Feasibility Study For Automatic Calibration Of Transportation Simulation Models

Hong Liu[*], Qian Yu[†], Wei Ding[‡], Daiheng Ni[+], Honggang Wang[*], Stephen Shannon[*]
[*]University of Massachusetts Dartmouth, 285 Old Westport Road, N. Dartmouth, MA 02747, USA
[†]University of Massachusetts, 611 N. Pleasant St, Amherst, MA 01003, USA
[‡]University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125, USA
HLiu@UMassD.edu, QYu@geo.UMass.edu, Wei.Ding@UMB.edu, Ni@ecs.UMass.edu,
HWang1@UMassD.edu, SShannon@UMassD.edu

Keywords: sensor and geography networking, data mining, model calibration, traffic flow simulation, transportation network design.

Abstract

This paper presents a high-tech solution to meet the challenges in calibrating transportation simulation models. Like any simulation software, model calibration prior to its application plays a crucial role in producing reliable results. However, transportation professionals face difficulties in performing the daunting tasks of calibrating a model for each transportation network design to satisfy the targeted traffic flow demand, especially during data collection and distillation. Our innovative approach utilizes sensor and geography networking technology to seamlessly collect data about real world network, traffic, and driver behavior. This data is then distilled as needed by data mining before feeding the data to a simulation model. The data is validated automatically to instantaneously reflect the real world and to avoid typographical errors often involved with human intervention, resulting in a more accurate model.

We conduct a feasibility study for our vision of model calibration automation. The research flexes multidisciplinary expertise in traffic flow simulation, geosciences, sensing/networking, and knowledge discovery. As a proof of concept, we implement a prototype that demonstrates how to convert sensor data about traffic flow collected by a state department of transportation into a format taken by CORSIM, a popular traffic simulation model. A running example shows encouraging results.

1. INTRODUCTION

Transportation simulation tools have been used intensively by transportation professionals to aid their decision making of modifying or building infrastructures for increased traffic capacity and improved level of service. Computer models capture the real world by the environmental characteristics such as the number of lanes on a highway and by the traffic flow at microscopic or macroscopic views. However, there is a wide consensus that traffic models, often not properly calibrated or validated with real-world traffic conditions, produce unrealistic or misleading results. Without model calibration to validate that the base-year model reproduces the existing traffic conditions with an acceptable accuracy, any transportation simulation tool would fail to predict the future for a sound design-year model. Transportation professionals such as those in Wisconsin Department of Transportation (DOT) recently issued a guideline that assesses a model with realism tests in terms of mathematical targets and overall traffic patterns [Wisconsin DOT 2010].

Model calibration/validation accompanies a model life cycle as a part of testing processes to compare the model with actual system behavior and to improve the model iteratively. A model life cycle contains four stages: *Problem Definition* to collect data on the real system behavior and to identify system's inputs/outputs, *Model Conceptualization* to design the system structure associating inputs/outputs and to develop theories explaining the system behavior, *Model Implementation* to computerize the conceptual model on a platform with a programming language, and *Model Application* to assist transportation professionals with solving real-world problems or making decisions. Specifically, model calibration should be done after Model Implementation and before Model Application, where various parameters in model structure are adjusted to reach an acceptable level of model accuracy [Ni et al. 2004]. Model developers, model users, and decision makers should be aware of a spectrum of traffic flow modeling. This includes multiple scales such as macroscopic view of traffic operation in a region and microscopic view with details on vehicles interacting with each other to maintain safe positions in traffic stream [Ni 2010].

Traffic data collection coordinating within geographic environment plays an important role during the entire life cycle of modeling. It serves as the basis for the first stage, Problem Definition, continues to guide the rest of the stages, and highlights model calibration at the fourth stage, Model Application. Deploying different technology for traffic data collection and traffic processes has found wide applications in transportation operation and management. For example,

The work is partially supported by UMass President's Office for Science and Technology Initiative fund 2010.

Maryland DOT has piloted a real-time sensor web (integrating satellite remote sensing, low cost UAVs, fleet tracking devices, individual vehicle GPS collections, in addition to geo-based anchored sensors and traffic monitoring cameras) to infer traffic flow and enable dynamic pricing schemes to mitigate congestion [Halem 2007]. We envision automatic calibration of transportation simulation models by collecting traffic data automatically with sensor technology, transferring them seamlessly with wireless communication and the Internet, and using data mining to discover traffic patterns ready for use in model calibration.

In this paper, we propose a novel approach to calibrate transportation simulation models towards automation by seamless traffic data collection and processing using sensor technology, geography networking, and knowledge discovery. Currently, we report the results of our feasibility study. We begin with a historical perspective of transportation simulation over fifty years followed by an analysis of challenges in model calibration. We then present our solution to the challenges that integrates network simulator, traffic simulator, and driver simulator with geography technology, wireless sensor network, and data mining to fuse the real world transportation in its computer model. A quick prototype proves the concept. Finally, we conclude our study and identify the future work.

2. TRANSPORTATION SIMULATION REVIEW

There exists a wealth of literature in traffic flow modeling and simulation and, consequently, only a small portion is cited here to establish the state of the art. Omission of other work is not meant to imply that it is irrelevant or unimportant.

Generally, traffic simulation has been pursued at three levels: macroscopic, mesoscopic, and microscopic. Representing one extreme of the spectrum, macroscopic simulation models traffic flow as a one-dimensional compressible fluid where disturbances in traffic propagate like waves. Fundamental to this type of simulation is the conservation law. The first-order form of this law is mass/vehicle conservation, a model based on which is called a first-order model, such as [Lighthill and Whitham 1955], [Richards 1956]. In addition to vehicles, other quantities such as momentum and energy can also be conserved. A model is of higher-order if it incorporates the latter types of conservation, e.g. [Whitham 1974], [Phillips 1979]. Considering the limited benefit offered by higher-order models may not justify their added complexity [Daganzo 1995b], efforts of numerical solutions and simulators were mainly centered on first-order models. The following are some examples of macroscopic simulators: KRONOS [Michalopoulos 1984], KWaves [Newell 1993], CTM [Daganzo 1994], [Daganzo 1995], FREFLO [Payne 1979], FREQ [May et al 1991], and CORQ [Yager 1975]. Macroscopic simulation can model a very large (geographical) scope but with a low modeling fidelity.

