Using Artificial Intelligence to Improve Traffic flows, with Consideration of Data Privacy

September 2022

A Research Report from the Pacific Southwest Region University Transportation Center

Gwen Shaffer, Ph.D., College of Liberal Arts, California State University - Long Beach (CSULB)

Hossein Jula, Ph.D., College of Engineering, CSULB

Anastasios Chassiakos, Ph.D., College of Engineering, CSULB

Tyler Reeb, Ph.D., College of Professional & International Education, CSULB

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
PSR-20-13	N/A	N/A
4. Title and Subtitle **Using Artificial Intelligence to Improve Traffic flows, with Consideration of Data Privacy**		**1/31/2022** **6. Performing Organization Code** N/A
7. Author(s) Anastasios Chassiakos, 0000-0001-9514-0812 Gwen Shaffer, 0000-0001-6425-960X Hossein Jula, 0000-0002-6380-1839 Tyler Reeb, 0000-0002-8056-9939		**8. Performing Organization Report No.** TBD
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (8/2020 – 1/2022) **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes N/A		

16. Abstract

This project develops an artificial neural network (ANN), a class of Artificial Intelligence (AI) systems, to accurately model and predict future delays at an intersection. Developing such modeling and prediction systems raises considerable data privacy concerns and it is incumbent upon municipal, state, and federal branches of government to prioritize citizens and their concerns before the implementation of new smart community technologies that are fueled by unprecedented levels of data collection. The technique proposed in this study identifies nonlinear, time-varying mapping between the inputs to the ANN and its output, the predicted delay. The traffic data measured at a Long Beach intersection with heavy truck traffic are used to build a realistic simulation in Vissim, a microscopic traffic flow simulator. We designed and performed experiments on the developed Vissim model to train the ANN delay predictor and validate the generalization ability of the predictor. The simulation results agree with our on-site delay measurements. This suggests the ANN predictor can accurately predict the delay at the intersection with heavy-truck penetration.

Because smart technologies raise data privacy concerns, the research team led 32 study participants on "datawalks" designed to gauge comfort levels and attitudes toward devices that collect personally identifiable information. Study participants encountered public WiFi routers, surveillance cameras, automated license plate readers and other surveillance technologies. They used a custom app to respond to prompts related to data collection, sharing and analysis. Study participants' responses, along qualitative data collected during a "debriefing" conversation following each walk, provided insights into residents' attitudes toward smart communities technologies and identified privacy concerns.

The quantitative and qualitative findings in this study inform a series of recommendations that research teams can follow to implement real-world test labs at busy truck intersections while fostering public trust, installing these modelling and prediction systems, and ensuring the overall safety and efficiency of the intersection's traffic flow.

<table>
<tr><td colspan="2">17. Key Words
Data privacy; smart city technologies; qualitative methods; traffic delay prediction; artificial neural network, Vissim intersection simulation</td><td colspan="3">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td colspan="2">20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
72</td><td>22. Price
N/A</td></tr>
</table>

Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
CA	California
EB	East bound
HGV	Heavy goods vehicle
NB	North bound
NN	Neural network
PCH	Pacific Coast Highway (located in Long Beach, CA)
POLB	Port of Long Beach
SB	South bound
TEU	Twenty-foot equivalent unit
WB	West bound

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

California Department of Transportation (CALTRANS) Disclaimer

Disclosure

Principal Investigator Gwen Shaffer, Co-Principal Investigators Hossein Jula, Anastasios Chassiakos, Tyler Reeb others, conducted this research titled, "Using Artificial Intelligence to Improve Traffic flows, with Consideration of Data Privacy" at the Department of Electrical Engineering and the Department of Journalism and Public Relations, California State University Long Beach. The research took place from August 2020 to January 2022 and was funded by a grant from the U.S. Department of Transportation in the amount of $100,000. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

This research could not have been possible without enormous help and contributions from our graduate and undergraduate students:

- John Krallman, undergraduate, Electrical Engineering
- Christopher Rosas, undergraduate, Electrical Engineering
- John Lee, undergraduate, Electrical Engineering
- Nicole Balisi, undergraduate, Electrical Engineering
- Martin Garcia, undergraduate, Electrical Engineering
- Wyatt Garcia, undergraduate, Electrical Engineering
- Sheila Hayati, undergraduate, Math
- Jared Paz, undergraduate, Finance
- Jenny Benitez, undergraduate, Economics
- Adam Doyle, graduate, GIS
- Angelina Carballo, graduate, GIS
- Alondra Garcia, graduate, GIS

Abstract

This project develops an artificial neural network (ANN), a class of Artificial Intelligence (AI) systems, to accurately model and predict future delays at an intersection. Developing such modeling and prediction systems raises considerable data privacy concerns and it is incumbent upon municipal, state, and federal branches of government to prioritize citizens and their concerns before the implementation of new smart community technologies that are fueled by unprecedented levels of data collection. The technique proposed in this study identifies nonlinear, time-varying mapping between the inputs to the ANN and its output, the predicted delay. The traffic data measured at a Long Beach intersection with heavy truck traffic are used to build a realistic simulation in Vissim, a microscopic traffic flow simulator. We designed and performed experiments on the developed Vissim model to train the ANN delay predictor and validate the generalization ability of the predictor. The simulation results agree with our on-site delay measurements. This suggests the ANN predictor can accurately predict the delay at the intersection with heavy-truck penetration.

Because smart technologies raise data privacy concerns, the research team led 32 study participants on “data walks” designed to gauge comfort levels and attitudes toward devices that collect personally identifiable information. Study participants encountered public WiFi routers, surveillance cameras, automated license plate readers and other surveillance technologies. They used a custom app to respond to prompts related to data collection, sharing and analysis. Study participants’ responses, along qualitative data collected during a “debriefing” conversation following each walk, provided insights into residents’ attitudes toward smart communities technologies and identified privacy concerns.

The quantitative and qualitative findings in this study inform a series of recommendations that research teams can follow to implement real-world test labs at busy truck intersections while fostering public trust, installing these modelling and prediction systems, and ensuring the overall safety and efficiency of the intersection’s traffic flow.

Using Artificial Intelligence to Improve Traffic Flows, with Consideration of Data Privacy

Executive Summary

Heavy trucks transporting goods between marine terminals and warehouses negatively impact traffic near major ports, particularly at major intersections. Compared to passenger vehicles, trucks have significantly lower acceleration rates, take longer to stop, require longer stopping distances, and have bigger turning ratios. As delays propagate throughout the traffic network, heavy trucks passing through intersections also raise a range of safety and emissions issues. This project develops an artificial neural network (ANN), a class of Artificial Intelligence (AI) systems, to accurately model and predict future delays at an intersection. Developing such modeling and prediction systems raises considerable data privacy concerns and it is incumbent upon municipal, state, and federal branches of government to prioritize citizens and their concerns before the implementation of new smart community technologies that are fueled by unprecedented levels of data collection.

The technique proposed in this study identifies nonlinear, time-varying mapping between the inputs to the ANN and its output, the predicted delay. The inputs to the ANN are the measured traffic flows, as well as the status of the current traffic lights. To build and verify the accuracy of the delay predictor, we identified a Long Beach intersection with significant heavy-truck penetration. We measured traffic data from this intersection on multiple days and at various times—including traffic flow information regarding passenger vehicles and heavy trucks and timing of traffic lights—in conjunction with the geometry of the intersection. The traffic data measured are used to build a realistic simulation in Vissim, a microscopic traffic flow simulator. We designed and performed experiments on the developed Vissim model to train the ANN delay predictor and validate the generalization ability of the predictor. The simulation results agree with our on-site delay measurements. This suggests the ANN predictor can accurately predict the delay at the intersection with heavy-truck penetration.

In addition to heavy truck traffic concerns, we also needed to select an intersection located in a community that raised data privacy issues impacting a range of demographics. Such efforts are part of a larger recognition that smart technologies often collect personally identifiable information, and steps should be taken to ensure these platforms avoid violating personal privacy. For this project, these research questions emerged:

- **RQ1:** What attitudes and comfort levels do residents express regarding smart technologies that capture personally identifiable data?

- **RQ2:** Do residents' attitudes and comfort levels vary, depending on whether the technologies are deployed by the city or private entities?

- **RQ3**: Based on the datawalk findings, what implications exist for smart cities deploying technologies that collect, store, analyze and sometimes share personal data about residents?

To answer these questions, our research team designed and facilitated a 1.5 mile "datawalk" that encompassed a mix of residential streets, a commercial corridor lined with restaurants and stores, a police station, a public high school, an industrial block, and a city park in West Long Beach near the Port of Long Beach. Along the route, study participants walked by traffic cameras, Internet-connected bus kiosks, public WiFi and residential security cameras. The research team recruited 32 study participants. Prior to the walk, study participants downloaded a smart phone data collection app designed by the research team. The app enabled participants to geolocate smart technologies they observed, photograph them, and answer a few brief questions about their comfort levels with each technology. This information, as well as qualitative data collected during a "debriefing" conversation following each walk, provided insights into residents' attitudes toward smart communities technologies and identified their privacy concerns. The data informed recommendations that the City of Long Beach can integrate into future smart technology deployments.

The quantitative and qualitative findings in this study inform a series of recommendations that research teams can follow to implement real-world test labs at busy truck intersections while fostering public trust, installing these modelling and prediction systems, and ensuring the overall safety and efficiency of the intersection's traffic flow.

1. Introduction

Traffic signals are critical for ensuring a fast and smooth traffic flow in the cities. In a survey done by the National Transportation Operations Coalition, traffic signals accounted upwards of 10% of the 295 million vehicle-hours of delay on major roadways [1], [2]. It was also found that reductions to travel times of 25% can lead to commuting times to work and errands reduced by 50 hours per year per person, a 22% overall reduction in greenhouse emissions, and a reduction in fuel consumption of up to 10% [1], [2]. For that reason, decreasing delay caused by traffic signals have been the main goal of city planners and traffic engineers for many years.

One potential methodology to decrease delay at an intersection is using advanced Artificial Neural Networks (ANN). ANNs, a class of Artificial Intelligence (AI) systems, have recently become popular with the vast improvement in their computational capabilities. Many see AI methods in general, and ANNs in particular, as possible tools that can be integrated into traffic signals and further mitigate any delay that may occur. In [3], the authors created a neural network to estimate delay using parameters based on formulas used in traditional models found in the Highway Capacity Manual [4]. Zhen *et. al.* in [5] created a neural network delay estimator in which traffic link volume and capacity are the input parameters.

In this project, we discuss one such possible tool using artificial neural networks. Our delay estimator will differ from previously observed delay estimators in that we will be including heavy good vehicle information as a part of our input parameters.

It should be noticed that there are three main types of control schemes, pretimed, semi-actuated, and fully actuated [6]. Initial electronic traffic control systems were pretimed in which predetermined intervals were programmed into signal controllers. Eventually, cities began to incorporate sensors embedded throughout the intersection which led to actuated systems. Based on the input from the sensors, engineers would create algorithms that would determine what times should each light change. Finally, fully actuated signals were developed with greater number of sensors in which operated with some ability to react to traffic conditions. Fully actuated systems can better respond to dynamic situations and are typically reserved for the busiest intersections. Although the incorporation of sensors has improved intersections within cities, fully actuated signals may be unable to respond appropriately to situations such as vehicle accidents or a large influx of heavy good vehicles. It is worth noting that many intersection control schemes, within a metropolitan area, tend to be semi-actuated. For example, as seen in Figure 1, in the Downtown Long Beach area, most intersections are indeed semi-actuated. The downtown area of Long Beach has 163 traffic signals where only 56 traffic signals are pre-timed [7] with the remaining containing some sort of actuated control.

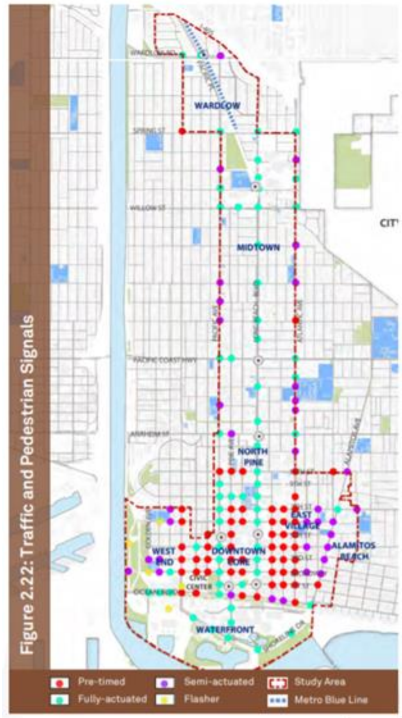


Figure 1. Traffic signal types in downtown Long Beach.

The issues presented within an intersection are the immense number of potential inputs that can be used for a traffic signal controller. Some examples are the number of vehicles present, velocity of the vehicles, weather, time of day, vehicle types, reaction times and pedestrians. With the extensive amount of variability among the inputs, the problem becomes extremely complex for traffic engineers. One such variable that can prove to be significant is heavy goods vehicle (HGV) traffic. HGV are vehicles used to transport freight and have much different dynamics compared to conventional passenger vehicles. The City of Long Beach for example, being a port city, has a large HGV volume percentage upwards of 17.4% [8]. Between 2010 and 2018, trends have shown that HGV increased by 22.9% overall in the US [9]. With the growing trend of HGV use, it is important that we create a traffic signal system that can accommodate such growth.

Creating a model that can accept input parameters such as vehicle volumes of both passenger and HGV and then output cumulative delay would prove to be useful for traffic engineers. Such a model could then predict delay times based on concurrent traffic data. Using a fairly accurate delay estimator, the traffic signal controller would then be able to have a more proactive signal plan rather than a plan that reacts to traffic conditions which can be slow. Utilizing AIs, more specifically ANNs, may help create such a delay predictor for use in traffic signal controllers.

2. Delay and Traffic Flow

The nonlinear nature of traffic makes it extremely difficult to predict and thus to mitigate its negative impacts on our daily life. One of the biggest issues in traffic is the delay produced at intersections. Delay can be used to measure the efficiency of an intersection. An intersection which causes a lot of delay is inefficient and may cause the bottleneck in a transportation network. Conversely, an efficient intersection, will control the flow of traffic with less induced delay.

Traffic delay can be viewed as wasted time during the trip from one point to another. Delay can take place at stop signs, speed bumps, natural hindrances, and intersections. In this document, we only consider delay induced by traffic lights at intersections. An intersection can cause traffic delay in a variety of ways. Figure 2 shows three different ways traffic delay at an intersection can be defined and measured [10], [11].

