Abstract—Traffic congestion in urban areas is posing many challenges, and a traffic flow model that accurately predicts traffic conditions can be useful in responding to them. With the limitation of infrastructure and the difficulty for real world experiment, traffic simulation is a good tool for the validation and future potential applications of the stochastic mode. This research presents a stochastic traffic flow model which uses a stochastic partial differential equation to describe the evolution of traffic flow on the highway. The stochastic model is calibrated and validated by real traffic data and is proved to have better predictive power than the deterministic model. Also, a microscopic traffic simulation is built and calibrated by real world traffic data on a highway, and the validation result shows that the simulation is capable to represent the real traffic. By using traffic simulation to further validate the capability of the stochastic model, the prediction work good at most locations, and the interpolation error may be improved by considering the influence of ramps and the number of lanes.

I. INTRODUCTION

Just as the normal blood circulation necessitates a healthy body, the smooth traffic flow is necessary for the healthy business and community development, in a city as well as a region. Yet traffic congestion that inflicts uncertainties, drains resources, reduces productivity, stresses commuters, and harms environment is haunting many cities and communities. A study estimated that 32% of the daily travel in major US urban areas occurred under congested traffic conditions [1]. Also, in 2007, congestion caused urban Americans to travel an extra 4.2 billion hours and to purchase an additional 2.8 billion gallons of fuel thus incurring the congestion cost of 87.2 billion dollars - an amount more than 50% over the costs incurred a decade ago [2].

To cope with the problem, major efforts at reducing traffic congestion have been undertaken. Various forms of Intelligent Transportation Systems (ITS) that take real-time traffic data for decision support have been developed for this purpose, and the success of these efforts at controlling congestion requires an accurate prediction of the evolution of traffic flow. For this purpose, reliable and robust traffic flow models are indispensable.

Macroscopic and microscopic are two main mathematical approaches to model the traffic flow. The macroscopic approach studies properties induced by the interaction of a group of vehicles and ignores the detailed identities of individual vehicles. The fluid-dynamical models treat traffic flow as a compressible fluid. In such models, the core variables for vehicles, flow rate \( q \), traffic density \( \rho \) and speed \( v \), are taken as functions of space and time. The basic idea is to build a partial differential equation for the variables and then solve the equation for the variables. A classical macroscopic model is the Lighthill-Whitham-Richards (LWR) model [3], [4].

In recent years, abundant traffic data has become available with the extensive use of the detection and surveillance devices on the road. In many states of the United States, the administrators of the transportation departments have constructed databases that can provide historical and real-time traffic data to the public. Thus, these data sets provide researchers an opportunity to build data-driven models for predicting the traffic flow.

Noting the special stochastic properties of traffic flow, we propose a stochastic partial differential equation (SPDE) model for traffic flow that captures the variation of congestion with space and time, and gradually reverts to the mean values after any random perturbations. Building on the classic LWR traffic flow model, our SPDE model describes the evolution of traffic flow by taking real-life data to predict traffic flow across space and time. The results from the numerical tests suggest that the SPDE model is capable of making accurate predictions of the traffic flows. It can be used as a primary building block within an Intelligent Transportation System (ITS) to provide intelligent decision support.

Microscopic approach which takes the view to study the movements of individual vehicles is another way to model and predict traffic flow. The approach considers the interaction among individual vehicles and the stochastic driving behavior of drivers[5]. One example of the microscopic approach is the car-following mode[6] which usually uses simulation models, such as the cellular model, to understand the traffic flow. For example, Boel and Milhaylova[7] proposes a compositional stochastic model for real time freeway traffic simulation. Chrobok et al.[8] predicts the traffic flow at a fixed location in the highway using an online cellular-automation simulator. Park and Schneeberger[9] create a simulation based model with the detail of model calibration and validation. Simulation model based on a pheromone communication model has been used to control traffic light [10], but it relies on sensors at each site to perform the prediction.

