중심합성계획법 기반 이중유전자알고리즘을 활용한 차량추종모형 정산방법론 개발

Calibration of Car-Following Models Using a Dual Genetic Algorithm with Central Composite Design

배 범 준* · 임 현 섭** · 소 재 현***

 * 주저자 : 한국교통연구원 민자도로관리연구센터 부연구위원
 ** 공저자 : Oak Ridge National Laboratory, Center for Transportation Analysis, R&D STAFF
*** 교신저자 : 한국교통연구원 스마트시티교통연구팀 부연구위원

Bumjoon Bae* · Hyeonsup Lim** · Jaehyun (Jason) So***

 * Center for Private Financing Highway, The Korea Transport Insititue
 ** Center for Transportation Analysis, Oak Ridge National Laboratory
*** Division for Smart City and Transportept, The Korea Transport Insititue
 † Corresponding author : Jaehyun (Jason) So, jso@koti.re.kr

요 약

　미시적 교통류 모형의 정산은 시뮬레이션 분석에 있어 매우 중요한 요소이다. 유전자 알고리즘은 교통류 모형의 정산에 널리 활용되어 왔으며, 일반적으로 이러한 최적화 문제에 있어 높은 효율성을 보이는 것으로 알려져 있다. 하지만 제한된 시간내에 신속한 의사결정을 위한 시뮬레이션 분석에 있어 유전자알고리즘의 모형 정산속도는 여전히 느리다. 이에 본 연구에서는 정산 효율 향상을 위해 중심합성계획법 기반의 이중유전자알고리즘을 활용한 차량추종모형 정산방법론을 개발하였다. 개발된 정산 방법론에서는 실험계획법 중 하나인 중심합성계획법과 유전자알고리즘을 결합하여 준최적해를 찾고, 이를 다시 유전자알고리즘의 초기 값으로 하여 모형 파라미터의 최적해를 찾는다. 개발된 방법을 활용하여 Gipps의 차량추종모형을 정산하였다. 선행연구에서 사용된 단일 유전자알고리즘을 활용한 방법과 비교한 결과, 본 연구에서 개발한 방법이 더 짧은 시간내에 최적해를 찾는 것으로 확인되었다. 개발된 방법론은 유전자알고리즘을 사용하는 다양한 교통분석에 활용될 수 있을 것으로 기대된다.

핵심어 : 미시교통모형 정산, 최적화, 중심계획합성법, 유전자알고리즘, Gipps 차량추종모형

ABSTRACT

　The calibration of microscopic traffic simulation models has received much attention in the simulation field. Although no standard has been established for it, a genetic algorithm (GA) has been widely employed in recent literature because of its high efficiency to find solutions in such optimization problems. However, the performance still falls short in simulation analyses to support fast decision making. This paper proposes a new calibration procedure using a dual GA and central composite design (CCD) in order to improve the efficiency. The calibration exercise goes through three major sequential steps: (1) experimental design using CCD for a quadratic response surface model (RSM) estimation, (2) 1st GA procedure using the RSM with CCD to find a near-optimal initial population for a next step, and (3) 2nd GA procedure to find a final solution. The proposed method was applied in calibrating the Gipps car-following model with respect to maximizing the

likelihood of a spacing distribution between a lead and following vehicle. In order to evaluate the performance of the proposed method, a conventional calibration approach using a single GA was compared under both simulated and real vehicle trajectory data. It was found that the proposed approach enhances the optimization speed by starting to search from an initial population that is closer to the optimum than that of the other approach. This result implies the proposed approach has benefits for a large-scale traffic network simulation analysis. This method can be extended to other optimization tasks using GA in transportation studies.

Key words : Simulation calibration, Optimization, Central composite design, Genetic algorithm, Gipps car-following model

Ⅰ. Introduction

Traffic analyses, such as traffic signal optimization and predicting network-wise travel time, cannot be solved by using simple regression models. In addition, field experiments and implementations are often very limited due to safety issues and tremendous costs. Microscopic traffic simulation has been used to describe traffic flow characteristics and solve those issues based on pre-defined driving behavior models embedded in such tools (e.g. a car-following, lane-changing, gap-acceptance models, and so on). They can represent realistic traffic situations including normal and abnormal traffic conditions if their parameters are properly calibrated. Hence, calibration of the models is a key procedure to obtain realistic and reliable results from a microscopic traffic simulation analysis.

A variety of calibration approaches have developed using parametric/nonparametric statistical models or machine learning algorithms(Kim et al., 2005; Eriksson, 1999; Kesting and Treiber, 2008; Van Hinsbergen, et al., 2009; Vasconselos, 2014; Treiber and Kesting, 2013; Park and Qi, 2005; Park et al., 2006; Cunto and Saccomanno, 2008) while there has been no standard. A genetic algorithm (GA) is a search algorithm that emulates biological evolutionary theories. A GA searches solutions at multiple locations in a parameter space based on three probabilistic rules including selection, crossover, and mutation over generations (or iterations). Thus, the main advantage of GA is that it has a low likelihood of getting stuck in local optima while providing high computational efficiency(Kim et al., 2005; Howe and Bozdogan, 2010). This tool has become useful for finding near-optimal values of parameters embedded in a traffic simulation model(Kesting and Treiber, 2008; Vasconselos, 2014; Treiber and Kesting, 2013; Park and Qi, 2005; Park et al., 2006; Cunto and Saccomanno, 2008; Cheu et al., 1998). Although a GA has provided reliable solutions for calibration, improvement of its computational efficiency in traffic simulation study has less been paid attention. Traffic simulation studies generally require to calibrate a large number of model parameters and each simulation run takes longer computation time than deriving an analytic solution. That is, a large number of parameters with each parameter having multi-level values leads an excessive amount of calibration time (e.g., 5 parameters with 5 values per each makes 3,125 (=55) possible combinations).