The next level is mesoscopic simulation. This type of model tries to model macroscopic behaviors of traffic flow from the understanding of its microscopic basis and thus have to introduce a statistical approach - such as the kinetic theory [Prigogine 1961]. Though simulators of the kinetic type are rare, another type of mesoscopic simulator, TRANSIMS [LANL 1999] is well-known. Rather than treating traffic as a continuous fluid, TRANSIMS models traffic as discrete particles without masses and personalities. These particles traverse a discretized time-space grid, hopping from one cell to another governed by some pre-defined local rules such as maximum speed constraints. Mesoscopic simulation represents a balance between modeling scope and fidelity.

Next in the spectrum is microscopic simulation which models driver-vehicle units as particles without masses but with personalities. The behaviors of these particles are governed by car-following, lane-changing, and gap-acceptance models. Several approaches have been identified in modeling car-following behavior, among which are stimulus-response models [Gazis et al. 1961], desired measure models [Pipes 1953], [Gipps 1981], psycho-physical models [Wiedemann 1974] , intelligent driver model [Treiber and Helbing 2002], and rule-based models [Kosonen 1999]. Approaches to lane-changing include mandatory and discretionary lane-changing (MLC/DLC) [Gipps 1981], adaptive acceleration MLC/DLC [Hidas and Behbahanizadeh 1999], and autonomous vehicle control [Sukthankar 1997]. Models for gap-acceptance include deterministic models [Velan and Van Aerde 1996], probabilistic models [Hewitt 1992], and neuro-fuzzy hybrid model [Rossi and Meneguzzer 2002]. After continuous evolution and refinement over about half a century, some of these car-following, lane-changing, and gap-acceptance models have resulted in a variety of transportation simulators including CORSIM, VISSIM, INTEGRATION, Synchro, AIMSUN, and S-Paramics, just to name a few. Federal Highway Administration's (FHWA) Next Generation Simulation (NGSIM) program maintains a fairly complete taxonomy of microscopic models and enhances microscopic simulation by adding behavioral models. Representing the current state of the art of traffic simulation, microscopic simulation can achieve a high modeling fidelity but with a limited modeling scope.

3. ISSUES IN MODEL CALIBRATION

A traffic simulation model is a computerized model which is created within a traffic simulator and meant to represent part of the real world system. One of the benefits of creating a computer model is to allow repeated testing of the target system in a safe and controllable environment which is otherwise infeasible in the real world.

Consequently, the trustworthiness of the test results depends largely on how close the simulation model resembles its real world counterpart, and the procedure to ensure such resemblance is called *model calibration*.

Since a real world transportation system consists of roadways, vehicles, and drivers, the major task to calibrate a simulation model is to obtain or fine tune the sets of data that represent the above transportation components. More specifically, items to be calibrated typically include but are not necessarily limited to:

- Data that describe the transportation network: roadways, intersections, and traffic control devices. Details for roadways include number of lanes, speed limit, curvature, road type, etc.; details for intersections are intersection type (at grade, grade separated, and interchange), priority rules, number of approaches, and outlets; details of traffic control devices can be type of device (e.g. signals, signs, and markings), implied rules (e.g. priority rules), and control logic (e.g. timing plans for signals).
- Data that describe the traffic: origin-destination (O-D) flows, traffic mix (types of vehicles in the traffic), and details of vehicles (e.g. power, mass, and dimension).
- Data that describe the driving population: the selection of underlying car-following, lane-changing, gap-acceptance, and route choice models and parameters within these models.

Challenges of traffic simulation model calibration are frequently brought about by data collection and distillation. More specifically, the following issues are identified in the process of model calibration:

- ➢ Resource consumption: Calibrating a simulation can be very time-consuming and labor-intensive. For example, field data collection can go anywhere from a few hours to many years, from one location to a regional network, and from one person to the collaboration and coordination of several teams.
- ➢ Calibration methodology: A conventional calibration process adjusts model parameters until reasonable resemblance between model and the real system is obtained. Such a process is not only tedious but also prone to subjective error. A more systematic approach would be to treat the process as an optimization problem which searches for the best fit between the model and its real world counterpart [Chu et al 2003].
- ➢ Measures of performance: This issue is related to how a calibration process is evaluated. For example, what indicates a successful calibration and when a model is sufficiently calibrated [Hellinga 1998].
- ➢ Data availability: Depending on the nature of the problem, the collection of the data that is required for calibration may be very difficult or sometimes simply infeasible. For example, the model involves many origins and destinations so that the determination of O-D flow becomes a daunting task. In addition, some simulation requires such a fine level of detail that it becomes very challenging to calibrate model parameters, e.g. the composition of road vehicles (including their static and dynamic properties) and driver population (including parameters that describe the perception and reaction process of each driver group). These problems can be even more challenging when the size of the network under analysis becomes large.
- ➢ Timing issue: A typical calibration process minimizes estimation errors (OD-matrix and/or parameters) and uses data obtained from a certain period of time in a typical day. However, this data may not represent a wide range of all likely demand conditions observed at a facility [Lee and Kaan 2008].

To address these issues, a number of calibration procedures are found [Jayakrishnan et al 2001], [Chu et al 2003], [Hourdakis et al 2003], [Turley 2007], [Balakrishna et al 2007], [Lee and Kaan 2008]. In addition, some of them have provided techniques that are able to deal with one or more of the above issues.

4. PROPOSED SOLUTION FOR CALIBRATION

The accuracy of the real world input data to the traffic simulation model should directly correlate to the accuracy and usefulness of the simulation output. As discussed before, these can be broadly categorized into three types of data and measured in real world situations in the following ways:

- Transportation network data: Acquired from satellites and maps. Used to provide roadway data and traffic control device data.
- Traffic flow data: Acquired from either video cameras or inductive loop detectors. Used to represent the amount of traffic flow in a given time interval.
- Driving population data: Acquired from sensors, demographic information, and data mining of other information to create underlying car-following, lane-changing, gap-acceptance, and route choice models and parameters within these models.