In recent years, Microscopic traffic simulation modelling has been an increasing popular tool in the area of transportation. It has been used for wide range applications in
network design, analysis of transportation problems, and the evaluation of ITS and traffic management strategies. One benefit of having simulation model is that it provides traffic information from system to individual level. Especially for the individual vehicle information, some researches used experiment on the real road to obtain individual vehicle speed [11], [12], but it is limited by sensors or probe vehicles, and the process may cost a lot of time and money. In order to have better insight of the transportation system with the inherent complicity, stochastic and dynamic nature, traffic simulation model becomes an efficient approach. A traffic simulation with appropriate assumption and calibration can be used to validate traffic congestion model [10] and to observe macroscopic phenomena in order to understand the influence of individual driver’s behavior [13].

There are many well developed microscopic traffic simulation models like AIMSUM, MITSIM, PARAMICS and VISSIM. In this paper, PARAMICS which is developed by Quadstone Limited is used for traffic simulation. It is a very comprehensive stochastic simulation model with a wide variety of traffic modelling application. Three PARAMICS product suites were used, including Modeller, Estimator, and Prosessor. Modeller is the modelling tool for building network, simulating with visualization and data output. Estimator is a tool to estimate origin-destination demand matrix by count data from observation. Processor allowed the user to set up multiple simulation runs in batch mode.

The objective of this paper is to present the stochastic traffic flow which can predict traffic flow, and how to use a microscopic traffic simulation to validate the stochastic model. The next section of this paper describes the proposed stochastic model, and the third section is about the modelling and calibration of simulation model. Validation results using real traffic data are presented in the fifth section, which is followed by the last section of the conclusion.

II. STOCHASTIC TRAFFIC FLOW MODELING

Our macroscopic model is built on the classical LWR model. The classical LWR model is a partial differential equation on traffic flow derived from conservation law, and we generalize it to a stochastic version by introducing a stochastic forcing function. The forcing function incorporates a Brownian Sheet[14] and a mean reverting term. The notation is described as below:

- \((x, t)\) is the space and time pair, \(x \in [0, L]\) and \(t \in [0, T]\);
- \(Q(x, t)\) is the volume (i.e., number of vehicles passing through per unit time) at location \(x\) and time \(t\);
- \(\rho(x, t)\) is the density (i.e., number of vehicles per unit distance) at location \(x\) and time \(t\);
- \(v(x, t)\) is the average velocity of all vehicles at location \(x\) and time \(t\);
- \(W(x, t)\) is the Brownian Sheet, a Gaussian process indexed by two parameters \(x\) and \(t\). The definition and properties of the sheet are developed in Walsh [14].

For simplicity, sometimes we use \(Q, \rho, v\) in subsequent discussion with the understanding that these quantities are dependent on \(x\) and \(t\).

A. The LWR Model

Since proposed in 1950s, the classical LWR model has been the building block of many macroscopic traffic flow models. Equation (1) describes LWR model, which consists of two equations. The first one is derived from the conservation law: the difference of the inflow and the outflow in a cell is equal to the increment of the vehicles in the cell. The second one is called fundamental flow relationship between \(Q, \rho\) and \(v\).

\[
\frac{\partial}{\partial t} \rho + \frac{\partial}{\partial x} Q = 0
\]

\[
Q = \rho \cdot v
\]

B. The Speed-Density Function

As shown in the second one of equation (1), given any forms or values for any two of \(Q, \rho\) and \(v\) the form or value of the remaining third can be determined. If the speed-density function \(v(\rho)\) is known, the SPDE is determined only by the functional \(\rho\).

Several functional forms of the speed and density relationship have been proposed in the literature. Table I summarizes the relationships that have been frequently used in the literature. Such relationships are generally used for different purposes; e.g., Greenberg is more appropriate for congested traffic and Underwood for free-flow. It is possible to mix these relationships in application; e.g., adopting the Greenberg relationship in congested space-time sections and Underwood for the free-flow sections. Our SPDE model adopts the fourth function, \(v = \min\{v_f, \alpha \rho^m\}\), which we call the log piecewise linear model, as the speed-density function.