This paper proposes a hybrid calibration procedure in which response surface methodology (RSM) with central composite design (CCD) is employed with GA to improve efficiency of the calibration process. Using field-collected vehicle trajectory data, the procedure is applied to calibrate the parameters of Gipps car-following model in order to fit spacing distribution between lead and following vehicles.

Ⅱ. Background on Methodology

1. Gipps Car-Following Model

Car-following models are one the most important models embedded in microscopic simulation tools. Most of them can be classified in a collision avoidance class aiming to specify a safe following distance behind the lead vehicle. Gipps car-following model is one of the most commonly used models in the class and mostly known for being a key factor of the Aimsun microscopic simulation tool. It consists of two components(Gipps, 1981): acceleration and deceleration models, corresponding to the empirical formulations defined by equations (1) and (2). These models calculate the speed of each vehicle at a given time t on the basis of its speed at the previous time step.

$$v_n^{acc}(t+\tau) = v_n(t) + 2.5a_n\tau(1-(v_n(t))/(v_n^d))\sqrt{(0.025+(v_n(t))/(v_n^d))} \tag{1}$$

$$v_n^{dec}(t+\tau) = -\tau d_n^2 + \sqrt{(\tau d_n^{2+d_n}2[x_{(n-1)}(t)-x_n(t)-S_{(n-1)}]-\tau v_n(t)+(v_{(n-1)}(t)^{2)/(d}_{(n-1)'}))} \tag{2}$$

where, τ :   driver's reaction time

            $n, n-1$:   follower and leader, respectively

            x :   space traveled by vehicle

            $v(t)$:   vehicle's speed at time t

            v_n^d, a_n :   follower's desired speed and maximum acceleration, respectively

            d_n, d_{n-1}' :   most severe braking that follower wishes to undertake and his estimate of leader's most severe braking capability

            $S_{(n-1)}$:   leader's effective length, which is the leading vehicle length plus the follower's desired inter-vehicle spacing at a stop

The speed of vehicle n at time t + τ is given by the minimum of Equation (1) and (2)

$$v_n(t+\tau) = \min \quad (\ [\![v_n^{acc}(t+\tau), v]\!]_n^{dec}(t+\tau)) \tag{3}$$

Then, the position of the vehicle n inside the current lane is updated taking this speed into the movement equation

$$x_n(t+\tau) = x_n(t) + v_n(t+\tau)\tau \tag{4}$$

Gipps model includes five parameters in order to compute acceleration and deceleration of a following vehicle. <Table 1> shows the model parameters to be calibrated.

<Table 1> Parameters of Gipps Car-Following Model

Parameters	Description
v_n^d (X1*)	Desired speed (m/s)
a_n (X2)	Maximum acceleration (m/s2)
d_n (X3)	Most severe braking that the follower wishes to undertake (m/s2)
$\dot{d}_{(n-1)}$ (X4)	Follower's estimate of the leader's most severe braking capability (m/s2)
$S_{(n-1)}$ (X5)	The leader's real length + the follower's desired inter-vehicle spacing (m)

* X1 – X5 are the parameter names to be used in the case study of this study.

2. Kernel Density Estimation and Genetic Algorithm for Model Calibration

A kernel density estimation (KDE) is a nonparametric method to estimate the probability density of a traffic flow variable, such as density and speed, without assuming a certain distribution of the data. This KDE is used to calculate probability densities of data points obtained by the simulation in this study.

Since the task of simulation in this study assumes that no disaggregated level attribute is available, the simulation model should be calibrated in a way to represent similar distribution of the aggregated attributes that can be observed in collected data. In this sense, the objective of this optimization task is to maximize the likelihood of having simulation results same as values in observed data. For each time point, a probability density of having a predicted value from the simulation using the KDE from the observations can be calculated. Then, the likelihood here is defined as the joint probability that the all predicted data points of the attributes given a certain parameter estimates vector given by Equation (5).

$$Maximize \quad L(\vec{\beta}) = prob(\widehat{\vec{y_1}}(\vec{\beta}) = (\vec{y_1}, \cdots, \widehat{\vec{y_n}}(\vec{\beta}) = \vec{y_n}) \tag{5}$$

Then, log transformation is used to make the calculation easier shown by Equation (6).

$$Maximize \quad \log(L(\vec{\beta})) = \log(prob(\widehat{\vec{y_1}}(\vec{\beta}) = \vec{y_1}) + \cdots, + \log(prob(\widehat{\vec{y_n}}(\vec{\beta}) = \vec{y_n}))) \tag{6}$$

A genetic algorithm (GA) is a heuristic searching technique for both constrained and unconstrained optimization problems. It generates solutions by mimicking the natural selection process. This algorithm continuously evolves a population of candidate solutions toward an optimum over generations by three main methods: selection, crossover, and mutation. In each generation a new solution is generated from its "parents" in a previous generation. The GA has been used widely in calibration of simulation models, to solve the problem using an automated procedure. In this study, a GA solver of the optimization toolbox in Matlab was used to find the best solution for the Gipps model.