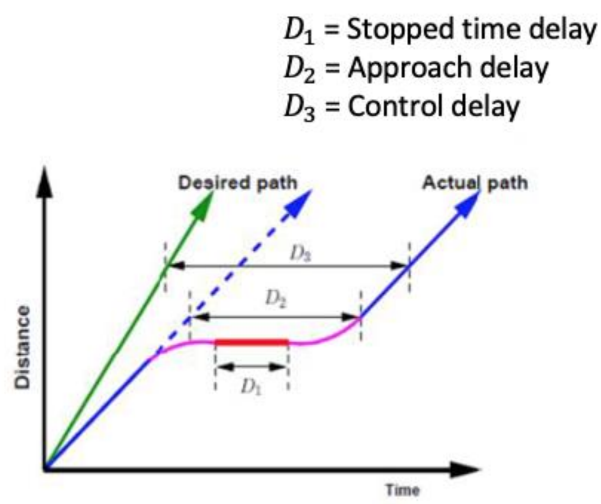


Figure 2. Traffic Delay at an Intersection.

As shown in Figure 2, $D1$ is the *stopped delay* which is defined as the duration of time a vehicle is fully stopped at the intersection due to red lights. $D2$ is the *approach delay* in which acceleration and deceleration times have been added to the stopped delay. $D3$ is called the *control delay* (a.k.a. the travel time delay), which is the total time a driver would have saved if the intersection would have not been on his/her way. Control delay indicates that, at an intersection, delay begins when the vehicle starts reducing its speed and ends when the vehicle is back to its original speed. In this document, we adopt the control delay as the delay induced by an intersection unless otherwise is specified.

Traffic congestion has long been an issue that people have struggled with. The number of vehicles on the road are growing and our ways to combat traffic have not significantly changed. For many years, engineers have tried to better understand the delay, and find factors impacting delays (for instance see

[10], [12], [13]). Shatnawi *et. al.* in [14] used a method that relies heavily on detection devices and communication technologies to keep track of every vehicle entering and leaving an intersection. The group called their method the AVDET system. This system needs to keep track of the vehicles in three key sensor areas. One very far upstream (the point vehicles are approaching the intersection), one at each lane at the entrance of the intersection, and one downstream (where vehicles fully departed the intersection). It should be noted that, in the AVDET system, sensors were used for each lane, not just for each approach, and they are all synchronized.

Once the delay is measured, there are different ways on how to reduce it. This is usually done through the action of changing the timing of the traffic lights. Traffic engineers can change the cycle length and green time length to reduce delay [15]. For safety reasons, the red clearance and yellow lights are not changed in most occurrences. The best way to test a new lighting configuration is by simulation [11], [16], [17]. Researchers throughout the years have developed new ways to determine lighting configurations. Recently many are using robust algorithms to run through multiple trials and measure delay. Arshad Jamal, et al use a Genetic Algorithm approach to do just this [11].

2.1 Collecting Data

It is the objective of this document to accurately model and measure the control delay of an intersection. If an intersection can be modeled accurately, it would be much easier to simulate different traffic scenarios and, thus, identify factors affecting delay and estimate traffic flow parameters. Equivalently, an accurate simulation model would make it substantially easier to design, evaluate, and implement various measures to effectively reduce delay.

Many researchers believe that a traffic simulator is one of the best ways to model traffic flow at an intersection. One of the best simulation software packages to use, when modeling traffic at an intersection is Vissim, which is discussed in 2.2 **Error! Reference source not found.**. It should be noted that it is often difficult to find quantitative data ensuring that developed traffic simulator is reasonably accurate.

In this research, we have collected data from the Santa Fe Avenue and Pacific Coast Highway (PCH) intersection, an intersection located in Long Beach, California. This intersection was chosen because of its close proximity to the Port of Long Beach (POLB) and because the City of Long Beach has designated this intersection as a part of an official truck route by ordinance (see Appendix A “City of Long Beach Truck Routes”). POLB is the second busiest container port in the U.S., and in recent years, the port handles more than 8.1 million twenty-foot equivalent unit (TEU) containers each year [18].

The high density of heavy trucks carrying goods, and its overall high traffic flow rate make the Santa Fe Avenue and PCH intersection an intersection of interest for this research. The reduction of traffic delay at the intersection results in reduction of travel times of heavy trucks to/from the port, and therefore, reduction in fuel consumption, air pollution, and environmental noise. Figure 3 shows the area around the POLB and the intersection where we collected data.

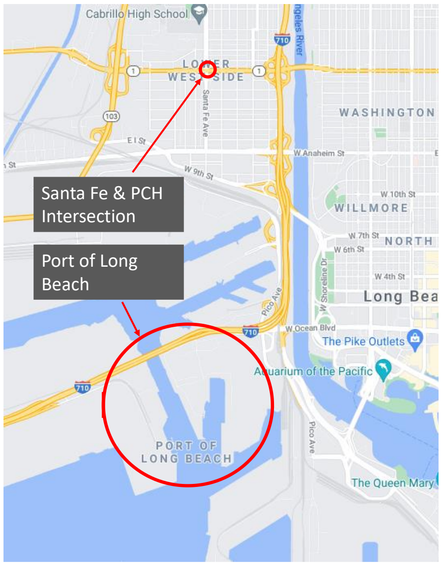


Figure 3. The area of data collection.

Figure 4 shows a view from top of the intersection.

Figure 4. A top view image of PCH and Santa Fe Ave.

We collected traffic data from the Santa Fe Avenue and PCH intersection during peak travel times on two discrete days. Data were then extracted from camera footage and other recording devices using image processing. It is important to note that no personal information was recorded during these times. The goal was to capture the flow of traffic at rush hours.

Table 1 shows the signal and traffic characteristics for the Santa Fe Avenue and PCH intersection for two different days and times (Day 1 and Day 2). On *Day* 1*,* we captured the traffic data at 5:30 pm on Wednesday 03/24/2021. On *Day* 2, we captured traffic data at 9 am on Thursday 04/22/2021. Two separate cameras were used, at different points of the intersection. One camera was used to capture traffic signal timings (by aiming at the traffic light signals), the other camera was pointed at the approaching vehicles of the intersection to capture queue lengths, arrival rates, and stop delays. Once the videos were recorded, they were carefully analyzed, and the following signal and traffic characteristics were extracted.

- *Cycle Length*: The amount of time required to serve all phases of traffic lights for each direction of the intersection before returning to the starting point.
- *Green Display Time*: The time that the green light indicators are displayed.
- *Red Clearance Time*: The time that the red-light indicators are displayed for every indicator at the intersection.
- *Yellow Clearance Time*: The time that the yellow light indicators are displayed.
- *Arrival Flow Rate*: The number of cars per unit time arriving at the intersection when the green indication is already displayed, and the queue of cars has cleared.

Table 1: Characteristics of Santa Fe and PCH intersection.

	PCH Day 1	**PCH Day 2**
Cycle Length	121.6 sec	114.5 sec
Green Display Time	63.7 sec	57 sec
Red Clearance Time	2.1 sec	2.3 sec
Yellow Clearance Time	4.2 sec	4.2 sec
Arrival Flow Rate	915 veh/hr	660 veh/hr

It is important to note that we employed a total of four approaches in this intersection: West Bound (WB), East Bound (EB), North Bound (NB), and South Bound (SB). Table I presents the signal and traffic characteristics for the West Bound (WB) approach of the intersection. Other approaches show very similar traffic signal characteristics (*i.e.*, all above listed characteristics except for the arrival flow rate). Arrival flow rate can differ drastically from one approach to another. When needed, we report arrival flow rate for other intersection approaches, as well.

The above measured signals and traffic characteristics from recorded traffic flows are used later in this report to model and calibrate our Vissim simulation model.

As mentioned earlier, three main types of control schemes for traffic signals exist: pretimed, semi-actuated, and fully actuated. The intersection of Santa Fe Avenue and PCH in Long Beach is a semi-actuated intersection. In semi-actuated traffic signals, traffic sensors are placed only on the minor streets, whereas major streets have green light priority. An example of a semi-actuated intersection can be seen in Figure 5.

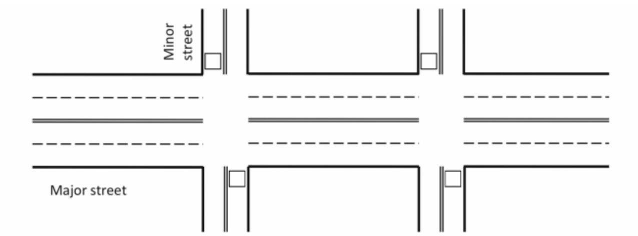


Figure 5. Semi-Actuated Intersection

The next important traffic signal characteristic that needs to be identified for the intersection of Santa Fe Avenue and PCH is the signal timing plan. A signal timing plan is the timing distribution of each phase of a traffic signal. It consists of a sequence of phases being actuated (becoming green). The intersection of Santa Fe Avenue and PCH specifically does have the normal 8-phase ring barrier diagram, which is typical for intersections throughout the United States. An example of an 8-phase ring barrier diagram is shown in Figure 6. The intersection of Satna Fe Avenue and PCH follows the same structure.

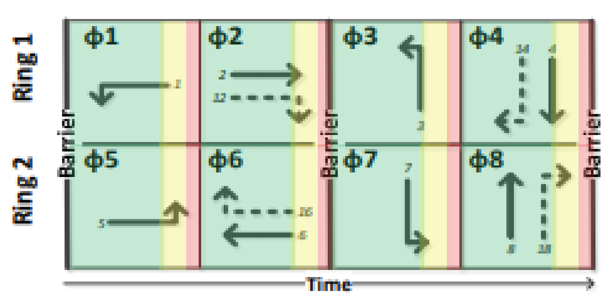


Figure 6. Example of a Ring-Barrier Diagram [6]

The ring barrier diagram can be looked at as the blueprint into how the vehicles move during the cycle. For example, Figure 6 indicates that phase 4 and phase 8 of the intersection can occur at the same time, indicating that arriving vehicles can both move through the intersection NB and SB simultaneously without any conflicts. These movements, along with the others shown in the diagram, need to be mimicked precisely within the Vissim simulation.

2.2 Simulation Model

Vissim is a microscopic traffic simulator developed by PTV Group in Karlsruhe, Germany. It is among the most common traffic simulators in the world [19]. Being a microscopic traffic simulator, the building blocks for a traffic simulation are the same throughout all simulators available on the market. These building blocks are labeled as infrastructure, vehicle inputs, and signal control. Infrastructure deals with static objects at an intersection such as the geometry of the road, signal heads, and detectors. Vehicle inputs are data collected on the field that is crucial to properly model the characteristics observed at the intersection. Information such as vehicle flow rates, relative vehicle flow rates, and percentage of vehicle class penetration fall under this category. Vehicle flow rates are the number of vehicles crossing the intersection per hour. Whereas the relative flow rates are the number of vehicles passing through a specific movement (left turn, right turn, thru) per hour. The vehicle classes we counted to input the vehicle class penetration are cars, heavy good vehicles (HGV), and buses. We neglected pedestrians and bikes, as they did not significantly impact our decision to focus on the intersection at Santa Fe Avenue and PCH. Lastly, the signal control is where we input the signal timing plan and method of control. As mentioned previously, this intersection is semi-actuated. There are giveaways that let us know which sides of the intersection have a detector and we can see this in Figure 7. The circle marking on the pavement typically depicts a detector location. Most detectors used by the industry are inductive loop sensors because of their low-cost and accuracy at detecting vehicles.

Figure 7. Detector location at Santa Fe and PCH.

Vissim offers two map providers which are Bing Maps Aerial and Open Street Maps. The former is useful to act as a canvas to “draw” our intersection; the latter is useful to find our intersection. We can see why this is true with Figure 8 because Open Street Maps has labels of the streets themselves and surrounding businesses, while Bing Maps Aerial shows a bird’s eye view of the intersection.

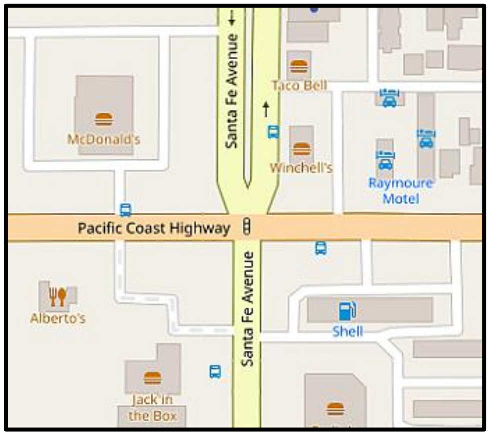


Figure 8(a). Bing Maps Aerial of the intersection of Santa Fe and PCH.

Figure 8(b). Open Street Maps of the intersection of Santa Fe and PCH.

Infrastructure is the first step in modelling our intersection since we need road network before adding vehicles and signals to it. We also then need to add something Vissim calls *Reduced Speed Areas* and *Conflict Areas.* Reduced Speed Areas are added where sharp turns are experienced such as right turns, left turns, and even merging onto left turn lanes. Conflict Areas require us to define which side of the intersection vehicles should yield when confronted with opposing traffic, to avoid collisions. We have done so for right turns for our intersection because it is allowable to do so. After drawing the roads, adding pavement markings, added detector locations, signal heads, *Reduced Speed Areas*, and *Conflict Areas*, the result of our intersection can be seen in Figure 9.

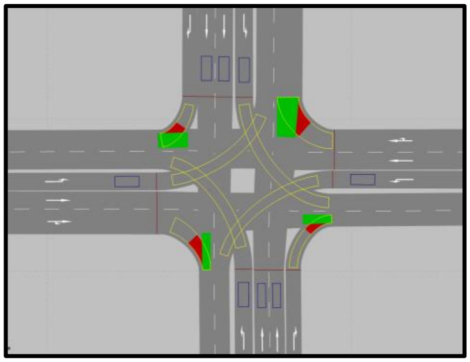


Figure 9. Intersection of PCH and Santa Fe Avenue drawn on Vissim

Next step is vehicle inputs. We used camera footage that lasted 10 minutes, which helped determine information (previously mentioned) that we will feed into Vissim. We multiplied each value by six since the vehicle flow rates are defined for an hour. Later, when we discuss validating the model, we will analyze only 10 minutes worth of simulation to make the comparison.