<table>
<thead>
<tr>
<th>Functions</th>
<th>(v(\rho))</th>
<th>(Q(\rho) = \rho \cdot v(\rho))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenshields</td>
<td>(v_f(1 - \rho/\rho_j))</td>
<td>(\rho v_f(1 - \rho/\rho_j))</td>
</tr>
<tr>
<td>Greenberg</td>
<td>(v_0 \ln(\rho_j/\rho))</td>
<td>(\rho v_0 \ln(\rho_j/\rho))</td>
</tr>
<tr>
<td>Underwood</td>
<td>(v_f \exp(-\rho/\rho_j))</td>
<td>(\rho v_f \exp(-\rho/\rho_j))</td>
</tr>
<tr>
<td>Log Piecewise Linear</td>
<td>(\min{v_f, \alpha \rho^m})</td>
<td>(\rho \min{v_f, \alpha \rho^m})</td>
</tr>
</tbody>
</table>

C. The Stochastic Modeling of the Traffic Flow

The LWR model is a first-order partial differential equation that assumes that the traffic flow reaches an equilibrium immediately. It predicts traffic flow well in relatively heavy traffic when the effects of individual driver behavior are minimal. However an effective traffic flow model must also account for the stochastic and time-varying nature of traffic flow (Saigal and Chu[15]). The traffic depends on the time of the day, the day of the week and the locations of segments of freeway. The exact amount is stochastic in nature, and is affected by macroscopic effects such as weather conditions, special events, etc. At any time, the traffic at a location deviates from its “nominal” value because of the unpredictable and uncontrollable microscopic phenomena like sudden acceleration/deceleration, lane shifts, lane surface conditions,
traffic density. The deterministic constant function \( a \) (c.f. Walsh[14]), is added to the classic LWR model on down. Our SPDE is expressed as:

\[
\frac{\partial}{\partial t} \rho(x,t) + \frac{\partial}{\partial x} Q(\rho(x,t)) = g(\rho(x,t), x, t),
\]
\[
Q(\rho(x,t)) = \rho(x,t) \cdot v(\rho(x,t)),
\]
\[
g(\rho, x, t) = a(x,t) + b(x, t) \cdot \rho + \sigma(x, t) \cdot W(dx, dt).
\]

A forcing term \( g(\rho(x,t), x, t) \), composed of deterministic functions \( a(x,t) \), \( b(x, t) \), and \( \sigma(x, t) \) and a Brownian Sheet \( W \) (c.f. Walsh[14]), is added to the classic LWR model on traffic density. The deterministic constant function \( a(x,t) \) is the drift term designed to capture the effects of the means of inflow/outflow and other factors that may affect flow conservation. We expect it to be positive at locations \( x \) near entrances and times \( t \) when traffic is entering, and negative at times \( t \) when traffic is leaving the highway. The function \( \sigma(x,t) \) is designed to capture the magnitude of the resulting volatility of the disturbance to the flow conservation due to the microscopic effects along the highway. The Brownian sheet \( W \) is a Gaussian process indexed by two parameters \( x \) and \( t \) with mean 0 and covariance \( E(W(x,s)W(y,t)) = \min(x,y) \cdot \min(s,t) \), and the term \( U(dx, dt) = b(x,t) \cdot \rho(x,t) + \sigma(x,t) \cdot W(dx, dt) \), borrowing from the nomenclature of the stochastic differential equation literature, converts the Brownian Sheet into an Ornstein-Uhlembek (OU) Sheet, \( U \). This choice makes the process "mean-reverting", i.e., after a disruption moves the process way from the its mean behavior, the effect of this disruption dissipates in time and space and the process returns to its mean behavior at a rate determined by the magnitude of the parameter \( b(x,t) \), which we expect to be negative. Related and other properties of the OU Sheet can be found in Walsh[14]. The calibration and validation of this model will be shown in section IV.