3. Response Surface Methodology with Central Composite Design

In general a first-order model from a full- or fractional-factorial design is less useful for searching an optimum since it can only identify linear relationships between explanatory variables and a response variable. Box and
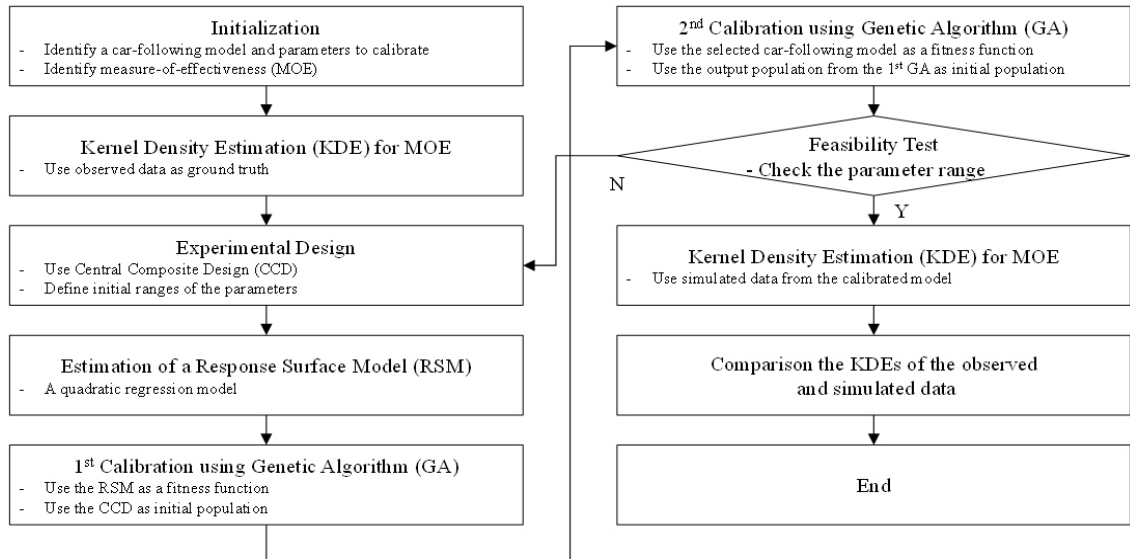
Wilson(1951) devised the response surface methodology (RSM) using a second-order model to investigate or optimize processes sequentially. The second-order model, expressed by Equation (7), contains linear interactions of two-factors and pure quadratic terms, as well as an intercept and linear main effect terms in a first-order model.

$$y = \beta_0 + \sum_{(i=1)}^{k} \beta_i x_i + \sum_{(i=1)}^{k} \beta_{i\,.\,i} x_i^2 + \sum_{(i=1)}^{k} \beta_{i\,.\,j} x_i x_j + \epsilon \tag{7}$$

A central composite design (CCD) is the most frequently utilized second-order design for RSM. This design consists of three types of experimental runs: a two-factor factorial design, center points and axial points. A fractional factorial is generally used in CCD for five or more factors(Mee, 2009). From the center point runs, the pure quadratic effect can be captured.

Ⅲ. Proposed calibration procedure

The overall procedure of the proposed calibration framework is shown in <Fig. 1>. The major characteristics of this approach is to use CCD for estimating a quadratic response model and to employ dual GA on the basis of the response modeling result.

<Fig. 1> Overview of proposed calibration procedure

1. Central Composite Design for Response Surface Model Estimation

First of all, CCD was used to build a second-order model for the response variable. The 5 parameters (v^d, a, d, $\acute{d}$, s) were considered as factors in <Table 1>, and half fraction of full-factorial was used. This 5-factor CCD is a 36-run design, including 16 of fractional factorial, 10 axial points, and 10 center points. <Table 2> shows both

coded and natural units of the design. The ranges of each factor in the design was selected on the basis of authors' experience with the car-following model. These values define the initial searching space for the proposed calibration framework. For each experiment run, the Gipps car-following model returns the negative sum of log-likelihood for spacing distribution and it is presented in the column Y as a response variable in <Table 2>.