Table 2 presents the information regarding the 10 minutes of recording fed into Vissim for vehicle inputs.

Table 2: Vehicle Class Count from camera footage.

Direction	Vehicle Class Count		
SN	Car: 16	HGV: 5	Bus: 2
SE	Car: 12	HGV: 6	Bus: 0
SW	Car: 12	HGV: 1	Bus: 0
NS	Car: 21	HGV: 2	Bus: 0
NE	Car: 4	HGV: 5	Bus: 1
NW	Car: 17	HGV: 1	Bus: 1
EW	Car: 62	HGV: 14	Bus: 2
EN	Car: 5	HGV: 7	Bus: 0
ES	Car: 6	HGV: 1	Bus: 1
WE	Car: 91	HGV: 16	Bus: 3
WN	Car: 5	HGV: 4	Bus: 0
WS	Car: 8	HGV: 2	Bus: 1

In

Table **2**, letters S, N, E, and W indicate South, North, East and West directions, respectively. The movement SN indicates the movement of vehicles from north to south. Each row of

Table 2 shows how many different types of vehicles passed through the intersection within a 10-minute period.

As

Table **2** shows, most vehicles passing through the intersections are cars. These include sedans, coupes, sport cars, etc. It should be noted that the traffic flow is a stochastic process in nature resulting in different values each time we run the simulation.

The final building block is the traffic signal control. Vissim provides many controllers, but we chose and used a powerful plugin called Econolite ASC/3. Econolite manufactures external traffic controllers and is well known for its work. This plugin is basically a virtual version of its traffic signal controllers. This is where we input the signal timing plan and the ring-phase barrier diagram.

The Santa Fe Avenue and PCH intersection has 8 phases and is semi-actuated. We were able to input maximum and minimum green times we measured at the field and input fixed times for the phases that did not have detectors. Figure 10 shows the ring-barrier diagram we input into Econolite ASC/3.

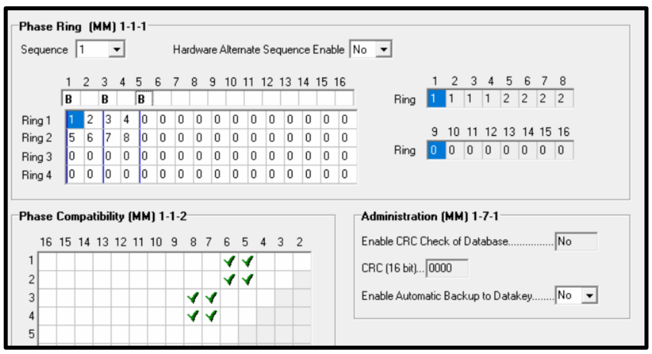


Figure 10. Ring-Barrier diagram in Econolite ASC/3.

We then proceeded with verifying the model to see if we are indeed capturing the dynamics of the intersection by making some comparisons to field measurements.

2.3 Simulation Results

As stated before, we collected data on two different days and times called *Day* 1 and *Day* 2 from the Santa Fe Avenue and PCH intersection. The delay data considered in this section are taken from the WB approach of the Santa Fe Avenue and PCH intersection. It should be noted that, once an approach is modeled accurately, the same methodology can be extended to other approaches, as well. By comparing measured (real-world) delay to simulated delay we can then evaluate the accuracy of the developed simulated model for the intersection.

The actual (real-world) measurements were taken from videos that captured the characteristics of arriving vehicles. These videos were 10 minutes long each.

Table 3 below shows the maximum control delay per cycle detected for day 1 for each of 10-minute video of the WB approach.

Once the delay measurements were collected, the Vissim simulation was run to find the simulated control delay. To improve the accuracy of the simulation model, we further calibrated the Vissim simulation parameters. The parameters that were changed were arrival flow rate, speed, and the placement of the sensors inside the simulation. This last parameter is one of the most important. Inside Vissim, when measuring delay, we must choose where to place “sensors” in the model so that the simulation knows when to start timing the delay. With control delay we found that the further apart the sensors the better. This means once a virtual car enters the simulation on the WB approach it passes over a sensor. Vissim begins to calculate the delay and stops when it passes over a second sensor on the other side of the intersection. Vissim can run a calculation in the background to determine how long a virtual vehicle would take to arrive at its destination if the intersection did not exist. The control delay is then found by calculating the difference between the two.

$$Control\ delay = Actual\ travel\ time - Theoretical\ travel\ time$$

We used a very similar approach when collecting the field delay. The first value was how long it took the vehicle to pass the intersection with no queue of cars at a green light. This means how long it takes a vehicle to pass through the intersection with no delay (also called *free flow condition*). Then we found how long it takes the vehicle to pass through the intersection when arriving at a red light. The difference between these two measurements is then found as the control delay.

Table 3 and Table 4 show the maximum control delays measured at the Santa Fe Avenue and PCH intersection compared to the maximum control delays measured from the Vissim simulation.

Table 3: Measured Control Delay vs. Vissim Simulated Control Delay for Day 1.

Trials	Day 1: measured control delay (sec)	Day 1: Simulated control delay (sec)
1	41	37
2	62	68
3	50	48
4	38	31

Table 4: Measured Control Delay vs. Vissim Simulated Control Delay for Day 2.

Trials	Day 2: measured control delay (sec)	Day 2: Simulated control delay (sec)
1	19	13
2	60	62
3	40	35
4	36	40
5	58	51

Figure 11 and Figure 12 show the measured (actual) and simulated control delays for Day 1 and Day 2, respectively. Each peak stands for one phase during the simulation. The flat parts of the graph correspond to when the light is green, and the vehicles can pass through the intersection without any delay. The peak of each triangle is the maximum delay experienced delay by vehicles during a phase. The blue line corresponds to the field measured delay whereas the orange line shows the control delay of the simulation after calibration. We noticed that the results obtained after further calibration (when parameters such as arrival flow rate, speed and sensor placement are further tuned) are much closer to the measured delay.

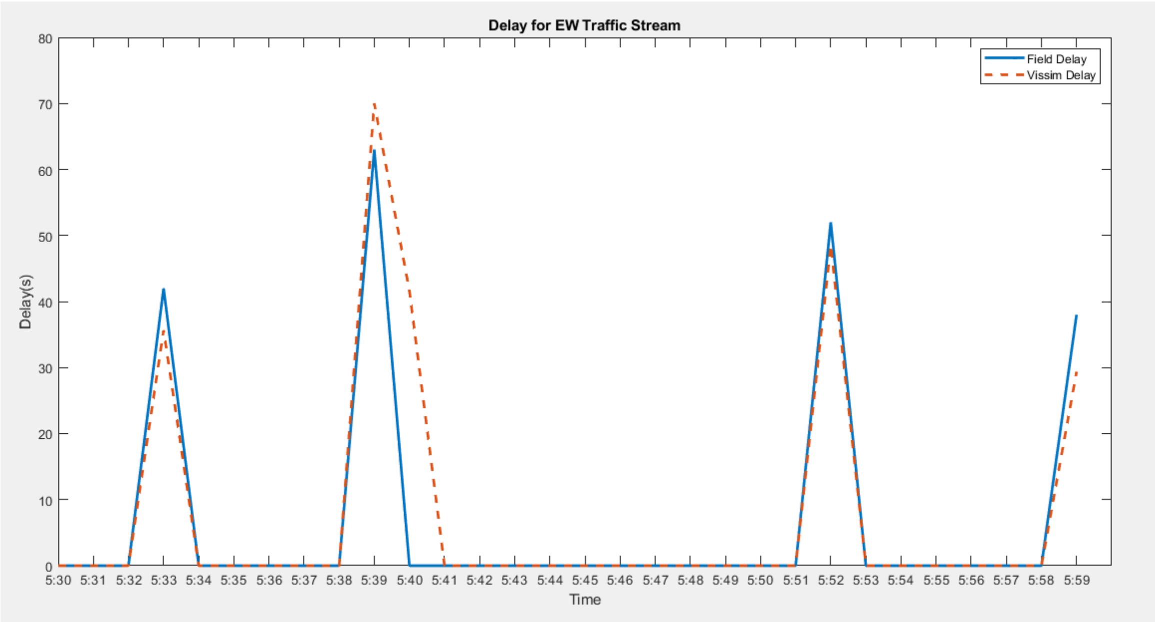


Figure 11. Control delays for Day 1: actual traffic measurements (blue), Vissim simulated traffic (orange).

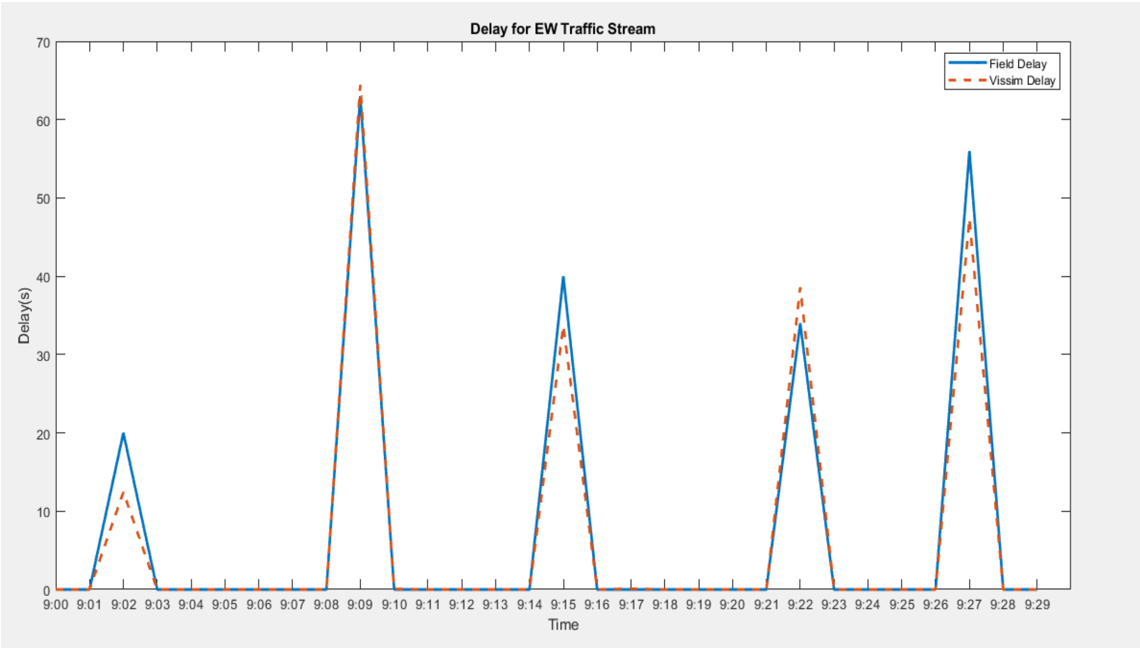


Figure 12. Control delays for Day 2: actual traffic measurements (blue), Vissim simulated traffic (orange).

It is important to note that we changed each parameter only slightly to keep the simulation as realistic as possible. We wanted to avoid drastically changing a parameter to make the measured and simulated delay similar. For instance, speed limit is an input inside the Vissim model, and Visim software will move vehicles at this speed unless otherwise specified. The speed limit at the Santa Fe Avenue and PCH intersection is 35 mph. However, it is very unlikely to see vehicles moving at the speed limit during 5:30 pm on a weekday. For this reason, the speed was set to 30 mph (based on field observation). Likewise, after processing the videos captured, it seems that vehicles tend to speed when allowed to do so. This means when there was less traffic, such as in the morning, vehicles moved faster than 25 mph. We then changed the speed limit slightly for this time. Lastly, as mentioned earlier, the sensors for the delay measurment were set to wide distance to ensure we captured the entire control delay correctly. This is important because as soon as the vehicle begins to decelerate, we needed to account for the delay.

Figure 11 and Figure 12 also show that the calibrated Vissim model has improved substantally. The simulated control delay, after calibration, demonstrate 10 – 20% less error compared to the base simulation model. That is, if we define the percentage delay error as

$$\%\ delay\ err = \frac{|actual\ control\ delay - simulated\ control\ delay|}{actual\ control\ delay} \times 100,$$

then, the percentage error of simulated model was reduced by 10 to 20% by some simple calibrations of the model.

3. Neural Network Delay Predictor

Artificial neural networks (ANNs), or simply neural networks (NNs), are often used in modeling systems which are nonlinear by nature. For example, NNs are usually used for image recognition mainly because of its ability to estimate the nonlinear behavior of the process. The nonlinear nature of traffic at an intersection makes NNs a promising candidate to model this system and predict the future behavior of traffic [5], [20].

A typical model of NNs can be seen in Figure 13.

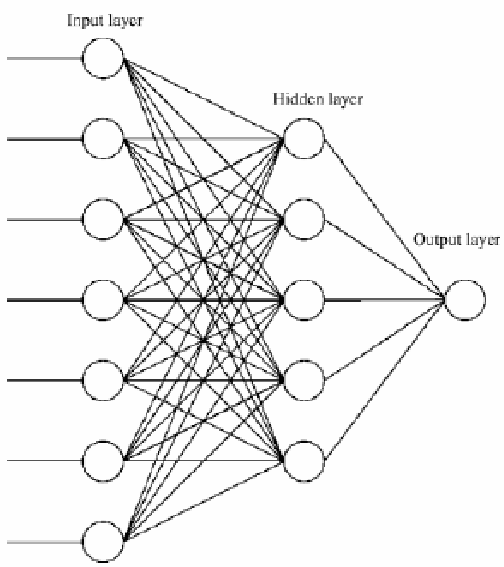


Figure 13. A schematic of a neural network.

The neural network has three parts: the input layer, hidden layer, and lastly the output layer. The input layer, as the name describes, is where the input features are first fed into the neural network. The hidden layer is where all the connections between the features are made. Finally, the output layer is where the output of the neural network is given. There is typically an activation function also included in the calculations. Activation functions are used to create a decision boundary to help better approximate nonlinear systems. We utilized an activation function known as sigmoid function and described by

$$s(x) = \frac{1}{1 + e^{-x}} \cdot$$

To train a neural network, we must collect data from the system we would like to model. In our case, we would like our neural network to estimate delay given concurrent conditions taken from an intersection. For us to accomplish the desired result of predicting delay, we will need to collect delay data and traffic data. Once the data are collected, they are used to train a neural network that will then estimate delay given parameters measured from the intersection. This form of training is the basis of supervised learning in which a feature dataset is connected to a target value set.

