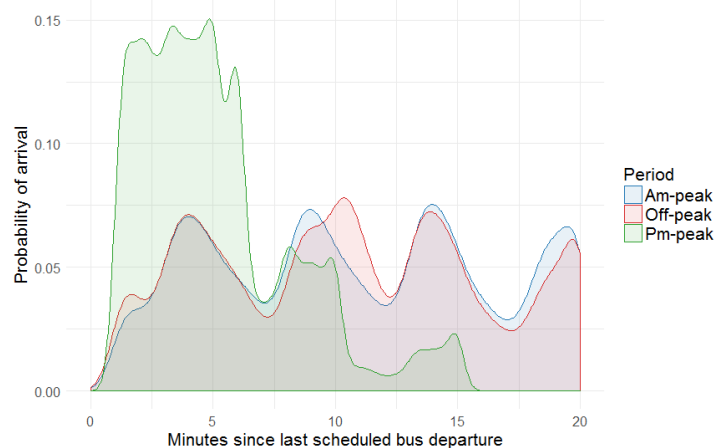
Figure 1: Distribution of Passenger Arrivals

Most transit riders use the bus terminal to commute to work and leave the city after working hours. During the PM peak period, the average bus headway is shorter, creating the expectation that passengers will dismiss the schedule information and arrive randomly. The sensor data confirm this expectation, showing that the arrival probability is relatively consistent under 6 minutes in the PM period. It is also expected that the majority of passengers will arrive closer to the scheduled bus departure time when the headway is longer. Although these probabilities provide general information about the passenger arrival behavior, more detailed analysis of the arrival intensities is required to understand the time-of-day and headway effects. More detailed analysis of the bursty arrival behavior and arrival probabilities using the same dataset for various peak and off-peak periods are provided in [17]. The next section explains the methodology proposed to infer the arrival intensity curve from the sampled data.

3 Methodology

A doubly stochastic model is a type of model in which an observed random variable is modeled in two steps. In the first step, the distribution of the outcome is represented in one of the standard probability distributions with one or more parameters. In the second step, these parameters are treated as being random variables and are modeled using stochastic processes or time-series.

A gamma process with log intensity functions drawn from a Gaussian process is used to model the arrival intensity during the off-peak period in this paper. This model is also called a modulated renewal process event model, which assumes that the inter arrivals are independent and identically distributed. Let's assume a set of event times, forming an event stream T, and this event stream can be modeled by using a modulated renewal process as follows:

$$P(T;\alpha,\lambda(t)) = \frac{1}{\Gamma(\alpha)^n} \prod_{i=1}^{n} \lambda(t_i)(\Lambda(t_i) - \Lambda(t_{i-1})^{\alpha-1} e^{-\Lambda(t_n)}) \tag{1}$$

Where $\Gamma(\cdot)$ is the gamma function, $\alpha > 0$ is the shape parameter, $\lambda(t) > 0$ is the modulating intensity function, and $\Lambda(t) = \int_0^t \lambda(u)du$. The equation above is actually a homogenous gamma process, which is a random process with independent gamma-distributed increments. If we replace $1/\lambda(t)$ with a constant b, which models the inter arrivals as:

$$\gamma(x|a,b) = P(x;a,b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} \tag{2}$$

$$x_i = t_i - t_{i-1} \qquad i = 1\ldots$$

The function $\Lambda(t)$ squashes the event times t_i into a new space where their inter-arrival intervals (x) become draws from the homogeneous gamma process. A gamma process is selected for simplicity without loss of generality because it models the relationship between neighboring arrivals, instead of assuming them to be independent or memoryless. The log intensity function is modeled as a draw from a Gaussian process prior with zero mean and the covariance function as:

$$C(t_i, t_j) = \sigma e^{-\left(\frac{t_i - t_j}{l}\right)^2} \tag{3}$$

where σ adjusts the magnitude and l adjusts the time scale of the Gaussian process. To make the final results smoother, the squared exponential covariance function is used.