<Table 2> Central Composite Design

Run	Coded Unit					Natural Unit					
	X1	X2	X3	X4	X5	X1	X2	X3	X4	X5	Y
1	-1	-1	-1	-1	1	15.0	0.8	2.0	2.0	12.0	5,111.5
2	-1	-1	-1	1	-1	15.0	0.8	2.0	3.0	8.0	5,215.7
3	-1	-1	1	-1	-1	15.0	0.8	3.0	2.0	8.0	5,087.8
4	-1	-1	1	1	1	15.0	0.8	3.0	3.0	12.0	5,285.5
5	-1	1	-1	-1	-1	15.0	1.2	2.0	2.0	8.0	4,987.3
6	-1	1	-1	1	1	15.0	1.2	2.0	3.0	12.0	5,158.6
7	-1	1	1	-1	1	15.0	1.2	3.0	2.0	12.0	5,028.0
8	-1	1	1	1	-1	15.0	1.2	3.0	3.0	8.0	5,155.9
9	1	-1	-1	-1	-1	25.0	0.8	2.0	2.0	8.0	3,528.6
10	1	-1	-1	1	1	25.0	0.8	2.0	3.0	12.0	3,449.9
11	1	-1	1	-1	1	25.0	0.8	3.0	2.0	12.0	3,338.0
12	1	-1	1	1	-1	25.0	0.8	3.0	3.0	8.0	3,473.8
13	1	1	-1	-1	1	25.0	1.2	2.0	2.0	12.0	3,404.3
14	1	1	-1	1	-1	25.0	1.2	2.0	3.0	8.0	3,365.1
15	1	1	1	-1	-1	25.0	1.2	3.0	2.0	8.0	3,367.9
16	1	1	1	1	1	25.0	1.2	3.0	3.0	12.0	3,514.3
17	-2	0	0	0	0	10.0	1.0	2.5	2.5	10.0	6,016.7
18	2	0	0	0	0	30.0	1.0	2.5	2.5	10.0	3,302.9
19	0	-2	0	0	0	20.0	0.6	2.5	2.5	10.0	3,304.3
20	0	2	0	0	0	20.0	1.4	2.5	2.5	10.0	3,290.6
21	0	0	-2	0	0	20.0	1.0	1.5	2.5	10.0	3,289.1
22	0	0	2	0	0	20.0	1.0	3.5	2.5	10.0	3,339.7
23	0	0	0	-2	0	20.0	1.0	2.5	1.5	10.0	3,613.0
24	0	0	0	2	0	20.0	1.0	2.5	3.5	10.0	3,455.0
25	0	0	0	0	-2	20.0	1.0	2.5	2.5	6.0	3,263.2
26	0	0	0	0	2	20.0	1.0	2.5	2.5	14.0	3,282.2
27	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,244.3
28	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,238.1
29	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,231.9
30	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,225.8
31	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,219.8
32	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,213.8
33	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,207.9
34	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,202.0
35	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,196.2
36	0	0	0	0	0	20.0	1.0	2.5	2.5	10.0	3,190.4

The estimation result of a quadratic model is shown in <Table 3>. Although there are many statistically insignificant estimates, all these were remained for the following two main reasons. First, the 5 factors must be included in the model because those are to be calibrated in the Gipps car-following model. Second, the objective of this calibration framework is to estimate the response variable as much as possible, so simplifying the model will not benefit in that sense.

<Table 3> Estimation Result of Quadratic Regression Model

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6,283.95	729.49	8.61	<.0001
X1*	-158.46	13.41	-11.82	<.0001
X2	-111.81	335.30	-0.33	0.743
X3	10.95	134.12	0.08	0.936
X4	37.46	134.12	0.28	0.784
X5	3.05	33.53	0.09	0.929
(X1-20)*(X1-20)	18.19	2.32	7.83	<.0001
(X1-20)*(X2-1)	14.50	82.13	0.18	0.862
(X2-1)*(X2-1)	2,853.49	1,451.91	1.97	0.068
(X1-20)*(X3-2.5)	-3.45	32.85	-0.11	0.918
(X2-1)*(X3-2.5)	169.66	821.32	0.21	0.839
(X3-2.5)*(X3-2.5)	473.49	232.30	2.04	0.060
(X1-20)*(X4-2.5)	-10.92	32.85	-0.33	0.744
(X2-1)*(X4-2.5)	29.66	821.32	0.04	0.972
(X3-2.5)*(X4-2.5)	112.58	328.53	0.34	0.737
(X4-2.5)*(X4-2.5)	693.11	232.30	2.98	0.009
(X1-20)*(X5-10)	-1.04	8.21	-0.13	0.901
(X2-1)*(X5-10)	54.68	205.33	0.27	0.794
(X3-2.5)*(X5-10)	3.31	82.13	0.04	0.968
(X4-2.5)*(X5-10)	17.96	82.13	0.22	0.830
(X5-10)*(X5-10)	26.99	14.52	1.86	0.083

* Variable description is shown in <Table 1>.

In the canonical curvature of RSM, if all of the eigenvalues were positive, the global minimum can be found by using the unique stationary point of the fitted surface. In this case, the global minimum was found as 2808.9 at X1=24.36, X2=1.01, X3=2.50, X4=2.51, and X5=10.02. However, instead of using this unique stationary point to find the optimum, GA was used to generate a population near the optimum, since the fitted surface is unbounded when there are both positive and negative eigenvalues.

2. 1st Genetic Algorithm

Once the fitted second-order model for the negative sum of log-likelihood is estimated, this regression model is

plugged into GA as a fitness function. The basic idea of running GA with the regression model is to find a near optimum and start 2nd GA with relatively good initial population (i.e., good parents). Thus, the better fitted regression model will bring the initial population closer to optimal parameter sets and improve the fitness value at start of 2nd GA.

3. 2nd Genetic Algorithm

The major difference of 2nd GA is the initial population. The initial population is used to seed in the GA. Traditionally, the initial populations are generated using pseudo random numbers, but in this proposed framework, the 2nd GA uses the final population of 1st GA as its initial population. The advantage of this approach is that we can start GA from near optimum sets or relatively better populations, compared to starting from the pseudo random numbers.

As a number of options in Matlab's GA tool affect resultant solutions, the most of them were remained as default set-up with a few changes (e.g. population size = 20, maximum number of iterations = 2,000) to consistently compare the results between the conventional GA and the proposed approach.