**III. TRAFFIC SIMULATION MODEL**

Our microscopic traffic simulation model is based on a real world traffic network. Interstate 95 (I-95) provides south-north transportation between eastern Virginia and Washington DC, and about 18 miles of I-95 Northbound in northeastern Virginia was modelled in PARAMICS Modeller. Figure 1. shows the map of I-95 and the illustration of ramp and loop detector locations. Since there is no dramatic curve for the highway segment, the geometry of the highway can be simplified as a straight road, and other network characteristics including ramp locations, number of lanes, and speed limits were modelled according to the information from Google Map. Each exit of this highway segment is assigned a zone area which vehicles are released into or removed from the network model.

Traffic data, including vehicle counts (vehicles per minute) and speed (miles per hour), was collected in every minute from 23 inductive loop detectors located along the highway segment from February 1st to June 1st in 2009. Excluding possible malfunction loop detectors with abnormal record, data from 16 locations can be used as observation data. The GEH Statistic is the criteria used for comparing the traffic volumes between simulation and observation. The GEH was invented by a British transportation engineer, Geoffrey E. Havers, and is widely used for a variety of analysis purposes in traffic engineering, traffic forecasting, and traffic modelling. The formula for the **GEH Statisticis**:

\[
GEH = \sqrt{\frac{(VOL_{obs} - VOL_{sim})^2}{(VOL_{obs} + VOL_{sim})/2}}.
\]

\( VOL_{obs} \) and \( VOL_{sim} \) are traffic volume from the observation and the simulation respectively. For traffic modelling work in the "baseline" scenario, a GEH of less than 5.0 is considered a good match between the modelled and observed volumes, and 85% of the volumes in a traffic model should have a GEH less than 5.0 [16].

**A. Original-destination Demand Estimation**

Data from 10 of the functional loop detectors was used to estimate the original-destination demand matrix (OD matrix). The traffic counts in one-minute interval were converted into vehicles per hour (VPH). By assuming that the travel demand would be similar on every Monday, the VPHs on each Monday during the two months were averaged to represent VPHs for Monday. Since the travel demands change with time, in order to obtain a more accurate simulation model, we need to estimate time-dependent OD demands. For each hour, the VPHs at 11 loop detector locations would be used to estimate one OD matrix for this hour. Estimator of PARAMICS is the tool to estimate OD matrix. Periodic normalisation method was used to obtained initial OD matrix, then incremental method was performed in a 20-minute calculation period for 20 iterations. The aim of the iterative estimation process is to
converge to a solution that minimizes the difference between the modeled flows and the observed flows using arithmetic mean of the GEH values for each Flow. By having estimated OD matrix for each hour, the model is ready for initial tests and calibration process.

B. Model Calibration

The PARAMICS simulation model was calibrated in three stages, including volume-based calibration, speed-based calibration and OD demand calibration. The calibration involves repeated process of adjusting parameters, test run and result analysis.

For volume-based calibration, the GEH values at different location was used as the first scan of the problems. For every location with high GEH, the traffic volume from the simulation compared to observation data in minutes would be examined. Since the nature of driving habits would affect the simulation significantly, many driver behavior factors in PARAMICS can be changed to calibrate the model. The default value for mean headway factor is 1.0 second, but on congested freeways, vehicles tend to travel closer and maintain a smaller headway. From previous works, a range from 0.6 to 1.0 second were used as mean headway factor [17], [18]. In this research, the overall mean headway factor is set as 0.8, and 0.6-0.8 may be applied to a particular link based on the volume analysis.