The output of the simulation model can be compared against the real world traffic flow inputs to calibrate the system properly. By automating simulation inputs from real world sensors, the amount of variables left to calibrate and the time taken to do so will be decreased. Therefore, the challenges mentioned before in model calibration can be handled as follows:

- ➢ Resource consumption: By measuring as much real world data as possible through sensors, human labor will be minimized and the need to collaborate between teams will be eliminated as collection of data will be automated.

- Calibration methodology: The amount of variables left to calibrate will be decreased by increasing the amount of real world data inputs and the accuracy of the initial simulation model will be increased. Therefore, the amount of time and resources to accurately calibrate a best fit simulation model to the real world data will be greatly lessened.
- Measures of performance: By having more inputs and less dynamic variables, the calibration time will be significantly reduced and the performance will be much more efficient.
- Data availability: The very nature of having various sensors collect data rather than humans will eliminate typographical error and greatly increase the amount of available data and the speed it is received at.
- Timing issue: By solving the data availability and performance problems above, by having more data and having it available quicker, more data will be ready sooner to use in all demand conditions.

Figure 1 depicts our system architecture for automatic calibration of transportation simulation models such as CORSIM (Corridor-microscopic simulation program). CORSIM, developed by University of Florida, is relatively simple and easy to operate without sacrificing the fidelity of modeling. In addition, CORSIM is perhaps the best (if not the only) microscopic simulator that can be embedded as a simulation engine to be invoked by other applications. The network model is automatically supplied by available GIS (Geographic Information System) data or extracted from satellite images that include roadway structures and traffic control devices.

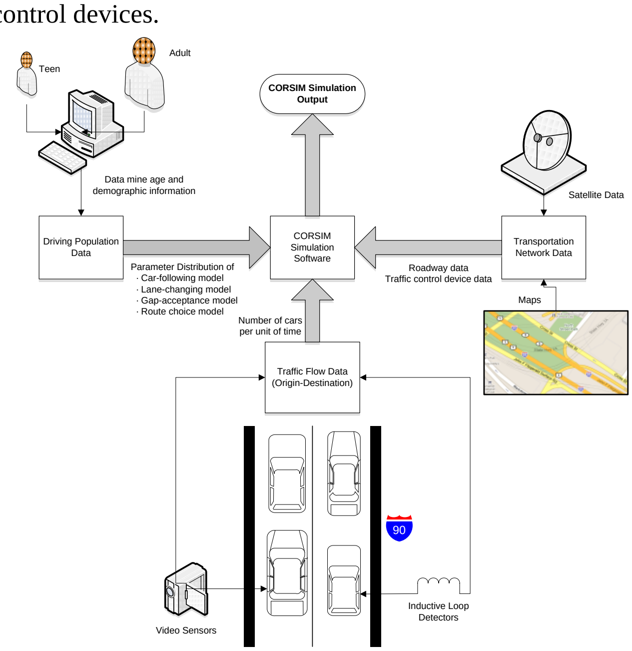

Figure 1. System Architecture

The following sections deal with issues and provide solutions for each of the components. Section 5 discusses the integration technology. The traffic flow model is extracted from data collected with a variety of sensor technologies and then filtered out noise for traffic pattern recognition using data mining methods. Section 6 presents seamless data collection while applying data mining to model calibration is discussed in Section 7. Apart from the current approaches of inferring Driving Population data from demographic statistics, we directly collect Driving Population data such as gap-acceptance from very high resolution satellite data or sensors with noise filtered out by data mining which are also discussed in Sections 5, 6, and 7. In Section 8, we present a prototype that focuses on the feasibility of traffic model calibration.

5. GEOGRAPHY INTEGRATION

Managing the sensing resources as well as interpreting the collected data requires a distributed infrastructure. A GIS based Multi-Agent Geo-Simulation architecture was proposed with the purpose of analyzing not only potential interactions between sensors but also with the real geographic environment where physical sensor are deployed [Mekni 2008]. At lower level, a GIS as the software platform is essential to reproduce real spatial data in the simulation environment. It actually serves as the spatial database, the functional background and the visual tool to supervise the geographic environment. The simulated environment is constructed from reliable GIS data, including raster/grid data (land cover, land use, elevation, etc) and vector data (transportation network).

ESRI ArcGIS supports a full structure of transportation data models using an object-oriented data modeling approach. Its transportation data model, UNETRANS (Unified Network for Transportation) developed by UC Santa Barbara provides best-practice templates for integrating static transportation network data and time-varying traffic flow data from a variety of sources, including characteristics of intersections, stop lines, signal timings, public transport stops, passenger flows, bus or transit lanes. In addition, ESRI ArcGIS Server 10 provides the ability to create, manage, and distribute GIS services over the Web to support desktop, mobile and Web mapping applications. It will assist in developing a large transportation infrastructure database with friendly Web access by decision makers, scientists, engineers and other users.

Aside from massive available GIS transportation data, remote sensing is a leading technology providing the update data at large spatial extent for many transportation applications, such as roadway delineation, traffic flow, road quality, and pollution assessment [Halem 2007]. The transportation application of remote sensing can be traced back to 80 years ago. In 1927, aerial photography was used to estimate traffic densities on a highway between Baltimore and Washington [Angel 2002]. Commercial satellite imaging has been available since late 1990s. In 1994, the U.S. government allowed civil commercial companies to market high spatial resolution remote sensor data. This resulted in the creation of a number of commercial consortia that have the capital necessary to create, launch, and market high spatial resolution digital remote sensor data. Since then, digital high spatial resolution remote sensing becomes increasingly available. The new sensors provide inexpensive and stable digital fine spatial resolution data from satellite platforms that enables better use of their applications which require a high level of detail, particularly transportation. The most notable companies are DigitalGlobe, Inc. and GeoEye, Inc. Multi-spectral imagery can be obtained at spatial resolutions 1.65 m and the panchromatic imagery can be obtained at spatial resolution at 0.41 m (GeoEye-1).