4. Simulation Set-up

The following is a pseudo code of the proposed framework, and it was run by Matlab scripts. Note that the conventional GA for model calibration skips the code line number 7-12 and uses initial population generated from pseudo random numbers.

```
A Dual Genetic Algorithm using Central Composite Design
1  : CCD_X = 36-run Combinations of 5 parameters using Central Composite Design (36-by-5 matrix)
2  : Y = Response variable (the negative sum of log-likelihood)
3  : Model_Gipps = Gipps Car-following Model (input=5 parameters, output= the negative sum of log-likelihood)
4  : X_1 = The local minimum from 1st genetic algorithm model
5  : P_1 = The final population from 1st genetic algorithm model
6  : X_2 = The local minimum from 2nd genetic algorithm model
7  : P_2 = The final population from 2nd genetic algorithm model
8  : CCD_X ← Construct CCD and convert to natural unit
9  : for i = 1 to size(CCD_X)
10 :        Y_i = Model_Gipps (x = CCD_X_i)
11 : end for
12 : Model_1 = Fit a (second order) quadratic regression model (x=CCD_X, y=Y)
13 : [X_1, P_1]= GA[Fitness function = Model_1;  Initial population = CCD_X]
     [X_2, P_2]= GA[Fitness function = Model_Gipps;  Initial population = P_1]
```

IV. Case Scenarios

The main contribution of the proposed calibration approach is to increase the efficiency of a calibration process. Therefore, the performance of the proposed approach, named CCD-GA, was compared with that of the conventional

GA approach in this study.

In order to calibrate a car-following model, ground truth data are required for validation of parameter estimates. Two case scenarios were established with different ground truth data: simulated trajectory using the Gipps model and sampled trajectories from the Next Generation Simulation (NGSIM) datasets. In the first scenario the purpose of using simulated data as ground truth is to avoid the accuracy issue of the selected car-following model. Hence the characteristics of the proposed calibration approach can be explored more clearly. In contrast, using real data, NGSIM, can give more realistic verification results for the proposed approach.

1. Scenario 1 Using Simulated Vehicle Trajectory

Data Collection for the Lead Vehicle

In order to generate the following vehicle's behavior using the car-following model, the lead vehicle's driving trajectory data is required. To get the real vehicle trajectory data, the latitude/longitude data and the corresponding speed of a vehicle were collected every second using a GPS data logger while driving. The data collection was carried out during off-peak hours of a weekday on an arterial in an urban area. Therefore, it includes comprehensive driving behaviors in terms of speed and acceleration. The 300-second data was selected to be used as the lead vehicle's driving behavior, in which stop-and-go, acceleration and deceleration characteristics with varying magnitude are included.

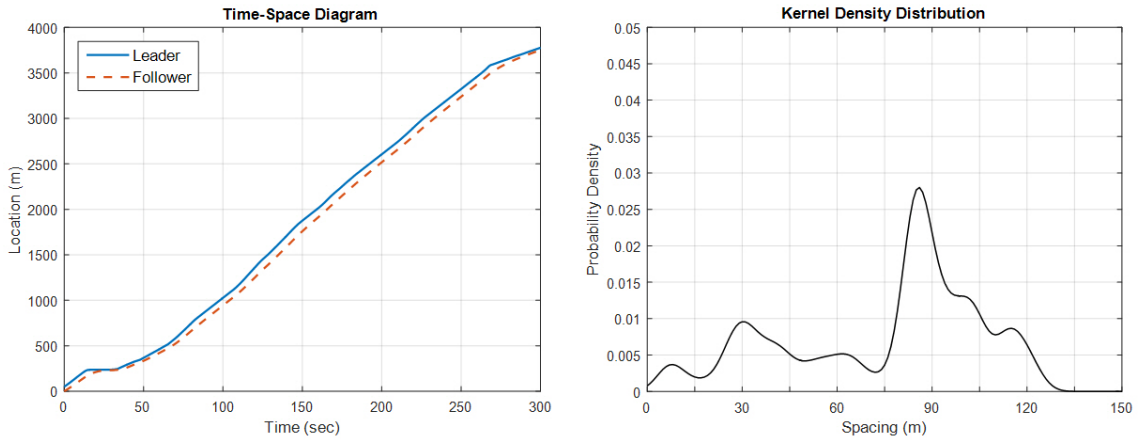### Data Generation for the Following Vehicle

With respect to the lead vehicle's data, the following vehicle's driving attributes were generated using the Gipps car-following model with the parameter values from Vasconcelos et al.(2014) (see Table 4). It was assumed that those parameter values are correct for this study. Since the purpose of this study is to propose a new approach for model calibration and verify its characteristics, this assumption allows to avoid judging the accuracy of the Gipps car-following model.

<Table 4> Assumed Parameter Values for Following Vehicle in Scenario 1

Parameters	v_n^d(m/s)	a_n(m/s2)	d_n(m/s2)	$d_{n-1}^{'}$(m/s2)	$S_{(n-1)}$(m)	τ(sec)
Values*	26.00	0.82	2.53	2.78	5.20	0.40

* Parameter values were imported from Vasconcelos et al.(2014).