The speed limit in PARAMICS is more like average target speed, though it was set as the speed limit on I-95, calibration for links may be required. Average speed for all the lanes at each loop detector location in the simulation was compared to the recorded speed from observation. Speed limit may be adjusted if there is a significant and consistent difference. If vehicles have unwanted slowdown for merge to other lanes, a shorter reaction time may allow vehicles to respond quicker.

The default reaction time factor is 1.0 second, and a range form 0.4 to 0.6 was used for different traffic models [17]. We use overall mean reaction time as 0.8, and 0.8-1.0 may be applied to particular links. The minimum ramp time is another factor that would influence the ability of drivers to merge with the main traffic at the ramp. Reducing the value allows vehicles to move more quickly into the main highway traffic and may reduce the occurrence of vehicles stopping on the ramp and then slowing down the main highway traffic when they accelerate back to the highway speed after leaving the ramp. 1 second is used for all the on-ramp links instead of the default value 2.

The estimated OD demand matrix represents travel demand for each hour, but the demand may not be uniform distribution during the one-hour period. For example, the traffic demand is very low before 5:00am, and then keeps increasing from 5:00am to 7:00am because of the commuters. The demand distribution during these two hours should increase accordingly. Therefore, the demand for each hour was divided into four 15-minute intervals, and then the percentages for each interval can be assigned. Figure 2. shows an example of how the OD demand distribution modification works to adjust traffic volume in the simulation to fit observation data.

Table II summarized all the parameters used for simulation calibration, including default values in PARAMICS, mean values applied for overall network and values applied for some particular links. These values are calibrated based on the traffic data we obtained from I-95, and may not applied to other instances.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Default</th>
<th>Overall</th>
<th>Each link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed limit (mph)</td>
<td>-</td>
<td>60</td>
<td>45-60</td>
</tr>
<tr>
<td>Reaction time (sec)</td>
<td>1</td>
<td>0.8</td>
<td>0.8-1</td>
</tr>
<tr>
<td>Headway (sec)</td>
<td>1</td>
<td>0.8</td>
<td>0.6-0.8</td>
</tr>
<tr>
<td>Minimum ramp time (sec)</td>
<td>2</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Demand distribution (% per 15-min interval)</td>
<td>Uniform</td>
<td>25/25/25/25</td>
<td>5am-6am: 12/20/20/39 6am-7am: 20/24/28/28</td>
</tr>
</tbody>
</table>

IV. RESULT

A. SPDE Model Calibration and Validation

The SPDE model is first calibrated by the observation data from loop detectors. Then the calibrated model will be validated by testing its prediction accuracy. The procedure of calibrating the parameters \(a, b\) and \(\sigma\) in the forcing function \(g\) in equation (2) is described as follows. For specific \(x\) and \(t\), the value of \(g(\rho, x, t)\) is calculated by finite difference method as equation (4), where observed density and volume are available at those grid points.

\[
g(\rho, x, t) = \frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}Q(\rho(x,t))
\]

\[
\frac{\partial}{\partial t}\rho(x,t) = \frac{\rho(x,t + \Delta t) - \rho(x,t)}{\Delta t}
\]

\[
\frac{\partial}{\partial x}Q(\rho(x,t)) = \frac{Q(\rho(x + \Delta x, t)) - Q(\rho(x - \Delta x, t))}{2\Delta x}
\]

After that, linear regression between \(g(\rho, x, t)\) and \(\rho\) for fixed \(x\) and \(t\) is applied to obtain the value of \(a, b\) and \(\sigma\) at specified \(x\) and \(t\). The data is from the collection of different days but with the same \(x\) and \(t\). One example of linear regression result is shown in figure 3. After the calibration, the SPDE model is validated by testing its prediction accuracy. The procedure of testing is as follows. The observed density at time \(T_{0}\) along the highway is set as the initial condition of
SPDE in equation (2). Then numerical method is applied to obtain the mean of density at future time $T_1$ with $T_1 > T_0$, and the mean is considered as the predicted density at time $T_1$. Comparison between predicted and observed density at time $T_1$ will illustrate the prediction power of the SPDE model.