6. SEAMLESS DATA COLLECTION

The transportation traffic can be measured by various types of traffic sensors. Inductive loop detectors can be installed into the pavement and are the most widely used sensors. The disadvantage of these sensors is that they can be easily damaged and are hard to replace and maintain. The traffic detection using video cameras that are installed in roadway or on overhead structure [Cheng, et al. 2005] is an alternative approach. The accuracy of traffic estimation in this approach depends heavily on the image processing techniques for traffic flow detection. The traffic counts detected by video cameras may be inaccurate because the video sensing and content analysis could be affected by lighting changes, headlight reflection and rainy/snowy weather condition and so on. In addition, there are specific requirements such as installing video cameras at heights above 25 feet, which limits using them in traffic applications.

There is increasing interest in using non-intrusive sensors such as active/passive infrared, acoustic array sensors, and ultrasonic sensors in traffic monitoring applications. The costs of these sensors may be a problem. Microwave radar has gained the popularity due to its insensitiveness to weather condition and capability of multilane detection. However, microwave radar has limitations to detect low speed (less than 15 mph) moving objects such as vehicles, which especially degrades the accuracy of estimating traffic congestion.

In our prototype, we develop a system that extracts traffic data from wireless/mobile sensors, classifies it, and stores it in a well-defined format. The stored data can be easily retrieved and used as the standard input to a traffic simulator for automatic calibration. Our previous work has demonstrated how static sensors such as Automatic Traffic Recorder (ATR) stations collect traffic data [Ni and Leonard 2004], [Ni 2010].

There are several challenges to seamlessly collecting the traffic data in real-time from wireless/mobile sensors. The first challenge is the power limitation of wireless sensors. In many applications, replacing the power supply source or recharging is close to infeasible. One approach to overcome this obstacle is to use environmental energy as the power supply. Solar energy has the advantage of predictability over other forms of environmental energy. Topology design is another important challenge for energy-efficient data collection. The network topology should be designed to minimize the power consumption and increase the throughput, and guarantee a uniform load distribution.

A new data collection approach using a mobile sink, in which the burden of data gathering is shifted from the sensor nodes to the mobile sink, results in improvements in energy efficiency and real-time performance. The mobile sink based approach is suitable for seamless transportation data collection in some applications. Energy supplies in mobile sinks are easily replaceable. It can be periodically returned to a support center for recharging, removing resource constraints. The sink can come in various forms ranging from a man riding a vehicle to a robot programmed to visit the network. It travels according to some pre-determined or random schemes through the network and gathers data from sensor nodes. One drawback of using mobile sink is that the sensor nodes have to wait for the sink to visit the area before they can transmit their data. Infrequent visiting of a certain area can also result in increased data delivery delays. On the other hand, having a mobile sink traveling inside the network helps in collecting data from sensors located in physically remote or isolated regions.

In the traffic applications, the data collection patterns can be classified in three main categories.

- Periodic sensing: Sensors monitor a predefined variable and transmit the sensed data to the control sink continuously.
- Event-driven: Sensors send a message whenever they detect a pre-determined event.
- Query-based: Sensors transmit data packets after receiving a query from the control center.

7. APPLYING DATA MINING TO MODEL CALIBRATION

Traffic models are expected to be properly calibrated with actual traffic system behavior. However, prediction of current models may deviate significantly from the ground truth without thorough understanding of the factors responsible for this discrepancy [Ni and Leonard 2004], [Ni 2010], [Lattner 2010], [Wang 2010].

To deal with the challenges of resource consumption and calibration methodology described in Section 4, we use association pattern mining to automatically identify controlling factors of model parameters that are associated with discrepancy between the model and ground truth. This improves on the traditional method of manually adjusting model parameters to calibrate the model to comply with the real system to calculate the difference between predicted and actual traffic flows. Data mining will yield a map of change between the two instances—actual and modeled—of the system. The potential drivers of model discrepancies are the model parameters. We then find controlling patterns of model parameters that discriminate between traffic flows for which there is a large discrepancy between the data predicted by the model and the actual traffic data. We summarize the patterns in super patterns of model parameters that are associated with discrepancy using our super-patterns concept. The goal of this evaluation is to provide model designers with controlling factors (super patterns) of model parameters that are associated with wrong predictions on traffic data. We then evaluate several, progressively improved, versions of the model.

Super-patterns are agglomerated clusters of similar patterns [Stepinski 2010], [Ding 2009]. While each pattern represents precise but local nuggets of information, their clusters provide more generalized but global information. To calculate the super-patterns, we first calculate a distance matrix between each pair of patterns and then perform agglomerative clustering directly from the distance matrix. We use agglomerative clustering because it would aggregate objects without breaking the patterns. The end product of this synthesis tool is a small number of super-patterns, each describing a bundle of change motifs. Such knowledge gives an in-depth insight into combinations of factors that drive the change between a predictive model and actual traffic flows.

Furthermore, to address the challenges of data availability and timing issue discussed in Section 4, density based clustering is applied to traffic flow data to remove noise and outliers. Cars that do not belong to any aggregated clusters are viewed as incorrect counts and will be discarded. Association rule mining is applied on driving population data to identify the relationship of the age/gender group with driving behavior to generate more accurate and coherent model input data for better calibration.

8. A QUICK PROTOTYPE

This prototype focuses on the two main challenges in calibrating traffic models. The first challenge is how to abstract real-world traffic conditions for base-year modeling. The other is how to prevent typographical errors such as those in the origin-destination matrix. As mentioned in Section 4, our solution uses CORSIM, a microscopic traffic simulation software, as the core of our automatic system. CORSIM contains two sets of microscopic simulators to represent an entire traffic environment: NETSIM for urban street traffic and FRESIM for freeway traffic. For a proof of concept, we use NETSIM.

Our goal is to automatically generate the input file to CORSIM from real world sensor data to calibrate the simulation model so that it accurately models the real world. The CORSIM input file contains characteristics that vary over time such as traffic volumes, turn movements, traffic regulations, and signal timing, as well as characteristics that vary over space such as traffic geometry, and types of roads [CORSIM 2006]. Real world data comes from traffic flow data collected by Minnesota DOT [Minnesota DOT 2009].