3.1 NN Training Data

We use VISSIM and its COM interfacing with MATLAB to generate data needed to train our neural network. The VISSIM model of the Santa Fe Avenue and PCH intersection can be seen in Figure 14. This intersection is near the Port of Long Beach, which results in increased heavy good vehicular traffic.

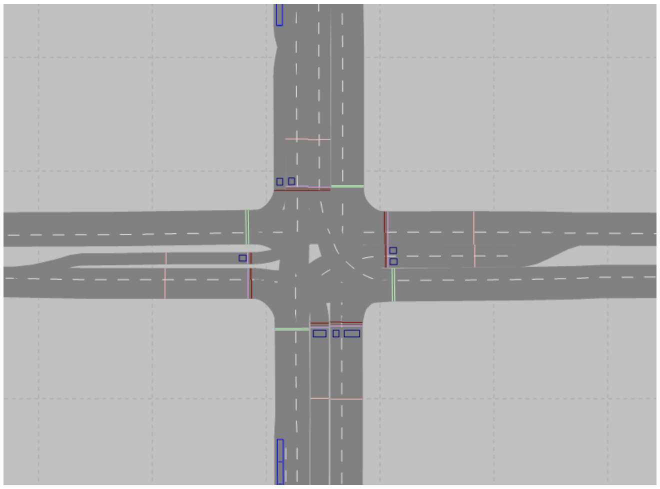


Figure 14. VISSIM model of Santa Fe Ave. and Pacific Coast Highway intersection.

The initial simulation data we collected from VISSIM are passenger vehicle and HGV counts, queue length, current signal state, and finally delay. These are the feature data sets we will use for our neural network training. The vehicle count feature was further expanded into two different classes of vehicles which are cars and heavy good vehicles (HGV). Cars are the typical passenger vehicles such sedans, coupes, SUVs, and pickup trucks. HGV vehicles are large tractor trailer vehicles and bobtails typically used to transport freight. We conducted the count at each leg of the intersection. Queue length is simply the total length of vehicles queued at the signal heads, and it is measured in meters. We also measured queue length at each leg of the intersection.

The output of NN is delay time, hence, we need to collect delay times from VISSIM. An example of the delay times collected can be seen in Table 5.

Table 5. DELAY TIMES FOR EACH LEG OF THE INTERSECTION.

	Interval 1 **Delay (s)**	**Interval 2** **Delay (s)**	**Interval 3** **Delay (s)**	**Interval 4** **Delay (s)**	**Interval 5** **Delay (s)**
WB	26.299	34.651	0.000	27.740	24.337
NB	0.000	14.610	32.990	25.480	36.658
EB	9.456	28.375	0.000	25.933	7.736
SB	0.000	34.114	50.896	50.505	26.462

VISSIM calculates delay time using the following formulas:

$$D_r = \frac{D}{T_{total}},$$

$$D = T_a - T_u$$

where D_r is the relative delay time, *D* is the total delay time, T_{total} is the total travel times of all vehicles within the measurement time interval, T_a is the actual travel time through the intersection, and T_u is the travel time if the vehicles are unimpeded (within the posted legal speed limit). The delay times are collected at each leg of the intersection along with the associated feature dataset. It is also worth noting that the data shown in Table 5 is for the non-turning sections of the intersection.

Another aspect needed to be discussed is the time interval of which the data needs to be collected. The time interval for the data shown in Table 5 is set at 100 seconds. We chose 100 seconds because this was the average cycle time for this intersection. When the time interval of measurement is set to a smaller value, the delay measurements contain many zeros, which may skew the result in the output of the neural network toward zero.

The data collected once the simulation period was completed amounted to a matrix of 4×2200 matrix for each feature. That is for the car count, truck count, queue length and delay measurements each resulted in a 4×2200 matrix. The simulation time was equivalent to 20 minutes of real time operation and 200 runs. Overall, the simulation period required approximately 8 hours to obtain all the data.

3.2 Neural Network Model Building

With the data we collected, we now can begin building a neural network model. As mentioned previously, we will use MATLAB's neural network toolbox. The MATLAB neural network toolbox simplifies much of the process of creating and training a neural network, allowing us to concentrate on the collection of data and data processing.

Recall the data shown in Table 5. We will utilize these datasets for our neural network. Specifically with the delay data, we organize the target data according to various approaches. The first approach is by adding the delay of each leg of the intersection for each time interval. Given that the average cycle time is 100 seconds, we can interpret this as the total delay per cycle. Observing Table 5, we would find the sum of each column with the new data now used as the target value. We can see the example new data values after calculating the sum in Table 6.

[RESUME HERE]

Table 6. DELAY TARGET VALUES W/ SUM PER INTERVAL.

	Interval 1 Delay (s)	Interval 2 Delay (s)	Interval 3 Delay (s)	Interval 4 Delay (s)	Interval 5 Delay (s)
W	26.299	34.651	0.000	27.740	24.337
N	0.000	14.610	32.990	25.480	36.658
E	9.456	28.375	0.000	25.933	7.736
S	0.000	34.114	50.896	50.505	26.462
Sum	**35.755**	**111.751**	**83.886**	**129.658**	**95.193**

The first neural network model can therefore be described in Figure 15. The inputs to the neural network are vectors of the passenger vehicle counts C, heavy good vehicle count T, and queue lengths Q of each non-turning link. The output of the neural network is the cumulative delay after one cycle given that the average signal cycle time is about 100 seconds.

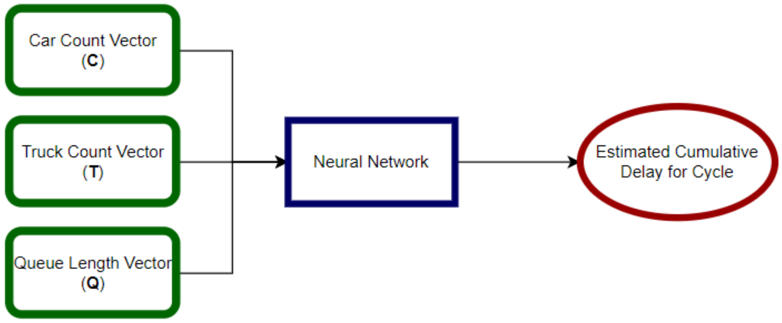


Figure 15. Neural network diagram of first preliminary model.

Using the neural network toolbox in MATLAB, we input the feature dataset and target data into the function and first run the settings with a single hidden layer with 10 neurons. We use the Levenberg-Marquart, Bayesian regularization, and finally the scaled conjugate gradient algorithms to taking the NN. The results of each training algorithm can be seen in Table 7 when training a neural network with a 10-neuron hidden layer.

Table 7. TRAINING ALGORITHMS W/ MSE.

Training Algorithm	Mean Square Error (MSE)
Levenberg-Marquardt	650.04
Bayesian Regularization	509.99
Scaled Conjugate Gradient	843.11

In **Table 7** the mean square error (MSE) is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where y_i is the actual or target delay value, $\hat{y}_i$ is a predicted delay value, n is the number of entry values and i is the index number. MSE quantifies the amount of error given by the estimator hence allowing to compare each individual neural network created. It is worth noting that the MSE only allows us to compare the estimator against other estimators and does not allow us to use MSE to assess the performance of the estimator.

Comparing the three training algorithms MSE, the algorithm in which yielded the least mean square error was the Bayesian regularization. Scaled conjugate resulted by far the worst performance. The Levenberg-Marquardt often performed nearer to the performance of the Bayesian regularization algorithm, though the Bayesian regularization algorithm more often outputted a neural network with a lower MSE. Regarding training time, Scaled Conjugate Gradient and Levenberg-Marquardt both took about the same time to train at only a few seconds (2-10 seconds). The Bayesian regularization training period often took upwards to 30 seconds to complete. Depending on the application of the neural network, it seems that of the three algorithms MATLAB recommends, the Bayesian regularization performed the best based on MSE. However, the Levenberg-Marquardt does drastically reduce training time but gives up a bit of performance of the predictions. For our current assessment, we will choose the Bayesian regularization to train the neural network while we compare hidden layer sizes. We can see some of the performance metrics of the neural networks in Table 8.

Table 8. COMPARISON OF HIDDEN LAYER SIZES.

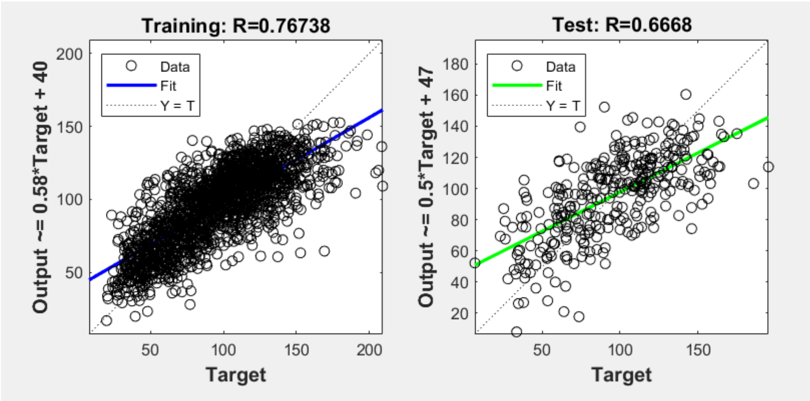
Hidden Layer Size	Mean Square Error (MSE)	R-value
7	544	0.7063
10	514.5	0.6996
12	500.75	0.7033
15	483	0.6960
20	447.5	0.6621
25	403	0.61278

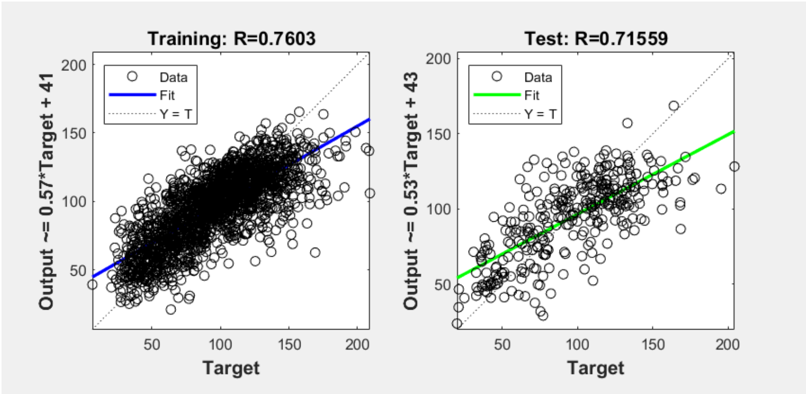
Anytime a regression model is created, one method to assess the performance of the model is using the correlation coefficient or R-value. The R-value is defined as

$$R = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{S_x}\right)\left(\frac{y-\bar{y}}{S_y}\right),$$

where x is the target value, $\bar{x}$ is the mean of the target delay values, S_x the standard deviation of the target delay values, y is the predicted delay values, $\bar{y}$ is the mean of the predicted delay values, S_y is the standard deviation of the predicted delay values, and lastly n is the number of entry values. The correlation coefficient will be a value ranging from -1 to 1. The R-value is a measure of how closely the predicted values are to the measured values. If a value of 1 is obtained, then the estimator is perfect at estimating the target values. If a value of 0 is obtained, this means that the estimator is extremely poor at estimating the target values. Note, it is possible albeit unlikely, to obtain an R-value of -1 which means that the estimator is estimating the exact of opposite or negative value with respect to the target values.

Each of the values obtained were averaged out of five different iterations of neural networks trained at each hidden layer size. Comparing the hidden layer sizes, it seems increasing it past passed 15 neurons does not improve the correlation coefficient of the test sets and appears to trend towards *overfitting*. Based on the above chart, it looks like using a hidden layer of 10 or 12 neurons give the best performance when considering the MSE and the R-value. A note on the R-value is that the higher the R-value is, the higher the linear correlation between the output of the neural network the target data. Figure 16 shows the regression plot for a neural network with 12 neurons in the hidden layer.





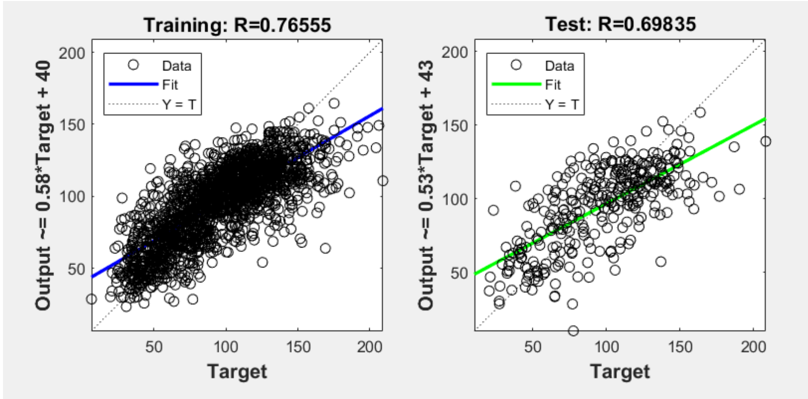


Figure 16. Regression plots given by MATLAB assessing neural network performance. versus target values.

In Figure 16, if the predicted delay value calculated from the neural network is equal to the target delay value, the point will be plotted along the dotted line (the dotted line shown to be at 45 degrees on the plots). The closer the line of best fit is to the dotted line on the test set, the better performing the delay estimation of the neural network.

These findings are still preliminary, and it appears that the neural network is showing promising results with the R-value being at approximately 0.7, which shows a strong correlation between the target value and the output of the neural network.

4. Consideration of Data Privacy

For this study, researchers led 32 study participants on a 1.5 mile community “datawalk” in West Long Beach—adjacent to the second busiest container port in the United States—to gain insights into the level of trust residents feel when encountering smart technologies that track their movements and capture their image. During the walk, study participants used a custom mobile app to map and photograph each smart technology observed. Participants then responded to several prompts on the app intended to gauge participants’ comfort and trust levels with the various smart technologies they encountered—including surveillance cameras, public WiFi routers and license plate readers. Relying on a “culture of trust” framework, the researcher found that study participants were far less concerned about specific technologies than about the data these technologies collect. Study participants expressed a lack of trust surrounding how government officials may use, store and share personal information collected through smart devices and software. Study participants also said local officials must take stronger measures to ensure accountability when personal data are exploited. Based on these findings, the study offers policy recommendations for local governments deploying smart technologies.