In this paper, two scenarios are analyzed at the bus terminal, where the data was collected. One is the condition in which the passenger arrivals are assumed to follow a Poisson arrival process with a fixed arrival intensity. This is called "fixed arrival intensity scenario" hereafter. The other is the scenario in which the arrival intensity varies with time, and this will be called "varying arrival intensity scenario" henceforth.
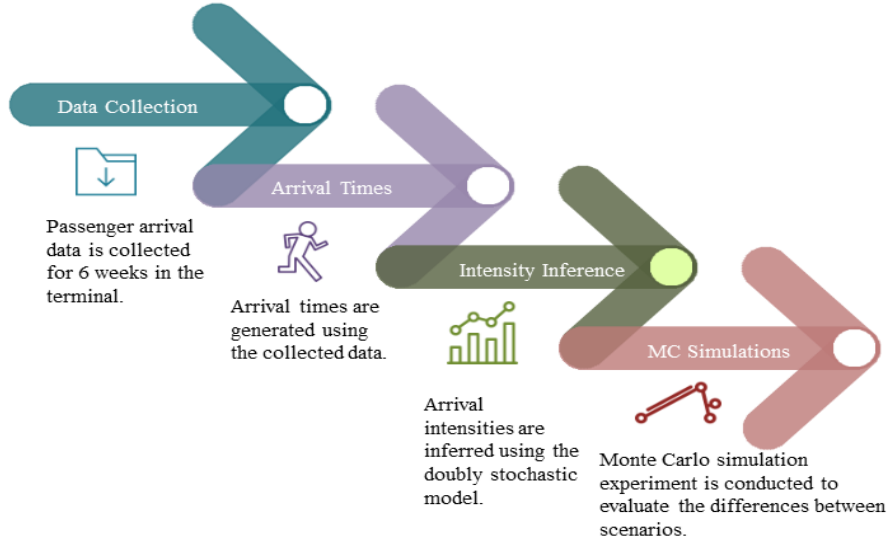
Figure 2: The Methodology of Modeling Transit Passenger Arrivals

Figure 2 shows the overall methodology proposed in this paper. First, observed arrival times are generated using the collected data in the transit terminal. Each unique arrival is recorded with a time-stamp. Then, arrival times are used as inputs for the arrival intensity inference for the doubly stochastic model. The inference method reveals the hidden arrival intensities and returns inferred variables that can be used to generate arrival times. To better understand the differences between arrival intensity functions, Monte Carlo simulations are run for both the fixed and varying intensity scenarios.

The MCMC process is used to simultaneously infer posterior distributions over the intensity function and its parameters given a set of event times. The gamma shape parameter is sampled using Metropolis-Hasting updates. The Gaussian process with zero mean and covariance function C, GP(0,C), is denoted as f in the algorithm below. Rather than updating "f" at n locations, this algorithm only uses "k" uniformly spaced points. Lasko [16] found out that k=200 works fine for nearly all of their data examples, regardless of the study period. The MCMC algorithm used in this study is given below:

MCMC Algorithm

Input: Arrival times T, Time grid $\hat{T}$, function f and parameters σ, l and α

Output: Updated $f, \lambda, \Lambda, \sigma, l$ and α

1. Update $f(\hat{T}), \sigma, l$ using slice sampling
2. Compute $f(T)$ by smooth interpolation of $f(\hat{T})$
3. $\lambda(T \cup \hat{T}) \leftarrow e^{f(T \cup \hat{T})}$
4. Compute $\Lambda(T)$ from $\Lambda(\hat{T})$ numerically
5. Compute $p = P(T; \alpha, \lambda(T))$ using Equation 1
6. Update the shape parameter and the likelihood with Metropolis Hastings

---

4 Results and Discussion

To reveal modulated renewal process variables, the MCMC inference is conducted using 2000 burn-in and 2000 inference iterations in roughly 5 minutes using the MATLAB code provided by Lasko [16].