8.1. CORSIM Input Methods

We will examine two methods of inputting data into CORSIM traffic simulator. One is the built-in GUI (Graphic User Interface) coming with CORSIM, and the other is facilitated by our automated prototype. Automated input minimizes the need for human interaction at traffic simulation, thereby reducing typographical errors involved in tedious manual work. More importantly, getting all inputs from real world sensors, the turnaround time from gathering information from sensors to gaining an output can be greatly decreased, allowing the system to be calibrated quicker and more accurately.

As shown in Figure 2, the CORSIM input GUI allows you to manually enter the traffic geometry, the amount of flow into each node, and all pertinent data to the system, one step at a time.

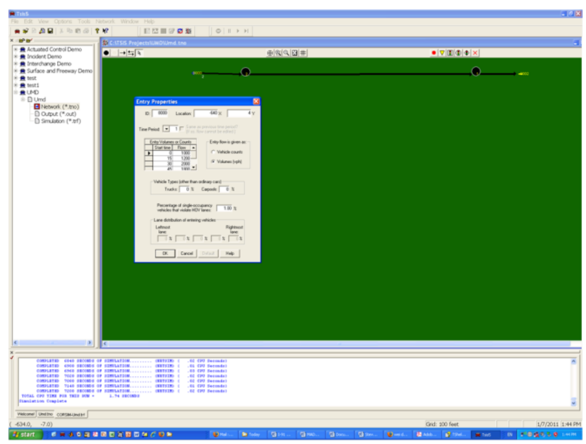

Figure 2. CORSIM Input GUI

Our prototype facilitates an automated GUI, entitled CORSIM Input Generator, which logically separates the input file to four sections.

(1) Header and run control data: This CORSIM input section contains run title, run identification, run control, time period specification, time interval, time steps per time period, and report information. Most of this data will be constant such as time intervals set to 12 two-hour intervals with measurements every hour. The remaining information of title, identification, date, and

description data will be inputted from the user as seen in Figure 3.

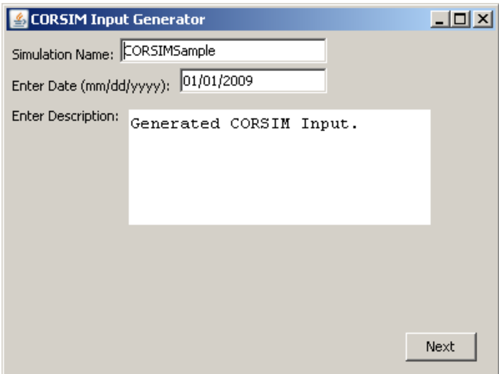

Figure 3. Header and Run Control Data Input

(2) Transportation network data: Obtained from satellite and map data, will provide the traffic geometry, traffic link, and signal control data. The input will be put into the top input file in Figure 4 to generate the geometry of the simulation network.

(3) Dynamic parameter distribution data: Obtained from car-following, lane-changing, gap-acceptance, and route choice models. These dynamic variables will be used in the model to simulate driver behavior by inputting them into the bottom file in Figure 4.

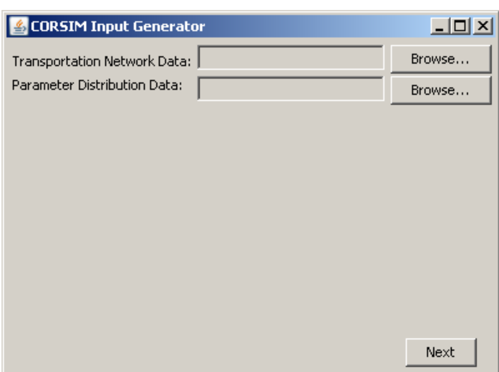

Figure 4. Transportation Network and Parameter Distribution Data Input

(4) Traffic flow data: Obtained from real world sensors such as ATR to input the amount of vehicles per hour entering a given node in the traffic network from outside the system. Entry and internal nodes are specific to the traffic geometry generated after the transportation network and parameter distribution data is entered. Figure 5 shows the GUI to select an entry node, internal node, and traffic data to generate the traffic flow input data to CORSIM. One or more files may be added to calculate the traffic flow for each entry node into the network.

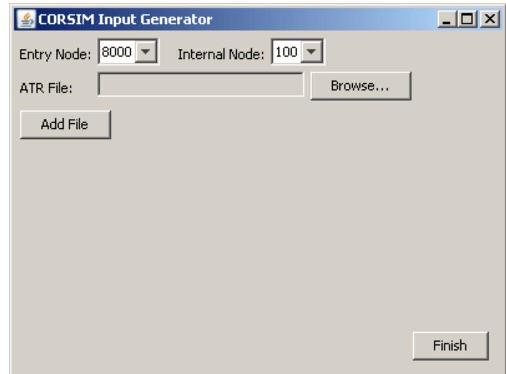

Figure 5. Traffic Flow Data Input

8.2. CORSIM Output Methods

The CORSIM software provides two outputs: graphical and text-based. The graphical output shows the geometry of the network and a visual representation of the traffic flow. For our automation purposes, the text output contains statistics and details that will be of greater use.

CORSIM's graphical output shows the traffic geometry in the form of roads. The traffic flow is shown by cars driving, with colors representing different direction, and incidents shown by yellow and red. In addition, the time of the simulation as well as the frame delay and frame/time step can be modified to change the simulation output. Figure 6 is a screenshot of CORSIM's graphical output.