4.1 Introduction

Increasingly, “smart” cities are deploying devices and implementing software platforms that collect massive quantities of data. Local officials then analyze these data—sometimes using artificial intelligence and even blockchain technology [20]—to improve efficiency and streamline functions ranging from waste collection and public safety to traffic management. The wide-ranging nature of smart city initiatives suggest that projects will, ultimately, impact all aspects of contemporary life, including communications; utilities such as water and power; transportation; and government services. These trends are exemplified by the fact that, in 2018, investment in smart city technologies was estimated to reach $22 billion [21]. While technology and data fuel smart cities, they are unsustainable without stakeholder trust in government officials, services and third-party vendors [22]. The need for trust extends beyond cybersecurity and privacy and into the realm of ensuring positive outcomes for constituents:

> When a city creates services that provide the outcomes residents and others expect and rely on, then the sense of trust is earned and reinforced. Residents expect that the bus service gets them to work and back home safely and on time every day. They expect police and emergency services will arrive quickly regardless of whether they live in a rich or poor neighborhood. They expect that red light cameras are used to catch traffic violators, and not to track their movements around the city [23].

This study builds on the notion that smart cities must collect, store, share and analyze personal data with transparency and integrity. Specifically, it relies on a “cultures of trust” framework to analyze and discuss the findings from three datawalks through West Long Beach. In October 2019, Long Beach Mayor Robert Garcia directed city staff to develop a smart city strategic initiative and guiding priorities. Garcia’s (2019) directive in [24] highlights the potential for technology to spur economic development, to make Long Beach safer and to develop a resilient workforce. Most residents realize that social media platforms and corporate websites track their online behavior, analyze that data, store those data and potentially share them. However, smart cities like Long Beach—the second largest city in Los Angeles County, with a population of nearly 500,000 people [25]—also deploy devices and software that collect personally identifiable information about residents.

For this study, we developed traffic sensors capable of capturing video images of vehicle drivers, passengers and license plates, as well as images of pedestrians. We then recognized the need to incorporate privacy protective features into this traffic management technology. In order to inform the design of such features, we sought to gain insights into residents’ awareness of various smart technologies that collect personally identifiable information, and to gauge residents’ comfort levels with these devices and platforms. The research also probes participants’ perceptions of whether local officials are transparent regarding how the City of

Long Beach uses personally identifiable data collected by these devices, and whether accountability measures exist. Therefore, we designed a 1.5-mile “datawalk” that both began and ended at the same congested West Long Beach intersection our research team selected for piloting traffic sensors meant to improve traffic flows (Santa Fe Avenue and Pacific Coast Highway). The route intentionally takes study participants past commercial businesses

including multiple fast-food restaurants and a liquor store; through a residential neighborhood; into a city park; past a large public high school; and past a Long Beach police station.

Both the theoretical and methodological frameworks guiding this research build on previous studies that employ “datawalks” as a means of knowledge sharing and civic engagement within the smart city phenomenon. Datawalks exemplify an interdisciplinary research approach that integrate communication studies, technology studies and urban planning to examine the increasing datafication of our lives [26]. As study participants cross paths with smart technologies—video cameras, license plate readers, public WiFi routers, cell phone network repeaters and more—they gain a heightened awareness of just how much personally identifiable data smart city residents generate while moving through their daily routines. Subsequent to completing a datawalk, participants may be motivated to resist future data collection efforts. This resistance can be as simple as no longer wearing a Fitbit activity tracker around one’s wrists or avoiding a traffic camera by taking an alternate route to work. In other cases, the datawalk experience may spur participants to influence policy by testifying against privacy-invasive technologies during a City Council meeting or writing a letter to the editor on the topic.

Given the ubiquity of smart technologies in our built environment and the sway they (often unknowingly) exert over our behaviors, datawalks shed light on “data as material and situated, and constitutive of everyday life” [26]. In fact, walking is personified in our daily lives and “a principal mode of perceiving and living (embodying) urban places” [27]. When walking, our state of mind is “both present and detached from the world around,” and a way of assuaging alienation in the modern city [28].

4.2 Literature on Datawalks and Trust Cultures

Datawalking as a methodology

The datawalking methodology was initially employed as a pedagogical tool meant to immerse students—physically, spatially and sensorially—in the realities of ubiquitous data collection. The exercise was devised to challenge the laudatory rhetoric surrounding big data and was coupled with critical readings on the topic [29] [30]. The datawalking methodology also borrows from the "flashmob ethnography" framework conceived by design researcher Laura Forlano [31]. Forlano's ideas focused on identifying and highlighting the role of values in our built environment—including plazas, shops, restaurants and other urban spaces. Specifically, she planned workshops that encouraged participants, collaborating in groups of 3 to 4 people, to observe and document "tensions, surprises and counter-intuitive findings" during an hour-long urban trek [31]. Each group member assumed a designated role: the navigator/sketcher was tasked with guiding the group, as well as mapping both the pathways taken and spaces observed; the note-taker/interviewer was responsible for taking detailed notes of observations and, as appropriate, interacting with passers-by; and, finally, the photographer/videographer was tasked with shooting still photos or video throughout the observation process.

For her own projects, data researcher Allison Powell merged Forlano's study design with concepts inherent in the "network walkshops" that urban theorist Adam Greenfield created to raise awareness of digital networks in cities [32]. Powell adapted these techniques and applied them to diverse collaborations with British artists, urban planners, activists and workers. Powell (2017) characterizes her "data walkshop" methodology as "a radically bottom-up process of exploring and defining data, 'big data' and data politics" (p. 2) from the perspectives of residents who consider and document linkages among data, processes of datafication, and the spaces in which residents inhabit. Powell views data walkshops as "especially powerful in connecting questions, concerns or investigations related to data" (p. 3) with other powerful social challenges [32]. Specifically, datawalk participants are exposed to issues ranging from racial inequality and environmental degradation to food deserts.

This outcome mirrored that of our participants during their 40-minute walk through West Long Beach. Specifically, many said they acquired a heightened awareness of various urban challenges. During group "debrief" sessions following each walk, study participants asserted that smart technologies are likely to more negatively impact residents of color. They also noted that smart technologies have the potential to cause unique harms for undocumented residents—a reality with far-reaching implications in Long Beach, where more than 25 percent of residents are foreign-born [25].

During her walks, Powell assumed the role of observer and consciously refrained from influencing participants' discoveries and reactions. Following the walk, though, Powell asked study participants how their knowledge of data had changed because of the walkshop. None responded with plans to alter their behavior in an effort to generate less data. Rather, they considered how “thinking and paying attention” transformed their perceptions of the city and provided “heightened awareness of potential places to intervene” ([32] p. 17). Comments such as these lead Powell to conclude that walkshops serve a purpose greater than deepening participants’ understanding of datafication and dataveillance.

Dutch scholars van Es and de Lange (2020) describe datawalking as “an embodied, situated and generative practice” (p. 278) that situates data as relevant and integral to our daily lives [26]. Further, they suggest that datawalking, as a methodology, enables communication and technology researchers to confront key challenges that exist because of a “computational turn” brought about by the widespread deployment of smart technologies. These challenges lead to the “invisibility, decontextualization and (lack of) accessibility of data and its infrastructures” (p. 281). By facilitating datawalks, media and communication scholars can lead projects both with and about data—even if they lack programming or coding skills. In line with Powell’s (2018) own conclusion that walkshops produce knowledge from the “bottom-up,” van Es and de Lange (2020) characterize the research method as “generative” (p. 282). This is because it spurs changes in the study participant, in the technologies under study and, finally, in the relationship between the two.

Grounded in the belief that city residents lack knowledge about the digital and datafied technologies deployed in smart cities—ranging from office keycards to parking tags and e-wallet apps—van Zoonen et al (2017) designed a datawalk meant to explore, develop and test forms of civic participation and engagement [33]. van Zoonen et al. recruited a diverse group of 80 people, comprised of urban residents and city staff, knowing that they had competing interests and possessed disparate knowledge bases. While walking through various sections of Rotterdam and The Hague in the Netherlands, the researchers asked study participants to contemplate four questions: Where do you see data? What happens with it? Who owns it? Would you like to have some say about it? (p. 17).

Ultimately, van Zoonen and her colleagues facilitated 14 datawalks, comprised of a maximum of five people each. The researchers avoided playing the role of “experts.” Therefore, when study participants encountered technologies, the researchers declined to provide details about how these technologies collected and used data. Based on discussions among datawalk participants, van Zoonen *et al.* conclude that neither residents nor city workers seriously reflect on the massive role data collection plays in their daily lives. Their findings also led them to believe we must “reconstruct digitisation and datafication as a social issue rather than as an individual responsibility” ([33], p. 19). This suggests that technology policymakers, as opposed to technology users, should create and implement data privacy protections.

Cultures of Trust Scholarship

Policymakers are unable to identify and prioritize risks without first understanding what needs to be protected and the urgency of doing so [21]. In fact, the implementation of any new smart city technology is associated with risk. Risk is typically expressed using some variation of the following equation: *Risk = consequence x likelihood* [34]. To elaborate, risk is determined by a person's understanding of the severity of a set of consequences, combined with the likelihood of experiencing those consequences. Trust and risk are frequently articulated as opposite sides of the same coin: while risk considers potential adverse outcomes, trust focuses on the desirable traits that need to be preserved [21].

Institutional trust is critical in order to maintain the stability of societies and cultures [35], [36], [37]. As local governments increasingly deploy smart technologies that collect personal data, residents recognize that unintended consequences, including privacy violations, are possible. This awareness of potential risk and abuse necessitates the need for trust [35]. Sociologists coined the phrase *trust cultures* to distinguish between societies based on their level of interpersonal trust and shared ethical values [38], [39]. More specifically, "trust cultures are groups of people with shared opinions, values and attitudes regarding whom and what to trust in a shared social and locational context" [35]. While trust cultures inevitably vary from country to country, due to historic events and societal norms, levels of trust differ among sub-populations within a single city. For instance, Black Americans are significantly less likely to trust local police to protect them and their families compared to white Americans [40].

Additionally, levels of trust within cultures typically differ according to specific circumstances—such as the type of data collected, how data are shared and who has access to the information. For instance, Nissenbaum (2010) proposes a contextual integrity framework [41]. This approach provides a holistic approach for understanding privacy expectations and their implications by drawing on scholarship situated in the law, public policy and political philosophy. Nissenbaum asserts that privacy "is preserved when informational norms are respected and violated when informational norms are breached." Information flows within a delineated context that fail to adhere to existing norms are perceived as privacy violations. Understandably, a weather app needs to recognize a user's location to provide an accurate forecast; the app's functionality depends on access to this detail. By contrast, a banking app does not need to track a user's location to provide an account balance; therefore, collecting geolocation information constitutes a privacy violation.

Previous research suggests that smart city residents will accept sharing their personal data only if officials establish trusted relationships among participants, and if participants can control the use of their data [42]. For instance, Julsrud & Krogstad in [35] explored the acceptance of government use of mobile phone data in various contexts and how these practices relate to different types of trust. Based on findings from a survey distributed to residents of the Oslo, Norway and Tallinn, Estonia—both smart cities—Julsrud & Krogstad caution smart cities against exploiting mobile phone data without first establishing safeguards that reassure residents their personal information is safe. Local government plans to use mobile phone data are likely to fail, and compound existing distrust, unless officials incorporate the “needs and wants” of residents when undertaking such practices [42].

We are all familiar with the adage that “perception is reality,” and this holds true in the context of smart city technologies. Often, user trust in devices and platforms is determined by belief and appearance, as opposed to actual performance records. Therefore, public engagement and education about a project’s intended purpose, functions and benefits are vital to instilling trust [21]. When local policymakers neglect to articulate the positive aspects of a project, the public is left to speculate and draw their own conclusions—which may clash with reality and foster distrust. Similarly, making documents and records easily accessible enables the public to verify how a system works, as well as which data it collects. This practice can build trust with stakeholders [21].

Chan (2019) also proposes a framework for building trust in the smart city. According to this approach, “Trust is created when services from the various value creators (cities, utilities, communities, corporations and individuals) in the smart city ecosystem create outcomes that are relevant, rendered reliably and with integrity, by service providers who are credible, transparent and have the capacity to execute.” Further, elements of trust must be taken into consideration and embedded into each phase of service implemented by the smart city service [23].

The following section elaborates on the term “smart city” and delves into the privacy guidelines that Long Beach has adopted as part of its own Smart City Initiative. Both bodies of literature influence my research questions.

4.3 Defining Smart Cities and Data Privacy

Conceptualizing smart cities for this project

No singular definition of a smart city exists. However, one conceptualization of the term that focuses on citizen involvement in the collection of data—and even in data governance—is relevant to datawalking. This holistic approach characterizes smart cities as ecosystems comprised of people, processes and solutions working collectively to prosper [22]. Residents collaborate with government entities to identify challenges and craft relevant fixes, as opposed to passively accepting policies fashioned by elected officials or corporations. Datawalks represent an opportunity for residents themselves to generate data. Also, according to this articulation of smart cities, residents assume an active role and participate in decision making [43], [44]. It stresses the importance of policymakers and technical experts soliciting input from key community stakeholders. In addition, citizen engagement such as community representation on committees that establish data governance practices, ensures that municipal leadership consider "big data" policies from the public's perspective.

City of Long Beach Data Privacy Guidelines

Long Beach City Council members adopted Data Privacy Guidelines on March 9, 2021, in an effort to establish parameters around how local government may—and may not—use residential data collected through city services, projects and programs. The city intends to use the guidelines as a basis for technical guidance for all departments and "embed them into local government policies, contracts, procedures, trainings, educational campaigns (for both City staff and Long Beach residents), software applications, and legacy systems" [45]. An abridged description of Long Beach's Data Privacy Guidelines follows:

- Long Beach will be publicly transparent and accountable in its collection and management practices of personal data, notwithstanding data requirements mandated by law.
- Long Beach will work to provide participatory, responsive feedback channels for residents to inform the City's data collection and usage practices, exercise privacy complaints, and ensure the City is held accountable to these Guidelines.
- Long Beach will advance digital equity and prioritize the needs of marginalized communities on matters pertaining to data and information management.
- Long Beach will use data in an ethical and non-discriminatory manner to not reinforce existing racial biases and prejudiced decision-making.

- Long Beach will practice ethical data stewardship throughout the data lifecycle to minimize misuse of personal data.