The threshold number of events is selected to be 10. This threshold establishes that the method requires more than 10 events to perform the inference. Using different weeks makes it possible to better capture the variation in the arrival intensity and its shape in the peak and off-peak periods. Figure 3 shows the inferred intensity function in red and 95% confidence intervals in blue. The inferred curves can be considered reasonable given that about 95% of varying intensities between different weeks are contained within them. The interpretation of these intensity functions is that there is a clear transition from the off-peak period to a peak period around 4 PM. In addition, the arrival intensity follows a similar pattern on Wednesdays among different weeks.
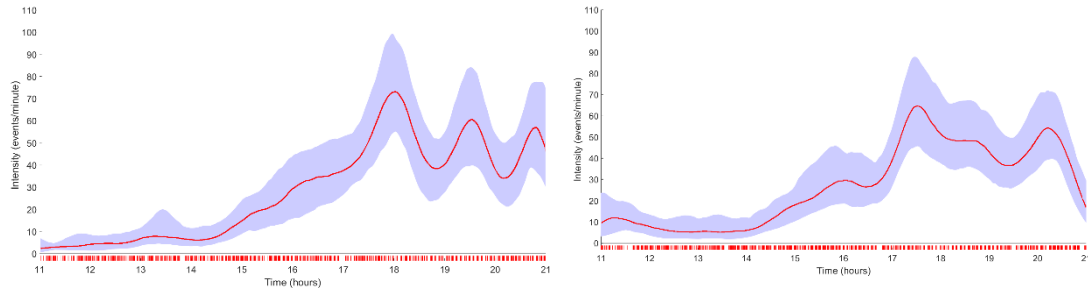

Figure 3 Intensity functions for Wednesdays

To better illustrate the effect of using varying arrival intensity, we conducted 2000 Monte Carlo simulations to compare the arrivals generated by using both traditional homogeneous Poisson process and proposed modulated renewal process. The mean queue length and number of arrival events are calculated and stored for each of the 2000 runs. The queue length refers to the number of people queued up waiting for the next bus within 5 minutes. 5-min intervals are selected because the average headway is 5 minutes for the selected peak period.

The distributions of queue lengths from both methods are shown in Figure 4. While the mean values of simulated queue lengths obtained from the modulated renewal method and Poisson process are similar, 45 and 47 respectively, their distributions are quite different. Figure 4 highlights the inadequacy of just mean value analysis when modeling transportation systems such as this transit terminal where a group of passenger arrivals can suddenly occur during short periods. The more realistic distribution of queue lengths and arrivals obtained from the modulated renewal method can provide a much more accurate understanding of the state of the system. Our approach also provides additional information which is not available from the Poisson process about the infrequent occurrences of bursty arrivals by considering the tails of the distribution.
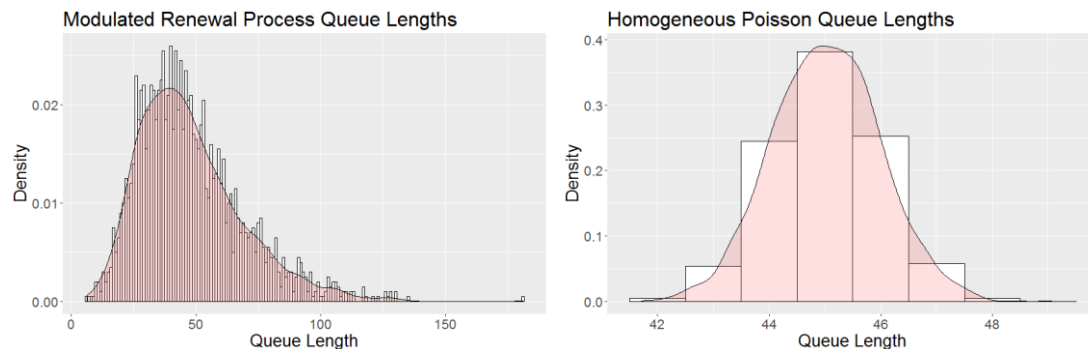
Figure 4: Queue Length Distributions: Non-Homogenous (Modulated Renewal Process) vs. Homogeneous (Poisson Process) Processes