The time-space diagram of the two vehicles is shown in <Fig. 2>. The spacing between two vehicles, i.e., the distance between two front bumpers of successive vehicles, for every second is measured so that the kernel density distribution was estimated as in <Fig. 2>. Finally, the negative log sum of the distribution was calculated, which was used in the GA in order to compare it with optimized statistics during the iterations.
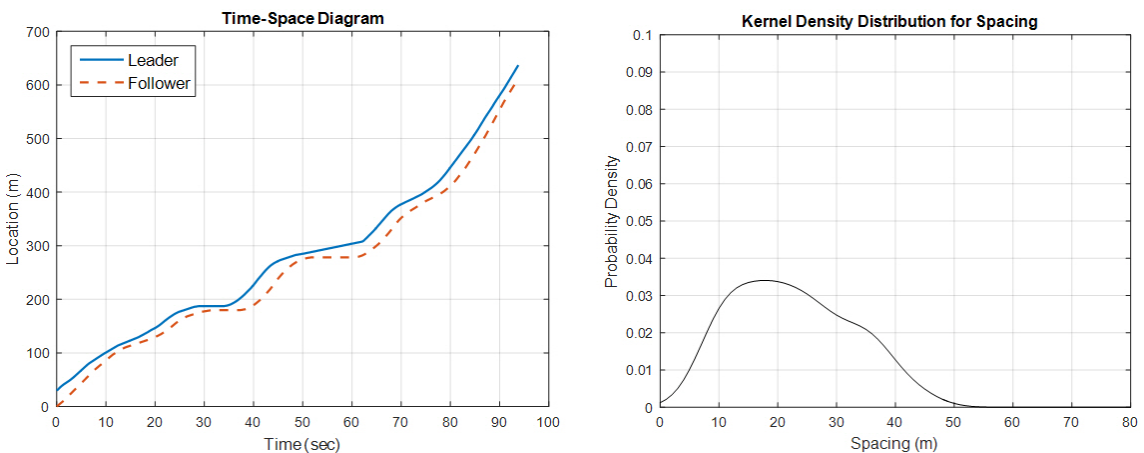
<Fig. 2> Trajectory and spacing distribution of ground truth data in Scenario 1

2. Scenario 2 Using Real Vehicle Trajectory

NGSIM US-101 Datasets

The United States Department of Transportation (US DOT) Federal Highway Administration (FHWA) launched the NGSIM program in 2005 to develop microscopic traffic behaviour models. As a part of this program, the US 101 trajectory dataset(FHWA,2016) was collected on southbound US 101, the Hollywood Freeway in California. Every individual vehicle's trajectory data were collected on the 640-meter (2,100-feet) length of study area during the morning peak hours.

The trajectories of two successive vehicles on the left-most lane were sampled because they shows deceleration and acceleration patterns repetitively. <Fig. 3> illustrates the time-space diagram and the kernel density distribution for the spacing between two vehicles.
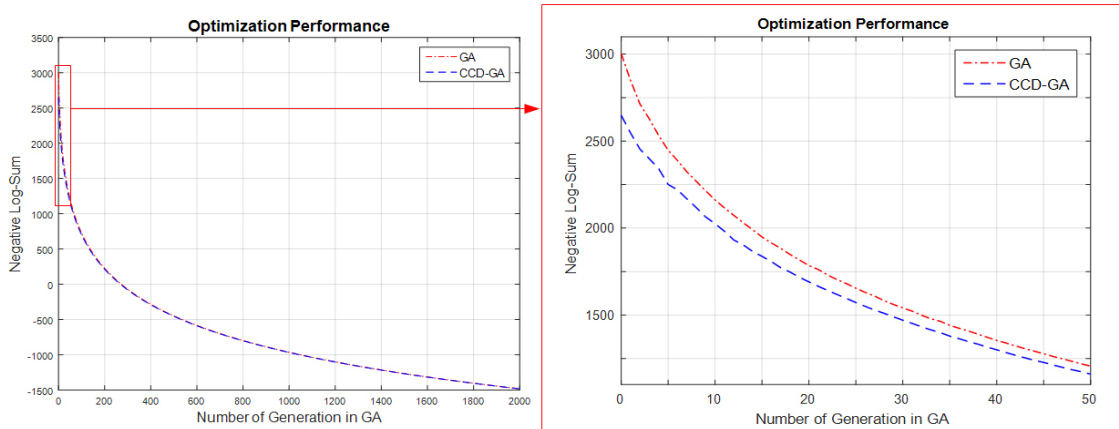

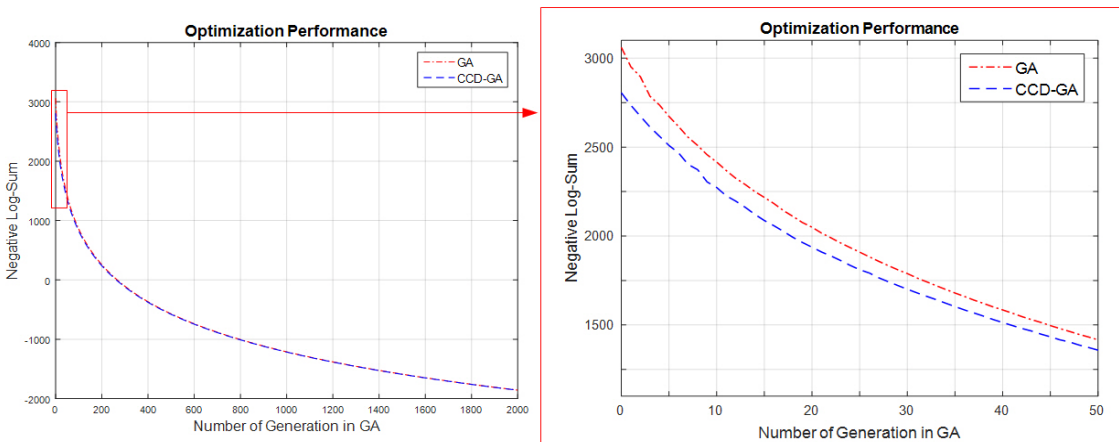<Fig. 3> Trajectory and spacing distribution of ground gruth data in Scenario 2