The City of Long Beach now faces the daunting task of operationalizing these guidelines in a way that broadly protects residents from data privacy violations. After studying literature based on previous datawalks and on smart cities, these research questions emerged:

- **RQ1:** What attitudes and comfort levels do residents express regarding smart technologies that capture personally identifiable data?
- **RQ2:** Do residents' attitudes and comfort levels vary, depending on whether the technologies are deployed by the city or private entities?
- **RQ3**: Based on the datawalk findings, what implications exist for smart cities deploying technologies that collect, store, analyze and sometimes share personal data about residents?

This research, partially, represents an effort to articulate tangible actions the city could take to ensure governmental data collection avoids leaving residents vulnerable to digital exploitation and fraud. The following section describes how the researchers designed the datawalks employed for this study; how we recruited Long Beach residents to participate; and how we coded and analyzed data.

4.4 Methodology

The researchers' initial step involved identifying an intersection near the Port of Long Beach—the second busiest container port in the United States—that would serve as an ideal location for their traffic sensor pilot project. As the paper already discusses, the researchers ultimately chose the intersection of Pacific Coast Highway and Santa Fe Avenue based on a combination of factors. First, a truck route map illustrated exceptional traffic congestion at the intersection during peak hours [46]. Second, information found in the City of Long Beach 2040 General Plan (2019) identified current levels of service (LOS) as "poor operation," with "longstanding vehicular queues" and delays "up to several minutes." Santa Fe and Pacific Coast Highway at present have received a grade of "E" for present conditions, and are projected to receive an "F" in 2035. A grade of "F" is defined as having "forced flow," which "represents jammed conditions," unpredictable volumes carried, and likely potential for stop-and-go traffic flow [47] (See excerpt from Appendix B Long Beach's General Plan). Finally, during multiple visits to Santa Fe Avenue and PCH, the researchers observed a mix of heavy truck, passenger vehicle and pedestrian traffic in all four directions. Considering plans to deploy traffic sensors there, the intersection of Santa Fe Avenue and Pacific Coast Highway was also the appropriate starting point for the datawalks. Recruiting participants

To recruit study participants, the co-PIs created a series of GIFs and posted them to multiple social media sites, along with a blurb describing the project (See social media post in Appendix C). Specifically, the co-PIs promoted the datawalks on about a dozen relevant Facebook, LinkedIn, Twitter and Instagram accounts between June 23, 2021 and June 30, 2021. In addition, a City of Long Beach staff member promoted the datawalks to stakeholders involved in ongoing "smart city" initiatives. All digital promotional materials noted that volunteers must be at least 18 years old, work or live in Long Beach, and have access to a smart phone or tablet. Recruitment materials also noted that each study participant would be eligible to receive a $25 Target gift card upon completion of the 90-minute study. Finally, the text included a hyperlink to an EventBrite registration page. Clicking on the hyperlink brought volunteers to a form enabling them to sign up for one of three datawalks scheduled for July 1, 2021.

Ultimately, 32 people registered through EventBrite. A co-PI emailed these volunteers on June 30, 2021, requesting that they complete an anonymous Qualtrics survey requesting basic demographic information (i.e. age, race, gender, zip code) prior to their chosen walk. The co-PI's email also included a URL for the cloud-based smart phone app participants would need to access during the datawalk, and encouraged them to test the link ahead of time.

Cloud-based mobile phone app

The study design for this project incorporates elements of initiatives led by multiple scholars, such as [32], [31], [26], and [33]. These scholars assigned roles to participants—i.e., map maker, photographer, notetaker. We adopted a different approach by collaborating with two geographic information systems specialists who work at the Center for International Trade and Transportation based at California State University Long Beach. For this project, they customized a mobile application initially developed by mapping software company Esri. Because the app is cloud-based, it does not require users to install anything on their mobile devices; rather, they access it by clicking on a URL (https://survey123.arcgis.com/share/f4b4553f9d494ceca90fa21670eb738a).

When opened while on the datawalk, the custom app prompted users to identify each type of smart technology they were observing. It provided radio buttons with the following options: “traffic camera,” “residential surveillance camera,” “public WiFi antenna,” “sensor,” “internet-connected bus kiosk” and “other.” Next, the app took advantage of mobile phone GPS functionality to automatically identify and “pin” the location of the technology being observed. Walkers confirmed the accuracy of the location data by pressing the “save” button. Next, the app requested that users photograph the technology currently being observed. The app then displayed a series of prompts (detailed below), accompanied by text boxes for entering responses.

Twenty-eight of the 32 walkers completed the demographic survey. Among those walkers, 20 were female and 8 were male. Nine were between 18 and 31 years old; 6 were between 32 and 50 years-old; and 13 participants were between 51 and 61 years-old. About 63 percent of participants were white; about 33 percent were LatinX; nearly 7 percent were Asian; and another 17 percent of study participants self-identified as “other.” More than two-thirds of participants reported living in Long Beach, while about 15 percent of walkers worked in Long Beach; and about 8 percent of participants attended school in the city.

Nine volunteers attended the first datawalk, which began at 8:30 a.m. on July 1, 2021. Twelve volunteers attended the second datawalk, which began at 10:30 am. Eleven volunteers participated in a walk that began at 12:30 p.m. Before starting each walk, a researcher gathered participants and disseminated informed consent forms for each participate to sign. She also explained the purpose of the study and general research goals.

We intentionally designed the 1.5 mile datawalk route so that study participants would pass by, and walk through, a variety of elements in West Long Beach’s built environment. The itinerary included commercial businesses, such as multiple fast-food restaurants and a liquor store; a residential neighborhood; a 12-acre city park; Cabrillo High School, which has an enrollment of more than 2,000 students; and the Long Beach Police Department’s West Patrol Division.

In order to obtain participants’ baseline understanding of data privacy-related concerns, the researcher asked participants four general questions prior to heading out for the walk itself:

1. Have you ever questioned your data privacy before? If yes, in what instances?

2. What types of data-collecting technologies do you know about that currently exist in Long Beach?

3. Do you believe there is too much technology installed around the city?

4. What do you expect to observe during the walk?

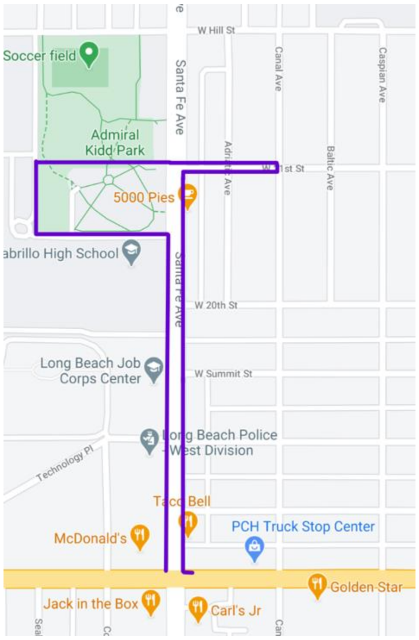


Figure 17: Map of datawalk route.

The researcher leading the datawalks took handwritten notes and audio recorded pre-walk conversations, which lasted approximately 15 minutes each. Then group members began the 1.5 mile route. At various points, the facilitator stopped the group to point out “smart technologies.” These included: commercial surveillance cameras, internet-connected bus kiosks/stops, residential surveillance cameras, license plate readers, public WiFi routers and city-owned surveillance cameras (mounted in Admiral Kidd Park and on traffic lights). In addition, study participants themselves identified internet-connected devices without the co-PI pointing them out.

During each encounter with a smart technology, study participants paused to use the cloud-based mobile phone application and respond to these questions:

1. Type of smart technology you are observing

2. What types of data do you believe this smart technology collects?

3. Does the technology surrounding you stand out more than it did prior to our discussion?

4. Are you comfortable with the idea of this particular smart technology collecting personal information about you?

5. Anything else you would like us to know about this smart technology?

Participants in all three groups spent approximately 45 minutes covering the datawalk route—which ended at the same intersection where it began, Santa Fe Avenue and Pacific Coast Highway. Once there, the researcher gathered study participants to "debrief" the experience and obtain responses for these questions:

1. Did anything you observed during the walk surprise you?

2. Is it the city's obligation to inform residents about data collection, or are residents responsible for their own awareness of the "smart" technologies that the city deploys?

3. Do the benefits of data collection, such as traffic management, outweigh potential privacy violations?

4. After observing the technology in the city, do you feel watched, unsafe, or violated?

Each post-walk discussion lasted approximately 20 minutes. Again, the researcher facilitating the datawalks took handwritten notes and audio recorded the conversations.

Datawalks generate a rich source of grassroots-level insights. They offer empirical evidence into both the deliberate and unintended consequences of privacy-invasive technologies, including how they manipulate culture and society [26]. The following section describes findings from the three datawalks conducted on July 1, 2021. It also explores how these findings can potentially inform data privacy guidelines articulated by the City of Long Beach—as well as efforts by other smart cities to safeguard residents' data privacy.

4.5 Datawalk Findings and Discussion

Video cameras

Study participants encountered dozens of video cameras during each datawalk. In response to the mobile app question, “What types of data do you believe this smart technology collects?” participants identified a range of data for all cameras, regardless of ownership, including: facial images, license plates, images with a time stamp, audio and geolocation. Specific to residential cameras, several study participants noted that cameras are recording package deliveries and “video of random Amazon delivery drivers.” Walkers also suggested that residential cameras collected images of “every bird and dog” that trigger a camera’s sensor and “water usage” from irrigation systems. Specific to a camera mounted near the doorway of a liquor store, participants suggested the recordings capture “suspicious activity” and “robbery suspects.” In response to city-owned cameras installed at intersections, study participants believed they captured the “number of cars and location during signal cycles” and “traffic violations.” Nearly all study participants said they realized surveillance cameras existed but that they rarely noticed them prior to the discussion a co-PI facilitated during the datawalk.

How study participants answered the question, “Are you comfortable with the personal info collected?” varied by context. While encountering a traffic camera owned by the Long Beach Department of Public Works, one participant typed that his or her acceptance level “depends” on *why* the camera is collecting personal data. “If I ran a red light, yes. Any other time, no.” Similarly, other walkers wrote, “It depends on the restrictions placed on the use of this tech” and “If the information is used for safety I am fine with it. But for selling information, no.” Another walker referenced the legal realities of public cameras by writing that Long Beach residents should have no expectation of privacy in a public space, while another acknowledged daily trade-offs with technology: “It’s a little uncomfortable but, I think, worth the compromise.”

Walkers were, for the most part, supportive of surveillance cameras in Admiral Kidd Park. Findings typed into the app include comments such as, “Never thought parks had cameras but happy to find out they do” and “child safety is important.” However, several participants again acknowledged trade-offs inherent in using smart technologies. Specifically, one participant wrote about feeling comfortable with cameras in the park, “but I have concerns about facial recognition of children.” Another commented, “I feel uncomfortable with the thought that kids are being watched, but I can understand that there’s a concern regarding predators or potential kidnapping.”

Study participants expressed few concerns about the residential surveillance cameras they passed during the datawalk, noting that “people have a right to secure and protect their property.” In fact, a number of participants conceded that they themselves installed surveillance cameras on their homes. Still, walkers noted unease specific to Ring cameras and doorbells after learning from the researcher that the Long Beach Police Department partners with Ring, a home surveillance company owned by Amazon. The agreement, in place since August 2019, enables Long Beach police to access the Neighbors Portal—where Ring camera owners post real-time safety alerts—and request surveillance footage from device owners near an active investigation. “I don’t mind people using [Ring cameras] but I have problems with what the company can be doing with the videos and where they are sending their videos,” typed one respondent. “I am fine with the people collecting information for safety but not with the actual company that sells the cameras,” wrote another. A third study participant noted, “It bothers me that police can access Ring.”

Public WiFi routers

The West Long Beach route required participants to walk by a McDonald’s and a Taco Bell, both fast food restaurants that host public WiFi networks. Participants noted that their mobile devices joined these networks even without intentionally logging on. Comments typed into the mobile app strongly suggest that many participants disliked being automatically connected—despite uncertainty surrounding which data these public networks collect and how network owners use data. One participant speculated that the restaurants may now possess “[my] cell phone unique identifier that could be collaborated with other data collection services to create a profile that could include my name, address, phone number, etc.” Others shared a belief that public WiFi networks collect everything from geolocation and browsing history to website passwords. “I don’t believe McDonalds is doing anything with my WiFi data, but maybe I’m naïve,” wrote one datawalk participant. By contrast, others said they consciously avoid public WiFi networks to protect their data privacy. “Provides public access but at what cost?” questioned a participant. Another typed, “If I choose to use the free WiFi then I ‘pay’ with my data. But I choose not to use it.” Still, multiple study participants said they accept privacy risks in exchange for the convenience of accessing the internet in a café or other public space. “I don’t want data abused,” stated a walker, but if a restaurant chooses to collect data “to offset costs, I’m OK with it.” Another respondent said he or she is comfortable with using public WiFi “as long as it’s not hacked for banking info, etc.”

When encountering public WiFi routers at Admiral Kidd Park, only two walkers entered concerns into the app. One walker noted that local government is not entitled to know if someone “posts to social media” from a “park bench.” The other study participant’s response indicates a concern about attachment to technology, rather than data privacy. “You’re outside. Be outside,” this person lamented in comments submitted through the app.

Internet connected bus stops and mobile app

When group members encountered a Long Beach Transit bus stop midway through the walk, the researcher drew their attention to two “smart” technologies widely used by transit riders.

First, the researcher pointed out “Text For Next” [48], a feature that allows riders to text their stop number (posted on bus stop signage) to Long Beach Transit and receive arrival times for the next bus. The researcher also mentioned that Long Beach Transit riders can download the Moovit mobile app, which provides real time updates and route schedules. According to the Google Play store, Moovit accesses users’ geolocation, contacts, camera, device accounts and media files [49].

In response to the prompt that asked participants which data they believe Long Beach Transit collects through its texting and mobile platforms, datawalkers cited IP addresses, geolocation, one’s current commute (starting point and destination) and the travel patterns of Long Beach transit users. Most study participants indicated they rarely ride the bus but appreciated that Long Beach Transit offers digital platforms. “Wow! I didn’t realize all of these helpful tech things exist at the bus stop,” one participant wrote. “As the spouse of an immigrant, I get a bit ‘prideful’ of cool things in this country…especially when it helps out the working poor,” another participant wrote. One walker praised the technology for its potential to “get more people on transit.” At the same time, study participants questioned why the Moovit mobile app accessed personal data on mobile devices. “Comfortable that it helps you get to point a to point b. Not comfortable with the fact that it could collect data from your personal phone,” a respondent typed. Others questioned whether Long Beach Transit aggregates or retains personally identifiable information; sells data; and how the agency informs users about its data practices.