Figure 5 shows the number of simulated arrivals using both methodologies. The average number of arrivals using the modulated renewal process is found to be 2345 passengers during the peak period. The Poisson process generated 2161 arrivals on average. Here, not only does the average number of arrivals differ significantly, but their distributions also show a different behavior under bursty conditions. Homogeneous Poisson process arrivals vary around the mean and do not exceed 2300. On the other hand, the non-homogenous process results in arrivals that can reach up to 6000 during the peak period. As seen from Figure 6, homogeneous Poisson process could not generate these extremely high but rather infrequent arrivals observed in real-world whereas the proposed modulated renewal process proposed in his paper was capable of capturing this particular bursty behavior. The differences between methodologies are summarized in Table 1.
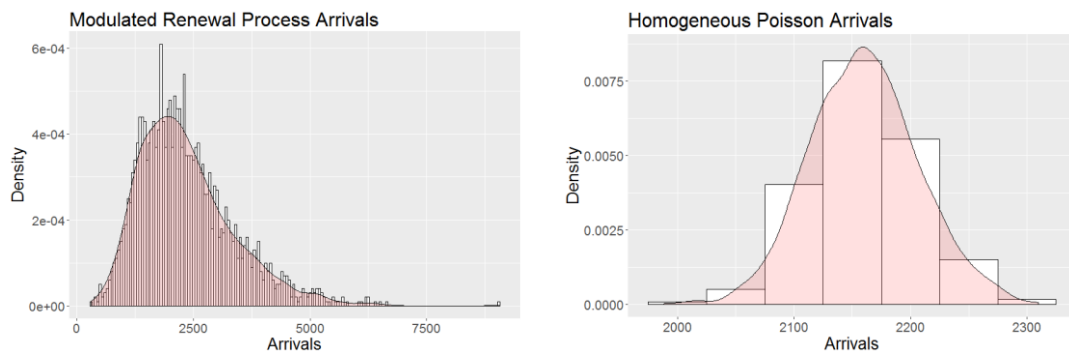
Figure 5: Arrival Distributions: Non-Homogenous vs. Homogeneous Processes

Generating the continuous function densities based on the observed data over a long period of time not only makes the in-depth study of arrival behavior in the time domain possible but also enables longitudinal analysis of arrival curves in time. In addition, with this approach it is possible to simulate the crowd density and queue lengths if the capacity of the station is known.

Table 1: Comparison of Simulated Queue Lengths and Arrivals

	Queue Length				# of Arrivals			
Arrival Process	Min	Median	Mean	Max	Min	Median	Mean	Max
Modulated Renewal	6.3	44.2	47.8	180.9	302	2172	2346	6068
Poisson	41.7	45.0	45.0	49.1	2002	2161	2193	2355

5 Conclusions and Future Work

The main objective of this study was to carefully study the actual arrival behavior of transit users to a major transit station in New York City. However, to acquire this kind of extensive arrival data at multiple locations in the transit station, there is a need to collect long-term disaggregate passenger arrival data. We first developed low-cost and portable Human Movement Detection (HMD) sensors that are capable of sniffing Wi-Fi and Bluetooth signals from mobile devices [9, 10, 17, 18]. These HMD sensors were deployed by the transit operator in the transit station at different key locations. Then, we proposed a novel methodology to relate passenger data collected by these HMD sensors to scheduled bus departures to study time-varying passenger arrival behavior. The sensor that recorded the most foot traffic at one of the bus gates is selected for conducting the analysis presented in this paper.

As a result of extensive analysis of this individual passenger arrival data by HMD sensors, we were able to show the need for the use of a nonhomogeneous arrival process called modulated renewal process event model instead of fixed intensity arrival function, such as a regular Poisson process to accurately model transit passengers' queue lengths and arrival behavior. If the arriving behavior is indeed a doubly stochastic process in which the intensity varies with time, then homogeneous Poisson-based arrival model estimations may result in inaccurate arrival numbers and queue length distributions.

The direct inference method of the arrival intensity makes it possible to transform discrete arrival events into a continuous function. One of the useful features of the modulated renewal process is that it enables a straightforward arrival generating mechanism that can be used both for simulating and predicting events. It provides a function to generate posterior predictive realizations of point patterns and examine the number of arrivals, the time between arrivals and the total duration of them. This proposed modeling approach can efficiently model bursty arrival behavior and then be used to better understand the differences between days, weeks, and months and to simulate alternative scenarios or predict future demand more accurately.