Ⅴ. Analysis Results

1. Minimization of the Negative Log-Sum in GA

Total 2,000 generations were searched successively in the GA procedure to find the optimal set of parameter estimates. Although the negative log sums do not seem to converge after the 2,000 generations, the difference between two approaches can be observed from the beginning of iterations. <Fig. 4(a)> and <Fig. 4(b)> display the optimization performance of both approaches for Scenario 1 and 2, respectively. The traditional method, to apply a single GA, starts above 3,000 of the log sum for both scenarios. However, the proposed approach, to use double GA with CCD, does at around 2,500, which is the value that the single GA approach can reaches after the first three iterations. This is the strong evidence that the CCD-GA method starts to search from the area which is closer to the optimum than the single GA does. It implies that the CCD-GA can reduce a number of iterations from the beginning of the calibration process compared with the traditional method.

(a) Scenario 1

(b) Scenario 2

<Fig. 4> Optimization performance over iteration: (a) Scenario 1 and (b) Scenario 2
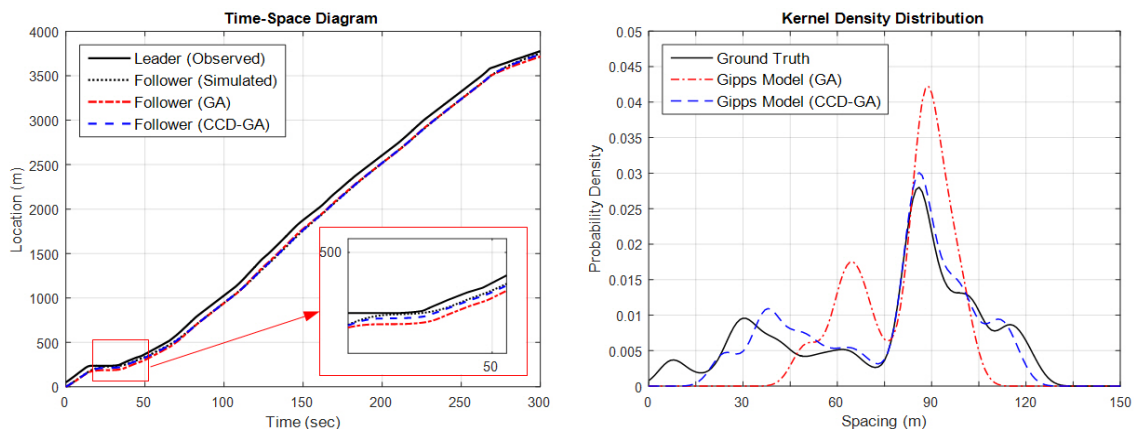
2. Comparison of the Follower's Driving Behaviors

<Table 5> shows the optimal parameter values from the CCD-GA approach for the Gipps model with respect to the ground truth data. Although the negative log-sum of both approaches seemed to converge after the 2,000 iterations, their solutions varied within the searching space, particularly for the maximum acceleration(a_n), the follower's estimate of the lead vehicle's maximum deceleration rate(d'_{n-1}) and the effective length($S_{(n-1)}$). If there was strong evidence that the solution should not be accepted, then one can go back to the experimental design step and modify the parameter ranges.

<Table 5> Optimal Parameters

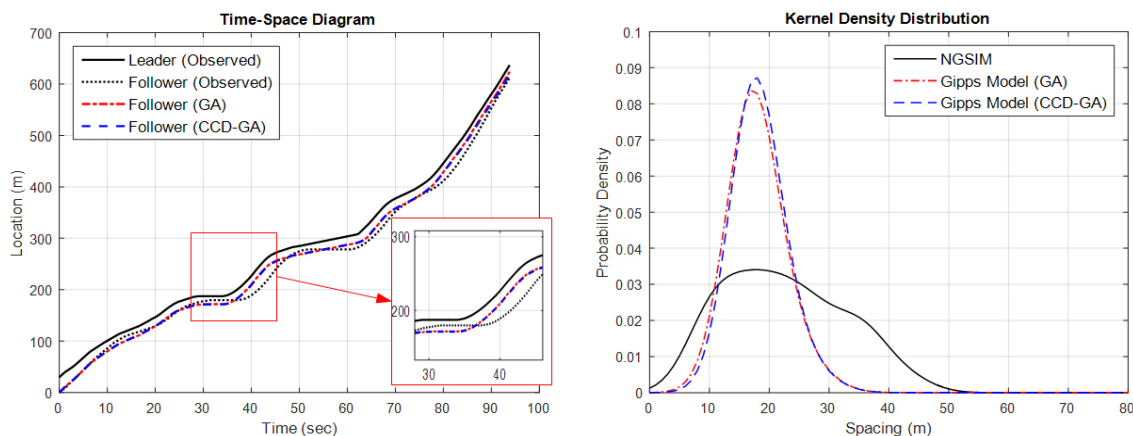| Parameters		v_n^d(m/s)	a_n(m/s2)	d_n(m/s2)	d'_{n-1}(m/s2)	$S_{(n-1)}$(m)
Scenario 1	GA	55.94	1.92	2.81	1.26	48.24
	CCD-GA	60.00	5.00	0.01	2.50	23.40
Scenario 2	GA	58.21	4.19	6.99	0.86	15.49
	CCD-GA	55.34	4.33	0.50	1.01	16.00

Interestingly, two reproduced trajectories of the following vehicle in Scenario 1 were almost overlapped with the ground truth data in the time-space diagram in <Fig. 5>, despite of the variance of parameter values between two methods. One potential cause of this result is the fact that the lead vehicle's driving trajectory is too smooth except the complete stop between around 20 and 35 second, so that there are less variations in the follower's reaction. On the other hand, the solution parameters of CCD-GA approach reproduced a more accurate pattern for the lead vehicle's stop than those of GA. As a result, this made the kernel density distribution of the proposed method more similar to the ground truth than that of GA.