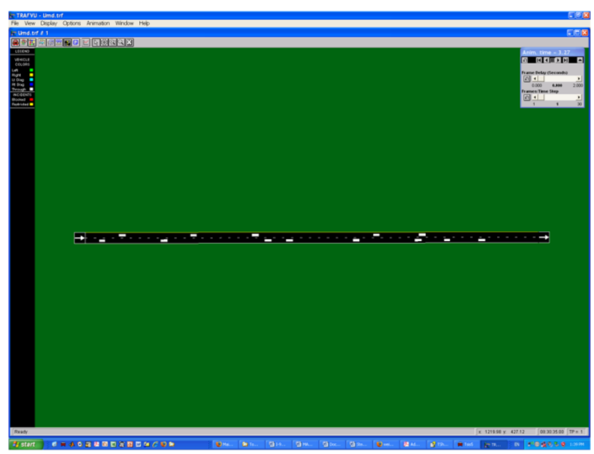

Figure 6. CORSIM Output GUI

CORSIM text-based output is one text file with a ".out" extension that echoes the input to the simulator and shows the resulting statistics from it. NETSIM simulation contains an output section that is time period specific.

The CORSIM text based output includes four main sections:

(1) Input data echo: Consists of a copy of the input file and tables stating the complete specification of the traffic environment and run options, including all the user-supplied inputs and default values. This can be used for

checking the validity and acceptability of values and parameters.

(2) Initialization results: Show how the vehicles filled the network at different time intervals prior to the network reaching equilibrium. The initialization results are not included in the cumulative results. CORSIM starts accumulating statistics after equilibrium has been reached.

(3) Intermediate results: Optionally printed out at the end of user specified intervals.

(4) End of time period results: Similar to intermediate results, but printed out at the end of time periods rather than at the end of option user specified intervals.

Intermediate and end of time period results include measures of effectiveness (MOE) that are used to evaluate traffic systems [CORSIM 2006].

8.3. A Running Example

Our simple running example consists of two nodes, an entry node, and an exit node. The system is only concerned with the entry traffic to the system, so we will input the traffic flow for that initial node. The traffic geometry contains one stretch of 1012 foot road with a 0% grade and a free flow speed of 30 mph.

8.3.1 Real World Traffic Flow

Traffic flow is obtained from the Minnesota DOT for January 1, 2009 at Station 101, North of Garfield Avenue in Duluth, St. Louis County; heading north.

Table 1 shows a small segment of the data. Each day has an entry for all 24 hours and a total column at the end [Minnesota DOT 2009].

Table 1. Sample Sensed Traffic Flow

DATE	DAY	1:00 AM	2:00 AM	3:00 AM	4:00 AM	5:00 AM
1-Jan-09	Thursday	220	211	216	248	211
2-Jan-09	Friday	92	74	53	49	69
3-Jan-09	Saturday	151	117	124	60	59
4-Jan-09	Sunday	79	75	79	29	41
5-Jan-09	Monday	45	68	50	39	110
6-Jan-09	Tuesday	88	94	52	43	90
7-Jan-09	Wednesday	89	88	54	60	99
8-Jan-09	Thursday	120	164	68	55	87

8.3.2 CORSIM Input File

The basic CORSIM input is a text file consisting of 80 one character columns. The 3 columns to the right (78-80) contain the record type number, which specifies what type of information the row is displaying to the CORSIM software.

Figure 7 displays the input file, corresponding to the sample Minnesota DOT traffic data, generated by our CORSIM Input Generator. It constructs the traffic geometry, traffic parameters, and traffic flow in a way that the CORSIM simulator recognizes. This sample of the input file contains the traffic for the first two hours. The full input would replicate the 50 and 53 record types (flow rate at the entry node in an interval and its variation) for all 12 two hour intervals.

```
Created by CORSIM Simulator Fri Jan 07 13:38:00 2011                           0
12345678 1 2345678 2 2345678 3 2345678 4 2345678 5 2345678 6 2345678 7 234567
CORSIM SIM                              1   72011UMD                        0  1
       1   0       3     7981 0000  0              3 000        7781    7581   2
720072007200720072007200720072007200720072007200                               3
                 120                                                           4
   0       0   0   0   0   0   0   0   0   0   0                               5
   2   11012       2     0100          8002               20  18  30   0      11
8000   2           2     0100             1               20  18       0      11
   2   1     100                                                              21
8000   2     100                                                              21
   1       2                                                                  35
   2    8000                                                                  35
   1 1                                                                        36
   2 1                                                                        36
Time Interval: 0000-0200
8000   2 220   0   0  100                                                     50
8000   2 220   0 211  60                                                    1 53
   0                                                                         170
   1    1266     26                                                          195
   2     254     24                                                          195
   0   3                                                                     210
   1   0   0                                                                 210
```

Figure 7. Sample Generated CORSIM Input File

9. CONCLUSION AND FUTURE WORK

Our study demonstrates that it is feasible to automate model calibration of transportation simulation. We have successfully implemented a quick prototype, called *CORSIM Input Generator*, which downloads traffic flow data from the Automatic Traffic Recorder (ATR) by Minnesota DOT and converts the data into the CORSIM Input File, ready for simulation. A running example on the prototype supports the fidelity of the proposed system architecture. Our system integrates the modeling components: network simulator, traffic simulator, and driver simulator for an environment of transportation simulation. It automates the process of data collection and distillation as part of the model calibration by utilizing a distributed infrastructure of various wireless sensors integrated with the real geographic environment and by applying association pattern mining methodology to identify controlling factors of model parameters for closing the gap between the model and the ground truth. Though it is a long way towards full automation since model calibration currently lacks standards to assess the accuracy of modeling, our work sheds light on its feasibility.

Automatic calibration of transportation simulation models promises a fertile field for multidisciplinary research. However, many technical details need to be worked out. Innovative sensors beyond inductive loop detectors and video cameras should be investigated for data collection at desired scopes with manageable resource consumption and prompt response timing. Geography integration of the real world environment with the computerized representation should make all transportation data, including network, traffic, and driver, available all the time everywhere and in ready-for-use format. Once data is delivered, it all boils down to knowledge discovery in particular domains with endless endures.

Acknowledgment

One of the authors, Hong Liu, thanks Marguerite Zarrillo at the Physics Department, University of Massachusetts Dartmouth, for her message on the problem in using traffic simulation software.

References

[Angel 2002] Angel, A. 2002. "Methods Of Traffic Data Collection Using Aerial Video." The IEEE 5th International Conference on Intelligent Transportation Systems.