While at the bus stop, multiple walkers took the initiative to inspect the physical structure. “Is that a camera? Are we being recorded while we wait for the bus?” one study participant typed into the app. Similarly, other walkers wondered, “Are there cameras on some of these [bus stops]?” and, “What is the weird box on the pole at the bus stop?” These inquiries demonstrate that, by encountering smart technologies within the context of an organized datawalk, study participants obtained both a heightened awareness and an enhanced curiosity about technologies in their built environment.

Cellphone repeaters

During each of the three walks, group members stopped to record their reactions to a cell phone repeater mounted on a utility pole along Santa Fe Avenue. In response to the prompt asking what type of data the technology collects, participants identified geolocation and mobile device identifiers for each device pinged. The majority of respondents said they rarely noticed cell phone repeaters in their daily travels. Comments included, “I’ve heard and read about this tech, and I find it unavoidable. It’s interesting to identify them around the city, though,” and “Knew they were out in the world, but not as prevalent.” For the most part, study participants indicated that they understand the need for cell phone repeaters to collect personal data in order to provide connectivity. Therefore, they accept that cell phone repeaters are ubiquitous in cities. But one participant wrote that if wireless carriers store or share phones’ unique identifiers with third parties, "that's an invasion of privacy." Another expressed feeling uncomfortable with the technology because "it just knows too much about my whereabouts at all times."

Automated license plate readers

The walk route took study participants past the Long Beach Police Department's West Patrol Division. Although this police station contains multiple smart technologies, including video cameras, the researcher pointed out that patrol cars are equipped with automated license plate readers. Also, in November 2020, Long Beach City Council approved a Parking Enforcement Division request for a $400,000 purchase of 17 automated license plate readers.

Some study participants were aware of the Long Beach Police Department's extensive use of automated license plate readers. Several said they read recent news stories exposing how the police department shared data from automatic license plate readers with Immigration and Customs Enforcement officials.

This cooperation occurred despite a 2018 ordinance barring city agencies from providing information to federal immigration officials [51]. In fact, the Long Beach Police Department shares license plate data with more than 1,000 agencies, according to public records [52]. In addition, in November 2020, Long Beach City Council approved a Parking Enforcement Division request for a $400,000 purchase of 17 automatic license plate readers [52].

While some study participants possessed a general awareness of license plate scanners, most comments indicated surprise about indiscriminate use of the technology. "I assumed police only captured license plates voluntarily in the case of an incident," typed one participant. Several walkers expressed concern upon learning that city vehicles, other than police cars, are equipped with automated license plate readers. "It's a little creepy…I understand why the police have it. But other cars, too…that's advanced," one comment states. Similarly, another participant wrote, "I didn't know that these existed and that they're getting implemented in more city vehicles besides just police cars." However, the majority of concerns typed into the datawalk app center less on license plate readers themselves and, rather, on how the Long Beach Police Department stores and shares data collected through the technology. Comments included:

- "Where do the data live?"
- "I'm comfortable with the technology. However, I have concerns about the retention and use of that data."
- "I'm nervous about with whom the data is shared."

- “Not ok to store my data! I’m a law-abiding citizen!”
- No way. I don’t feel comfortable with police collecting information about me for no reason without me knowing.
- I’m sure it’s used to solve some sort of crimes but it seems like a lot of information and invasion of our privacy.

Several study participants acknowledged the potential for law enforcement technologies to disproportionately impact people of color. For example, one participant typed, “I don't like that this is used by police at all. It reinforces existing racial biases and over surveils POC.” Another walker typed that, knowing the LBPD shared license plate data with federal immigration officials, “I am extremely concerned about LBPD overreach.” And a white study participant wrote, “Communities of color will have very different answers on these types of things.”

Limitations and future research

Because the COVID-19 pandemic created several challenges for the datawalks incorporated into this project, the findings are limited. First, CSULB’s Institutional Review Board paused approval for all human subject research during the pandemic. As a result, the researcher could not move forward with the project until just two months prior to the grant’s initial expiration date. Additionally, the researcher had planned to recruit datawalk participants during community events and meetings. However, nearly all public events were canceled March 2020 through Summer 2021. Therefore, the researcher relied primarily on social media platforms to promote the study. While the researcher ultimately led 32 people on datawalks, this number represents half her target sample size. Also of note, 63 percent of study participants were white and mostly lived in East Long Beach—clearly not representative of the city as a whole.

For the next phase of this project, the researcher is designing three additional datawalk routes in diverse Long Beach neighborhoods, including those where residents’ primary languages are Spanish and Khmer. Tentatively, the walk routes would start:

- At a community hub in North Long Beach, such as the Michelle Obama Neighborhood Branch of the Long Beach Library.
- At a community hub in the Washington neighborhood (bordered by PCH, Anaheim Street, Long Beach Boulevard and the Los Angeles River).
- At a community hub in Cambodiatown, such as the United Cambodian Community space on Anaheim Street.

The researcher plans to lead two walks through each of these neighborhoods, with a minimum of 30 residents participating. The findings and analysis from this subsequent phase of research will better inform efforts to implement the City of Long Beach's Smart City Initiative.

5. Conclusions

This report has presented two significant qualitative and quantitative developments related to smart communities technologies and policies that are relevant to a wide range of municipal planning scenarios. The Vissim simulation model was developed to calculate the control delay at the intersection where there are not enough sensors to measure all traffic data on site. We also described how the Vissim model can be finely calibrated to model the traffic flow at the intersection. Simulation results showed that, by carefully calibrating the simulation model, we were able to make the Vissim model follow the real-world traffic closely.

Another take away from this experiment is that to fully compare the real-world data and the Vissim simulation we need large amounts of data. Data from many different days and times are needed to better capture the dynamics of traffic flow and further improve the Vissim model. Traffic simulation models can be used as one of the most efficient ways to determine parameters and conditions which can improve traffic flows in cities. Accurate simulation models make it possible to simulate various traffic scenarios, as well as to effectively reduce delays, safety concerns, environmental impacts, and other community concerns associated with truck traffic.

The use of such technologies in public spaces requires public trust. The community datawalk conducted in this report showed a heightened awareness and an enhanced curiosity in participants' feedback regarding data-gathering technologies in their built environment. During conversations with the study facilitator, participants said the experience boosted their knowledge regarding the types of platforms and devices that routinely collect personally identifiable information. Similarly, participants said the community datawalks revealed the ubiquitous nature of smart technologies deployed by both the City and private entities. The three datawalks conducted for this project shed light on how some Long Beach residents perceive a broad range of smart technologies, as well as the data practices their owners employ.

Significantly, participants voiced concerns that these smart technologies can violate data privacy rights. While datawalk participants acknowledged use of personal computing devices and related online and social media platforms, they also expressed concerns about data collected by those technologies and platforms and, further, where that data is shared and how it is used. This sentiment is exemplified by comments such as, "I do not like people having videos/pictures of me. Who knows what they do with that [data]." This finding suggests a clear need for the City of Long Beach to develop policies that provide transparency and accountability surrounding the deployment of devices and platforms that collect personally identifiable information.

Furthermore, many participants recognized the potential for law enforcement technologies to disproportionately impact people of color and low-income communities due to long-standing tensions between residents and the Long Beach Police Department. These findings strongly suggest that, as cities deploy intelligent infrastructure, they must put equal effort into fostering trust, practicing transparency, and engaging the public. Long Beach officials are attempting to operationalize these goals through the Framework for Reconciliation adopted by City Council in June 2020 [52]. The framework proposes the

potential regulation of facial recognition technology and algorithms used in policing. These steps surrounding technology in policing is an example of the transparent policies and practices that should be applied to the widespread use of technology in the city.

Through its Smart City Initiative, Long Beach policymakers are in the early phases of creating a process for involving residents in the adoption and implementation of smart technologies. The initiative is centered around four guiding principles: (1) design for equity; (2) earn public trust; (3) cultivate local expertise; and (4) build civic resilience [53]. By operationalizing these principles, City officials will help ensure that the smart technologies they deploy will protect people's privacy. Afterall, "the freedoms intrinsic to our public spaces make democracy possible" [54].

For most Americans, any discussion of a future where every move they make is monitored using next generation cameras and sensors evokes fears of a surveillance state where data privacy and civil liberties for citizens are diminished. The research methodology in this report and best practices for smart community governance featured in the City of Long Beach's Data Privacy Guidelines provides a blueprint for formal deployment of a real-world test lab at the intersection of Santa Fe and Pacific Coast Highway.

The recommended next steps for deployment of a real-world test lab at the intersection of Santa Fe and Pacific Coast Highway reflect extensive outreach with leadership at the City of Long Beach Public Works Department as well as Caltrans leadership with jurisdiction over data related to Pacific Coast Highway, which is a state route. The outreach conducted by the research team confirmed that neither the City of Long Beach nor the State of California can provide the mobility data required to comprehensively test the truck traffic technologies in this study. In order to implement real-world test labs capable of testing next generation technologies, research teams of the future will need to install cameras, microsensors, and/or drones (compliant with Federal Aviation Administration requirements) capable of gathering mission critical research data. Fortunately, there is precedent for this approach. During a research interview, the City of Long Beach Public Works Director recommended a formal process for our research team to submit a request for a Long Beach City Council permit approval to install cameras and censors at targeted intersections in the city to test next generation smart communities technologies (see Appendix D). The insights from this research effort have established a measurable, legal, and civically supported approach to impelenting real world test labs to test and refine the deployment of intelligent transportation systems that improve the safety and efficiency of mobility systems without sacrificing the data privacy and civil liberties of the citizens who rely on those systems.

6. References

[1] National Transportation Operations Coalition, "National Traffic Signal Report Card," Institute of Transportation Engineers, Washington, DC, 2012.

[2] Caltrans, "Traffic Signal Operations Manual," Caltrans, 2020.

[3] R. T. Sriram, T. Teja and D. N. K. Chikkakrishna, "Application of Machine Learning Techniques in Delay Modelling for Heterogenous Signalized," in *Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, 2020.

[4] Transportation Research Board, Highway Capacity Manual: A Guide for Multimodal Mobility Analysis, Washington, DC: National Academies Press, 2016.

[5] D. Zhen, Z. Fang and W. Yongqiang, "Artificial Neural Network Delay Model for Traffic Assignment Incorporating Intersection Delay Costs," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2132, pp. 25-32, 2009.

[6] F. L. Washbrun and S. S. Mannering, in *Principles of Highway Engineering and Traffic Analysis*, 5 ed., Hoboken, NJ, Wiley & Sons, 2013, pp. 225-279.

[7] City of Long Beach Development Service Department, "City of Long Beach & TOD Pedestrian Master Plan," Los Angeles County Metropolitan Transportation Authority, Long Beach, Feb. 2016.

[8] Federal Highway Adminisration, "FHWA Operations Support - Port Peak Pricing Program Evaluation," U.S. Department of Transportation, Washington, DC, 2009.

[9] US Dept. of Transportation Bureau of Transportation Statistics, "Freight Facts and Figures: Freight Transportation System Extent & Use," US Dept. of Transportation Bureau of Transportation, 2020. [Online]. Available: https://data.bts.gov/stories/s/Freight-Transportation-System-Extent-Use/r3vy-npqd. [Accessed 25 May 2021].

[10] A. Saha, S. Chandra and I. Ghosh, "A comparison of delay at signal controlled intersections based on Different methods," in *12th Transportation Planning and Implementation Methodologies for Developing Countries*, IIT Bombay, India, 2016.

[11] A. Jamal, M. Tauhidur Rahman, H. M. Al-Ahmadi, I. Ullah and M. Zahid, "Intelligent intersection control for delay optimization: Using meta-heuristic search algorithms," *Sustainability,* vol. 12, no. 5, p. 1896, 2020.

[12] L. Pengkai, Y. Xiaoguang, W. Wei and Y. and Yawen, "Measuring control delay at signalized intersections based on vehicle infrastructure integration system," in *International Conference on Optoelectronics and Image Processing*, 2010.

[13] J. a. D. S. B. a. M. D. a. I. I. a. G. P.-O. Haddad, "Optimal steady-state control for isolated traffic intersections," *IEEE Transactions on automatic control,* vol. 55, no. 11, pp. 2612--2617, 2010.

[14] I. Shatnawi, P. Yi and I. and Khliefat, "Automated intersection delay estimation using the input--output principle and turning movement data," *International Journal of Transportation Science and Technology,* vol. 7, no. 2, pp. 137-150, 2018.

[15] D. a. N. H. Ma, "Cycle length optimization at isolated signalized intersections from the viewpoint of emission," in *Traffic and Transportation Studies 2010*, 2010, pp. 275 - 284.

[16] A. Baz, P. Yi and A. and Qurashi, "Intersection control and delay optimization for autonomous vehicles flows only as well as mixed flows with ordinary vehicles," *Vehicles,* vol. 2, no. 3, pp. 523 - 541, 2020.

[17] Y. Liang, S. Zhang and Y. and Wang, "Data-driven road side unit location optimization for connected-autonomous-vehicle-based intersection control," *Transportation Research Part C: Emerging Technologies,* vol. 128, pp. 1 - 20, 2021.

[18] "Port of Long Beach, Ports Facts and FAQs," Port of Long Beach, [Online]. Available: https://polb.com/port-info/port-facts-faqs#facts-at-a-glance. [Accessed June 2021].

[19] "PTV-Group," [Online]. Available: https://libraryguides.vu.edu.au/ieeereferencing/webbaseddocument. [Accessed June 2021].

[20] P. Ioannou, "Design and Evaluation of Impact of Traffic Light Priority for Trucks on Traffic Flow," METRANS Transportation Center, Los Angeles, 2015.

[21] "CB Insights," [Online]. Available: https://www.cbinsights.com/research/what-are-smart-cities/ . [Accessed 15 december 2020].

[22] K. Heidelberg, Trust in smart city systems: Characteristics and key considerations, Cybersecurity and Infrastructure Security Agency. , 2020.

[23] B. Chan, "A smart city ecosystem framework for building sustainable smart cities," Strategy of Things, 24 January 2018. [Online]. Available: https://strategyofthings.io/smart-city-ecosystem. . [Accessed 7 September 2020].