<Fig. 5> Comparison of the trajectories and kernel density distribution for Scenario 1

The reproduced trajectories for Scenario 2 are shown in <Fig. 6>. Both calibration approaches generated the overlapped trajectories each other despite of the different optimization performance at the beginning of the

iterations, shown in <Fig. 4(b)>. However, those trajectories were not overlapped with the ground truth. This result stems from that the Gipps model tends to fall short of flexibility and sensitivity to reproduce an irregular spacing pattern in such repetitive accelerating and decelerating traffic flow condition in <Fig. 6>.

<Fig. 6> Comparison of the trajectories and kernel density distribution for Scenario 2

VI. Conclusions and Discussions

Calibration of a car-following model is an essential part of microscopic traffic simulation analyses. And GA has been frequently used for it. Despite the high performance of GA, it starts with the initial population that is assigned as random.

In this study, the dual GA combined with CCD has been proposed in order to improve the efficiency of the calibration process. Using the CCD, GA can start to search from the area that is closer to the optimum so that it can skip initial iterations of a simulation run. The higher efficiency of the proposed method was identified by comparing with the single GA approach which is the convention in traffic simulation area with the simulated and real vehicle trajectory data.

The proposed approach has benefits for the following cases. First, when a single simulation run takes a very long time (e.g. for a physically large-scale network analysis), reducing the number of simulation runs gives higher efficiency. Second, when one needs to find the near-optimum within a relatively short operational time to support fast decision making, this framework can reach the goal more quickly than the conventional approach does. This can be a case of real-time traffic analyses.

Many other options in the proposed method should be investigated for future studies in order to enhance the improvement in this paper. Running GA twice in the proposed algorithm brings additional complexity in computation, compared to running a single GA. Although the impact was not significant in the case study, improvement in the first GA may be necessary for a simulation analysis in real-time traffic operations. Adding more experiment runs in CCD or using a full factorial CCD may be able to estimate better population potentially. Changing GA parameters is another possibility. This includes to change searching boundary dynamically over

generations, to use different selection methods, and so on. Applying more complex car-following models may show more evident improvement compared with the conventional GA approach. Lastly, another optimization method as a substitution of the second GA should be tested and compared with the proposed method. Neural-network-based models are known to perform well to find a solution in nonlinear optimization problems.

ACKNOWLEDGEMENTS

REFERENCES

Box G. E. P. and Wilson K. B.(1951), "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 13, no. 1, pp.1-45.

Cheu R. L., et al.(1998), "Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm," *Journal of Transportation Engineering*, vol. 124, no. 6, pp.526-535.

Cunto F. and Saccomanno F. F.(2008), "Calibration and validation of simulated vehicle safety performance at signalized intersections," *Accident Analysis & Prevention*, vol. 40, no. 3, pp.1171-1179.

Eriksson A.(1999), "Calibration of a traffic simulation model by using factorial experiments and response surfaces," LiTH-MAT-Ex-2000-14.

FHWA(2016), *Next Generation Simulation (NGSIM) US Route 101 Dataset*, FHWA, Editor, U.S. Department of Transportation Intelligent Transportation Systems Joint Program Office (JPO): Washington, D.C.

Gipps P. G.(1981), "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp.105-111.

Howe A. and Bozdogan H.(2010), "Predictive Subset VAR Modeling Using the Genetic Algorithm and Information Complexity," *European Journal of Pure and Applied Mathematics*, vol. 3, no. 3, pp.382-405.

Kesting A. and Treiber M.(2008), "Calibrating car-following models by using trajectory data: Methodological study," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2088, pp.148-156.

Kim S. J., Kim W. and Rilett L.(2005), "Calibration of microsimulation models using nonparametric statistical techniques," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1935, no. 1, pp.111-119.

Mee R. W.(2009), *A Comprehensive Guide to Factorial Two-Level Experimentaion*, Springer Science & Business Media.

Park B. B. and Qi H. M.(2005), "Development and Evaluation of a Procedure for the Calibration of

Simulation Models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1934, pp.208-217.

Park B. B., Won J. and Yun I.(2006), "Application of Microscopic Simulation Model Calibration and Validation Procedure," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1978, pp.113-122.

Treiber M. and Kesting A.(2013), "Microscopic Calibration and Validation of Car-Following Models – A Systematic Approach," *Procedia-Social and Behavioral Sciences*, vol. 80, pp.922-939.

Van Hinsbergen C., et al.(2009), "Bayesian calibration of car-following models," *IFAC Proceedings Volumes*, vol. 42, no. 15, pp.91-97.

Vasconcelos L., et al.(2014), "Calibration of the Gipps Car-following Model Using Trajectory Data," *Transportation Research Procedia*, vol. 3, pp.952-961.