[Balakrishna et al 2007] Ramachandran Balakrishna, Constantinos Antoniou, Moshe Ben-Akiva, Haris N. Koutsopoulos, and Yang We. 2007. "Calibration of Microscopic Traffic Simulation Models: Methods and Application". Transportation Research Record: 198-207.

[Cheng, 2005] Cheng. 2005. "Development and Field Test of a Laser-Based Non Intrusive Detection System for Identification of Vehicles on the Highway", IEEE Transactions on Intelligent Transportation Systems, Vol. 6, No. 2, (June): 147-155.

[Chu et al 2003] Lianyu Chu, H. X. Liu, Jun-Seok Oh, and W. Recker. 2003. "A Calibration Procedure for Microscopic Traffic Simulation." Proceedings of IEEE Intelligent Transportation Systems. Vol.2: 1574–1579.

[CORSIM 2006] ITT Industries, Inc. 2006. "CORSIM User's Guide." www.ptr.poli.usp.br/lemt/CORSIM/CORSIMUsersGuide.pdf (accessed January 2011).

[Daganzo 1994] Daganzo, C. F. 1994. "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory." Transportation Research B 28(4): 269-287.

[Daganzo 1995] Daganzo, C. F. 1995. "The Cell Transmission Mode, Part II: Network Traffic." Transportation Research B 29(2): 79-93.

[Daganzo 1995b] Daganzo, C. F. 1995b. "Requiem For Second-order Fluid Approximations of Traffic Flow." Transportation Research B 29(4): 277-286.

[Ding 2009] Ding, W.; T. F. Stepinski; and J. Salazar. 2009. "Discovery of geospatial discriminating patterns from remote sensing datasets." Proceedings of SIAM International Conference on Data Mining (SDM), (Nevada, April).

[Gazis et al. 1961] Gazis, D. C., R. Herman, et al. 1961. "Non-Linear Follow the Leader Models of Traffic Flow." Operations Research 9: 545-567.

[Gipps 1981] Gipps, P. 1981. "A behavioral car following model for computer simulation." Transportation Research, Part B 15: 105-111.

[Halem 2007] Halem, M.; A. Patwardhan; S. Dornbush; M. Seablom; and Y. Yesha. 2007. "Sensor web design studies for realtime dynamic congestion pricing." Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops '07). (White Plains, NY, March 19-23): 413-418.

[Hellinga 1998] Hellinga, B.R. 1998. "Requirements for the Calibration of Traffic Simulation Models". Canadian Society of Civil Engineering [CSCE], University of Waterloo, Ontario, Canada.

[Hewitt 1992] Hewitt, R. H. 1992. Using Probit Analysis with Gap Acceptance Data, Department of Civil Enginneering, University of Glasgow.

[Hidas and Behbahanizadeh 1999] Hidas, P. and K. Behbahanizadeh. 1999. Microscopic simulation of lane changing under incident conditions. Proceedings of the 14th International Symposium on the Theory of Traffic Flow and Transportation.

[Hourdakis et al 2003] John Hourdakis, Panos G. Michalopoulos, Jiji Kottommannil. 2003. "A Practical Procedure for Calibrating Microscopic Traffic Simulation Models". Pre-print CD-ROM, the 82th TRB Annual Meeting, Transportation Research Board, (Washington, D.C).

[Jayakrishnan et al 2001] R. Jayakrishnan, Jun-Seok Oh, Abd-El-Kader Sahraoui. 2001. "Calibration and Path Dynamics Issues in Microscopic Simulation for Advanced Traffic Management and Information Systems". Transportation Research Record 1771: 9-17.

[Kosonen 1999] Kosonen, I. 1999. HUTSIM - Urban Traffic Simulation and Control Model: Principles and Applications. Transportation Engineering, Helsinki University of Technology, . Ph.D. Thesis.

[LANL 1999] LANL. 1999. Transportation Analysis Simulation System (TRANSIMS), Los Alamos National Laboratory.

[Lattner 2010] Lattner, A.D.; T. Bogon; Y. Lorion; and I.J. Timm. 2010. "A knowledge-based approach to automated simulation model adaptation." Proceedings of the 2010 SCS Spring Simulation Multiconference: 43th Annual Simulation Symposium (SpringSim'10-ANSS'11). (Orlando, Florida, April 11-15): 200-207.

[Lee and Kaan 2008] Lee, Jung-Beom and Ozbay, Kaan. 2008. "Calibration of a Macroscopic Traffic Simulation Model Using Enhanced Simultaneous Perturbation Stochastic Approximation Methodology". Pre-print CD-ROM, the 87th TRB Annual Meeting, Transportation Research Board, Washington, D.C., [Paper #08-2964]

[Lighthill and Whitham 1955] Lighthill, M. and G. Whitham. 1955. "On kinematic waves II. A theory of traffic flow on long crowded roads." Proc. Royal Society of London, Part A 229(1178): 317-345.

[May et al 1991] May, A. D. and e. al. 1991. "Integrated system of freeway corridor simulation models." Transportation Research Record 1320: 177-189.

[Mekni 2008] Mekni, M. and B. Moulin. 2008. "Using multi-agent geo-simulation techniques for intelligent sensor web management." Proceedings of the Fifth IEEE/ACM International Conference on Soft Computing as Transdisciplinary Science and Technology (CSTST'08), Cergy-Pontoise/Paris, France.

[Michalopoulos 1984] Michalopoulos, P. G. 1984. "Dynamic freeway simulation program for personal computers." Transportation Research Record 971: 68-79.

[Minnesota DOT 2009] Automatic Traffic Recorder. 2009. "Minnesota Department of Transportation Continuous Count Station Edited Data for 2009." www.dot.state.mn.us/traffic/data/atr/atr.html (accessed January 2011).

[Newell 1993] Newell, G. F. 1993. "A simplified theory on kinematic waves in highway traffic, Part II: queueing at freeway bottlenecks." Transportation Research B 27(4): 289-303.

[Ni 2010] Ni, D. 2010. "A spectrum of traffic flow modeling at multiple scales." Proceedings of the Winter Simulation Conference (WSC). (Baltimore, MD, December 3-6).