**B. Simulation Validation**

By considering the stochastic nature of the traffic, the simulation was ran 20 times using different seed numbers which was random generated. The simulation started at 12:00am with an no-vehicle initial condition, and then after releasing the vehicles for an hour, the data was collected from 1:00am to 1:00pm. The results from 20 simulation runs were compared to the observation data from 8 Mondays during the two-month period. Validation result for both mean value and standard deviation (STD) of volume would be investigated. Using the mean value of volume from both simulation and observation, the GEH statistics was calculated for each minute at each loop detector location. By having all the GEH under 5.0, Figure 5. is average GEH for each minute and Figure 6. is average GEH for each station. There is no significant high GEH with respect to either time or location, so that the simulation is capable to represent the real traffic on I-95.

The standard deviation of the volume between 20 simulation runs and 8 observation days was also compared. Figure 7. is the average STD for every minute. The simulation seems to have slightly higher variation than in the real traffic, but the difference is usually less than 3.0, which is relatively small compared to the volume around 100 veh/minute. Both simulation and observation have higher variation when the traffic is more congestion after 6am.
C. Validation of SPDE Model with Simulation

In order to further investigate the prediction ability of the SPDE model, some locations without observation data would be explored by using the data from simulation. Another 20 locations were chosen to have virtual loop detectors in the simulation, which are numbered as v1 to v20 in Figure 1. The density at virtual detectors from simulation would be used as the criteria to evaluate the prediction from SPDE model. Density between 4am to 8am was investigated, because the traffic flow changes from very low density to congestion condition during this period of time. Data from 20 simulation runs at 4:00am was used as the initial condition for the SPDE model, and then performed the 5-minute prediction from 4:05am to 8:00am.

Fig. 8. Prediction error for the virtual detector locations

Figure 8. shows the average percentage error of the prediction for each virtual loop detector. Because the prediction for the virtual detector is the interpolation of the prediction at two closest real loop detectors, so higher errors are expected. Most of the locations have errors under 15%, but three locations which are v3, v7 and v19 have significant high errors caused by over-estimation from the prediction. Refer to the road condition around v3 and v7, they are right after off-ramps, so that the traffic density decreases significantly. However, when the SPDE model interpolate the prediction for virtual loop detectors, the dramatical change caused by off-ramps is not considered. Also, the number of lanes varies from 3 to 5 along the highway, and the traffic density would also increase/decrease suddenly when the number of lanes decreases/ increases, but the SPDE model may not capture this effect by linear interpolation. This may be the reason for higher errors at v11 and v19. The interpolation may also have larger error at the locations like v7, v14 and v15, where the closest real loop detectors are far apart.

V. CONCLUSION

This paper presents a stochastic traffic flow model for traffic prediction and a microscopic traffic model for simulating the real traffic, and both of the SPDE model and the simulation were calibrated by the same traffic data from the field. The methodology of modelling the traffic flow with the stochastic partial differential equation is described in detail, and the result shows that the proposed SPDE model which is capable to accommodate the variance of daily traffic has better predictive power than the deterministic model. For the simulation model built in PARAMICS, the calibration process for some parameters in the microscopic traffic simulation is presented, and the simulation is now an approved tool to represent the real traffic on the highway. The validation of the SPDE model with simulation shows that the prediction at most virtual detector locations can have errors between 9%-15%. For some high error locations, possible issues from the interpolation of the prediction between real detector locations are discussed. With the insightful validation process of this paper, with the simulation as a test platform and data resource, further improvement can be done by including the ramp condition and the number of lanes into the SPDE model.

VI. ACKNOWLEDGMENTS

The authors thank Dr. Francois Dion at Partners for Advanced Transportation Technology (PATH), University of California, Berkeley for his guidance and valuable support in the process of applying PARAMICS model.

REFERENCES