[24] B. Chan, "Building trust in the smart city: Think beyond cybersecurity and privacy," HOT World, [Online]. Available: https://www.iiot-world.com/smart-cities-buildings-infrastructure/smart-cities/building-trust-in-the-smart-city-think-beyond-cybersecurity-and-privacy/. [Accessed 20 March 2019].

[25] R. Garcia, " Smart City Initiative," City of Long Beach, [Online]. Available: http://longbeach.gov/smartcity/. [Accessed 22 October 2019].

[26] "QuickFacts: Long Beach City, California," U.S. Cencus, 2019. [Online]. Available: https://www.census.gov/quickfacts/longbeachcitycalifornia.

[27] K. a. d. L. M. van Es, "Data with its boots on the ground: Datawalking as research method," *European Journal of Communication,* vol. 35, no. 3, pp. 278 - 289, 2020.

[28] W. Matos, "Walking and rhythmicity: Sensing urban space," *Journal of Urban Design,* vol. 13, no. 1, pp. 125-139, 2008.

[29] R. Solnit, Wanderlust: A history of walking, New York: Viking Penguin, 2001.

[30] V. Mayer-Schönberger and K. and Cukier, Big data: A revolution that will transform how we live, work, and think, New York: Houghton Mifflin Harcourt, 2013.

[31] Boyd and a. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Informacios Tarsadalom,* vol. 12, no. 2, pp. 662-679, 2012.

[32] L. Forlano, "Flashmob ethnography workshop.," in *Ethnographic Praxis in Industry Conference Proceedings*, Tokyo, Japan., 2010.

[33] A. Powell, "Data walks and the production of radical bottom-up data knowledge," in *International Communications Association conference*, San Diego, California, USA, 2017.

[34] L. van Zoonen, F. Hirzalla, J. Engelbert, L. Zuijderwijk and L. and Schokker, "Seeing more than you think': a 'data walk'in the smart city," *Public engagement with the smart city. Bang the Table: Carleton, Australia,* 2017.

[35] M. Kassner, "Former NSA and CIA director recommends managing consequences instead of vulnerabilities," TechRepublic. , 2 June 2016. [Online]. Available: http://www.techrepublic.com/article/former-nsa-and-cia-director-recommends-managing-consequences-instead-of-vulnerabilities/. [Accessed 2016].

[36] T. Julsrud and J. and Krogstad, "Is there enough trust for the smart city? Exploring acceptance for the use of mobile phone data in Oslo and Tallinn," Technological Forecasting and Social Change, 161, 120314. , December 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7501060/#.

[37] L. Zucker, "Production of trust: Institutional sources of economic structure, 1840-1920," in *Research in Organizational Behavior, Vol. 8*, Greenwich, JAI press, 1986, pp. 53-111.

[38] R. Hardin, Trust, Cambridge: Polity Press, 2006.

[39] F. Fukuyama, Trust: The social virtues and the creation of prosperity, New York: The Free Press, 1995.

[40] S. Natale, Corporate structures, business and the management of values, Lanham, MD: University Press of America, 1998.

[41] D. Nather, "Axios-Ipsos poll: Black Americans' police experiences are getting worse.," Axios, 22 May 2021. [Online]. Available: https://www.axios.com/axios-ipsos-poll-black-americans-police-dd91c883-1b3b-4eba-87e2-726357d108d7.html..

[42] H. Nissenbaum, Privacy in Context: Technology, policy and the integrity of social life, Standord, CA: Stanford University Press, 2010.

[43] Q. H. Cao, I. Khan, R. Farahbakhsh, G. Madhusudan, G. M. Lee and N. and Crespi, "A trust model for data sharing in smart cities," in *IEEE International Conference on Communications (ICC)*, 2016.

[44] P. Cardullo and R. and Kitchin, " Being a 'citizen' in the smart city: Up and down the Scaffold of smart citizen participation in Dublin, Ireland," *GeoJournal,* vol. 84, pp. 1-13, 2019.

[45] S. Arnstein, "A ladder of citizen participation," *Journal of American Institute of Planners,* vol. 35, no. 4, pp. 216-224, 2007.

[46] City of Long Beach, "Data Privacy Guidelines.," 2021. [Online]. Available: https://longbeach.gov/smartcity/guiding-principles/data-privacy/.

[47] Long Beach Department of Public Works, "City of Long Beach truck routes," 2019. [Online]. Available: http://www.longbeach.gov/globalassets/pw/media-library/documents/resources/general/maps-and-gis/truckroutemap_web.

[48] Long Beach Transit, "Stay connected with LBT: Text for Next.," 2021. [Online]. Available: https://ridelbt.com/connect/.

[49] Moovit , "Moovit," 2021. [Online]. Available: https://play.google.com/store/apps/details?id=com.tranzmate.

[50] V. Osier, "After promising not to, LBPD says it accidentally shared license plate data with ICE.," Long Beach Post. , 21 December 2020. [Online]. Available: https://lbpost.com/news/immigration/lincense-plate-readers-ice-long-beach-police-share-accident..

[51] K. Flores, "City Council to decide whether to buy controversial license plate readers," Forthe, 17 November 2020. [Online]. Available: https://forthe.org/journalism/license-plate-readers/.

[52] City of Long Beach, "Framework for Reconciliation," 26 June 2020. [Online]. Available: https://longbeach.legistar.com/View.ashx?M=F&ID=8595273&GUID=107D5EFA-D10F-4444-B35A-3E7C272887BD. [Accessed January 2022].

[53] City of Long Beach, "Smart City Initiative," January 2021. [Online]. Available: https://longbeach.gov/globalassets/smart-city/media-library/documents/smart-city-initiative-strategy---approved. [Accessed 26 January 2022].

[54] R. Williams, "Whose streets? Our streets! Belfer Center for Science and International Affairs," Harvard Kennedy School, August 2021. [Online]. Available: https://www.belfercenter.org/sites/default/files/2021-08/WhoseStreets.pdf.

[55] M. Fellendorf and P. Vortisch, "Microscopic Traffic Flow Simulator VISSIM," in *Fundamentals of Traffic Simulation*, New York, NY, Springer, 2010, pp. 63-93.

[56] B. A. a. H. A. S. El Tanatwy, "Design of reinforcement learning parameters for seamless application of adaptive traffic signal control," *J. Intell. Transp. Syst. Technol. Planning Operation,* vol. 18, no. 3, pp. 227-245, 2014.

[57] S. S. a. B. Abdulhai, "Deep learning vs. discrete reinforcement learning for adaptive traffic control," *Proc. 21th Int. IEEE Conf. Intell. Trasp. Syst ,* Vols. 2018-Nov, pp. 1215-1220, 2018.

[58] J. Zeng, "Adaptive traffic signal control with deep recurrent Q-learning," *Proc. IEEE Intel. Veh. Symp. (IV),* pp. 1215-1220, 2018.

[59] X. Liang, "Deep reinforcement learning for traffic light control in vehicular networks," *IEEE Trans. Veh. Technol,* 2018.

[60] L. Y, "Intelligent traffic light control using distributed multi-agent Q-learning," *in Proc. 21st Int. IEEE Conf. Intell. Transp. Syst. (ITSC),* vol. 2018, pp. 1-8.

[61] L. Hao, H. Ruihua, Z. Ke and L. Jing, "A Neural Network Model for Travel Time Prediction," in *IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS)*, Shanghai, China, 2009.

[62] City of Long Beach Development Services, "Long Beach 2040 General Plan," 2019. [Online]. Available: https://www.longbeach.gov/lbds/planning/initiatives--programs/long-beach-2040-general-plan/.

[63] Electronic Frontier Foundation, "Atlas of surveillance. Long Beach Police Department: Automated license plate readers.," 2021. [Online]. Available: https://atlasofsurveillance.org/a/aos0865-long-beach-police-department-automated-license-plate-readers.

[64] E. Latham, "Three tips for boosting citizen engagement," Government Technology, 2015. [Online]. Available: http://www.govtech.com/opinion/3-Tips-for-Boosting-Citizen-Engagement.html.

[65] R. Robinson, "Why smart cities still aren't working for us after 20 Years. And how we can fix them," The Urban Technologist, 1 February 2016. [Online]. Available: https://theurbantechnologist.com/2016/02/01/why-smart-cities-still-arent-working-for-us-after-20-years-and-how-we-can-fix-them/.

[66] City of Long Beach, "City Manager Appoints Assistant Chief Wally Hebeish as New Chief of Police," 22 December 2021. [Online]. Available: https://www.longbeach.gov/press-releases/city-manager-appoints-assistant-chief-wally-hebeish--as-new-chief-of-police/. [Accessed 26 January 2022].

7. Data Management Plan

Types of data, samples, and other materials to be produced in the course of the project

This project develops an artificial neural network (ANN), a class of Artificial Intelligence (AI) systems, to accurately model and predict future delays at an intersection. Developing such modeling and prediction systems raises considerable data privacy concerns and it is incumbent upon municipal, state, and federal branches of government to prioritize citizens and their concerns before the implementation of new smart community technologies that are fueled by unprecedented levels of data collection. We collected qualitative data during “datawalks” designed to gauge comfort levels and attitudes toward devices that collect personally identifiable information. Study participants encountered public WiFi routers, surveillance cameras, automated license plate readers and other surveillance technologies. They used a custom app to respond to prompts related to data collection, sharing and analysis. Study participants’ responses, along qualitative data collected during a “debriefing” conversation following each walk, provided insights into residents’ attitudes toward smart communities technologies and identified privacy concerns.

Data files collected through the mobile app: Study participants used their personal mobile devices to access a custom app (stored in the cloud) during each datawalk. Prompts built into this app were designed to help researchers gain a better understanding of study participants’ comfort and trust levels with various “surveillance” technologies deployed in our built environment. A total of 32 volunteers participated in three separate datawalks. The mobile application “backend” saved their anonymous responses as a single Excel file, which PI Shaffer copied to the hard drive on her personal computer, which is password protected. PI Shaffer saved the file according to a format meant to be descriptive and consistent. This format will enable the researchers to accurately identify when the data were collected, in anticipation of future datawalks. PI Shaffer also facilitated “debriefing” conversations following each of the three datawalks. Each conversation lasted approximately 20 minutes. PI Shaffer audio recorded the interviews and, subsequently, transcribed them as separate MS Word documents. She saved each Word document according to a format meant to be descriptive and consistent, such as *Shaffer_20210701_Datawalk1.*

Data files generated through Vissim: To model traffic flow, the microscopic traffic simulation software (Vissim) developed by PTV Group will be used in this project. Vissim software license will be installed on a University owned server, and the team members will have access to the software. This software will be mainly used to simulate traffic flows of street intersections in the City of Long Beach where we have observed high volumes of heavy trucks are passing through.

Data generated by Vissim are all simulated data and follow the file format generated by the software. For some post processing, technical software MATLAB, developed by MathWorks will be used. This software has already been installed on a University wide server and all faculty and students have access to that software.

Documentation files: In order to provide clear context for a full explanation of data creation and analysis, the co-PIs will make a number of documentation files available as part of the data deposit. These include:

- the protocols/list of questions embedded into the mobile app, questions posed prior to each walk, and questions posed during “debriefing” conversations following each walk;
- aggregated demographic information about study participants, devoid of identifying information;
- coded transcripts;
- IRB approvals with accompanying consent scripts used in the data collection;
- this data management plan.

Management of the files during the data collection and analysis phase: All Word files will follow the same naming convention described above. As previously noted, the comments from study participants are anonymous. During analysis of the data, the PI manually coded each manuscript (identifying key themes and separating the data into representative categories). Coded transcripts were securely stored on Microsoft 365 OneDrive, a platform licensed by CSULB. The original materials were placed in a locked cabinet drawer in PI Shaffer’s CSULB office and serve as a third copy of the source material. She is the only person with access to this cabinet drawer. The paper files will be destroyed three years after completion of this smart technologies privacy project.

The researchers will take steps to ensure preservation of MS Word transcripts, Excel files and all other documents generated for this project. Because the text files are both anonymized and encrypted, the cloud storage platform Microsoft 365 OneDrive is a suitable back up storage option. Specifically, the co-PIs created a OneDrive folder for the entire traffic flow project. They created sub-folders for storing raw data, as well as for documents produced through analysis of the raw data. A major benefit of archiving these documents on OneDrive is the platform’s ability to recover files that may be inadvertently deleted or lost. Additionally, using a cloud storage service ensures data are routinely backed up. (The co-PIs' institution retains a license for Microsoft Office 365 OneDrive, so no grant funding was allocated for this storage service.)

Standards to be used for data and metadata format and content

Since the data will be shared via the Qualitative Data Repository (QDR), the metadata policy of the repository will be in effect for the published materials (project and file-level). Specifically, the repository uses Data Documentation Initiative (DDI 2.5) and DataCite 3.1 schemas. The co-PIs previously contacted the repository and will remain in ongoing consultations with QDR curation staff in order to collect/provide all the substantive information needed for the creation of the structured metadata.

Methods and policies for providing access and enabling sharing

The researchers will share all data, documentation and analysis files listed above The co-PIs promised confidentiality to all study participants, as articulated in the informed consent document each volunteer was required to sign. While the co-PIs will work with QDR staff to prepare the data deposit during the course of the research project, the publication of the shared files will be embargoed until six months after data collection is complete.

Provisions for re-use, re-distribution, and the production of derivatives

Materials generated via the project will be disseminated in accordance with CSULB policies. According to QDR's policies, the 4 co-PIs will retain full ownership of the materials created during the project. No special access or re-use provisions will be applied to the shared files, aside from the repository's standard requirement for registration and login before a secondary user may access the files.

Methods for archiving and preserving access to data and materials

All data generated an/or processed and by software packages Vissim and MATLAB will be saved on the university-licensed OneDrive (a file sharing platform) for access of team members.

In order to ensure the social science research community has long-term access to the data, the full collection of text-based interview transcripts (and any appropriate metadata) will be stored with the QDR. The QDR is one of the few dedicated archives for storing and sharing digital data generated or collected through qualitative research in the social sciences. The QDR provides search tools to facilitate the discovery of data (e.g., the repository's catalog is shared via DataCite for easier discovery; the project will be issued a digital object identifier (DOI) for appropriate citation). Given that this educational technology project is relevant to the digital privacy impacts of smart city technologies—and the political economy of digital technologies, generally—the QDR is a particularly appropriate repository for archiving these data, research tools, and resulting policy and public education materials.