For stations at which AFC systems are not available, HMD sensors can be easily deployed because they are simple to operate, easily transported, and require minimal maintenance. Thus, many transit agencies with limited budgets can adopt this kind of low cost and ubiquitous sensors and collect data needed for better planning of bus and train operations as well as long-term planning of terminals, stations and implementing gate improvements to reduce wait times and overcrowding. In the future, there is a need to study more transit locations with different arrival characteristics and also capture the impact of inter-modal interactions. It is also probable that bursty behavior of arrivals might be due to unusual traffic demand patterns, significant events or adverse weather conditions. As part of our future research, these various external factors will be evaluated to understand underlying reasons of stochasticity affecting arrival functions.

References

[1]   L. Sun, J. G. Jin, D.-H. Lee, K. W. Axhausen, and A. Erath, "Demand-driven timetable design for metro services," *Transportation Research Part C: Emerging Technologies,* vol. 46, pp. 284-299, 2014.

[2]   J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis," *Transportation Research Part C: Emerging Technologies,* vol. 90, pp. 292-306, 2018.

[3]   F. Zhou, J.-g. Shi, and R.-h. Xu, "Estimation method of path-selecting proportion for urban rail transit based on AFC data," *Mathematical Problems in Engineering,* vol. 2015, 2015.

[4]   D. Csikos and G. Currie, "Impacts of transit reliability on wait time: insights from AFC data," in *Transportation Research Board 86th Annual MeetingTransportation Research Board*, 2007.

[5]   J. Zhao, M. Frumin, N. Wilson, and Z. Zhao, "Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data," *Transportation Research Part C: Emerging Technologies,* vol. 34, pp. 70-88, 2013.

[6]   D. Csikos and G. Currie, "Investigating consistency in transit passenger arrivals: insights from longitudinal automated fare collection data," *Transportation Research Record: Journal of the Transportation Research Board,* pp. 12-19, 2008.

[7]   Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data," *Transportation Research Record: Journal of the Transportation Research Board,* pp. 58-67, 2012.

[8]   A. Chakirov and A. Erath, "Use of public transport smart card fare payment data for travel behaviour analysis in Singapore," 2011.

[9]   N. Shlayan, A. Kurkcu, and K. Ozbay, "Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, 2016, pp. 229-234.

[10]   A. Kurkcu and K. Ozbay, "Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors," *Transportation Research Record: Journal of the Transportation Research Board,* pp. 72-82, 2017.

[11]   K. Ozbay, N. Shlayan, H. Nassif, A. Kurkcu, S. Demiroluk, M. D. Maggio*, et al.*, "Real-Time Estimation of Transit OD Patterns and Delays Using Low Cost-Ubiquitous Advanced Technologies," 2017.

[12]   H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological,* vol. 76, pp. 117-135, 2015.

[13]   J. Yin, T. Tang, L. Yang, Z. Gao, and B. Ran, "Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach," *Transportation Research Part B: Methodological,* vol. 91, pp. 178-210, 2016.

[14]   G. Qi, G. Pan, S. Li, Z. Wu, D. Zhang, L. Sun*, et al.*, "How Long a Passenger Waits for a Vacant Taxi--Large-Scale Taxi Trace Mining for Smart Cities," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 2013, pp. 1029-1036.

[15]   D. Morgan, H. Koutsopoulos, and M. Ben-Akiva, "Simulation-based evaluation of advanced public transportation systems," in *Schedule-Based Dynamic Transit Modeling: theory and applications*, ed: Springer, 2004, pp. 95-112.

[16]   T. A. Lasko, "Efficient inference of Gaussian-process-modulated renewal processes with application to medical event data," in *Uncertainty in artificial intelligence: proceedings of the... conference. Conference on Uncertainty in Artificial Intelligence*, 2014, p. 469.

[17]   A. Kurkcu and K. R. Kaan Ozbay, "Investigating Transit Passenger Arrivals using Wi-Fi and Bluetooth Sensors," 2017.

[18]   K. Ozbay, A. Kurkcu, and H. Yang, "Portable and Integrated Multi-Sensor System for Data-Driven Performance Evaluation of Urban Transportation Networks," 2018.