[Ni and Leonard 2004] Ni, D. and J.D. Leonard. 2004. "Development of TrafficXML: the common vocabulary for traffic simulation." TRB Transportation Research Record: Journal of the Transportation Research Board, No. 1879: 30-40.

[Ni et al. 2004] Ni, D.; J.D. Leonard II; A. Guin; and B.M. Williams. 2004. "Systematic approach for validating traffic simulation models." TRB Transportation Research Record: Journal of the Transportation Research Board, No. 1876: 20-31.

[Payne 1979] Payne, H. J. 1979. "FREFLO: A macroscopic simulation model for freeway traffic." Transportation Research Record 722: 68-77.

[Phillips 1979] Phillips, W. F. 1979. "A Kinetic Model for Traffic Flow with Continuum Implications." Transportation Planning and Technology 5: 131-138.

[Pipes 1953] Pipes, L. A. 1953. "An Operational Analysis of Traffic Dynamics." Journal of Applied Physics 24: 271-281.

[Prigogine 1961] Prigogine, I. 1961. *A Boltzmann-Like Approach to the Statistical Theory of Traffic Flow*. Elsevier, Amsterdam.

[Richards 1956] Richards, P. I. 1956. "Shock waves on the highway." Operations Research 4: 42-51.

[Rossi and Meneguzzer 2002] Rossi, R. and C. Meneguzzer. 2002. "The Effect of Crisp Variables on Fuzzy Models of Gap-Acceptance Behaviour." Proceedings of the 13th Mini-EURO Conference:

Handling Uncertainty in the Analysis of Traffic and Transportation Systems.
[Stepinski 2010] Stepinski, T.; J. Salazar; W. Ding; and D. White. 2010. "ESTATE: Strategy for Exploring Labeled Spatial Datasets Using Association Analysis." Proceedings of the 13th International Conference on Discovery Science (DS10), (Canberra, Australia, October).
[Sukthankar 1997] Sukthankar, R. 1997. Situation Awareness for Tactical Driving. Robotics Institute. Pittsburgh, PA 15213, Carnegie Mellon University. Ph.D.
[Treiber and Helbing 2002] Treiber, M. and D. Helbing. 2002. Realistische Mikrosimulation von Straßenverkehr mit einem einfachen Modell. 16th Symposium Simulationstechnik ASIM, Germany.
[Turley 2007] Carole Turley. 2007. "Calibration Procedure for a Microscopic Traffic Simulation Model". MS thesis. Department of Civil and Environmental Engineering, Brigham Young University.
[Velan and Van Aerde 1996] Velan, S. M. and M. Van Aerde. 1996. "Gap Acceptance and Approach Capacity at Unsignalized Intersections." ITE Journal 66(3): 40-45.
[Wang 2010] Wang, H. and S. Wu. 2010. "Intelligent architecture through a supervised learning approach in wireless multimedia sensor networks." Proceedings of the 2010 SCS Spring Simulation Multiconference: 43th Annual Simulation Symposium (SpringSim'10-ANSS'11). (Orlando, Florida, April 11-15).
[Whitham 1974] Whitham, G. B. 1974. *Linear and Nonlinear Waves*. John Wileyand Sons Inc, New York, NY.
[Wiedemann 1974] Wiedemann, R. 1974. Simulation des Straßenverkehrsflusses, In Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Germany,.
[Wisconsin DOT 2010] Wisconsin DOT Microsimulation Model Calibration Guideline. 2010. "Model calibration: from traffic analysis and microsimulation." http://www.wisdot.info/microsimulation/index.php?title=Model_Calibration (accessed December 2010).
[Yager 1975] Yager, S. 1975. "CORQ - A Model for Predicting Flows and Queues in a Road Corridor." Transportation Research Record 533: 77-87.

Biography

Hong Liu is a Professor of Electrical and Computer Engineering at University of Massachusetts Dartmouth, Dartmouth, MA, U.S.A. She received her Ph.D. from Polytechnic Institute of New York University in 1990. Her research interests include computer networks, quality of service and security/privacy, applications with wireless and Internet technology.

Qian Yu is an Assistant Professor in the department of Geosciences at University of Massachusetts, Amherst, MA, U.S.A. since 2006. She got her Ph.D. in Environmental Sciences at UC Berkeley in 2005. Her research focuses on high resolution and hyperspectral remote sensing to extract land and water features. She published in both remote sensing/GIS journals and applied science journals of high quality. Her work is funded by NSF, USGS and UMass Faculty Research Fund.

Wei Ding has been an Assistant Professor of Computer Science at the University of Massachusetts Boston, Boston, MA, U.S.A. since 2008. She received her Ph.D. degree in Computer Science from the University of Houston in 2008. Her main research interests include Data Mining, Machine Learning, Artificial Intelligence, Computational Semantics, and with applications to astronomy, geosciences, and environmental sciences. Her research projects are currently sponsored by NASA, DOE, and NSF.

Daiheng Ni is an Assistant Professor of Civil Engineering at University of Massachusetts, Amherst, MA, U.S.A. Prior to that, he was a post-doctoral fellow at and received his Ph.D. from Georgia Institute of Technology, Atlanta, GA, U.S.A. His research interests include traffic flow theory and simulation, intelligent transportation systems, traffic sensing and information technology, and transportation logistics and optimization.

Honggang Wang received his Ph.D. degree in computer engineering from the University of Nebraska-Lincoln in 2009. He is currently an assistant professor at University of Massachusetts Dartmouth, Dartmouth, MA, U.S.A. His research interests include networking, wireless sensor networks, multimedia communication, network and information security, embedded sensory systems, biomedical computing, and pattern recognition. He won the Best Paper Award at the 2008 IEEE Wireless Communications and Networking Conference (WCNC). He serves as an Associate Editor of Wiley's *Security and Communication Networks Journal*.

Stephen Shannon is a Five-Year B.S./M.S. student in Computer Engineering at the University of Massachusetts Dartmouth, Dartmouth, MA, U.S.A. He is interested in computer network research with a focus on security and privacy.