Effects of microscopic traffic platform calibration on errors in safety and traffic metrics

David D.Q. Duong
PhD Candidate, Department of Civil and Environmental Engineering, University of Waterloo
Waterloo ON, Canada, e-mail: dqduong@uwaterloo.ca

Frank F. Saccomanno
Professor, Department of Civil and Environmental Engineering, University of Waterloo
Waterloo ON, Canada, e-mail: saccoman@uwaterloo.ca

Bruce R. Hellinga
Professor, Department of Civil and Environmental Engineering, University of Waterloo
Waterloo ON, Canada, e-mail: bhellinga@uwaterloo.ca

Submitted to the 3rd International Conference on Road Safety and Simulation,
September 14-16, 2011, Indianapolis, USA

Abstract

A Pareto Archived Dynamically Dimensioned Search (PA-DDS) algorithm is introduced for the calibration of microscopic traffic simulation platforms and compared to other single and multi-criteria calibration approaches. PA-DDS algorithm explicitly considers trade-off in errors among the various constituent fitting function such that any solution on the non-dominated curve cannot be improved without incurring a corresponding degradation in error in at least one of the constituent fitness criteria. For example, improvement in speed error cannot be achieved without increased error in either volume or safety performance or both. In this paper, PA-DDS is used to calibrate selected VISSIM model parameters based on observed traffic data obtained from the Federal Highway Administration (FHWA) Next Generation Simulation (NG-SIM) vehicle tracking study. The calibration seeks to obtain best-estimate parameter values that minimize residual mean square percentage error (RMSPE) for three fitness criteria: speed, volume, and Crash Potential Index per vehicle (a surrogate safety performance metric). This comparison clearly demonstrates that the multi-criteria PA-DDS algorithm yields best-estimate parameter values with acceptable residual errors for the two traffic factors of speed and volume, as well as for the safety performance criteria. The best estimate VISSIM parameters obtained from the PA-DDS application were found to differ significantly from default values and from values obtained based on other calibration methods that do not explicitly consider trade-off errors in fitness criteria. A number of solution sets were obtained from the PA-DDS algorithm, with a range of parameter values. The best estimate solution (lowest overall model goodness-of-fit) yielded parameters that differed from other PA-DDS solution sets (with higher overall error). This suggests that trade-offs (non-dominated sets) of parameter values can have significant implications for values that correspond to the lowest overall model goodness-of-fit.

Keywords: transportation, safety, microscopic, simulation, calibration

Introduction

Microscopic traffic simulation models have been receiving increasing attention as an effective means of analyzing traffic operations and safety for a wide spectrum of mitigating factors. Critics of these types of models, however, have argued quite effectively that the results obtained from simulation have not been adequately verified with regard to observational data, and hence may be suspect when compared to "real world" traffic conditions. A major challenge to a more extensive adoption of traffic simulation models remains bridging the gap between simulated and real-world driving experience (Sayed and Zein, 1999). To bridge this gap it is important that input model parameters are accurately determined such that simulated traffic attributes reflect observational data for different road and traffic conditions.

A number of recent calibration studies have adopted evolutionary-based search algorithms using single criterion fitness functions (e.g. normally travel time or flow). These are discussed by Ma and Abdulhai, 2002; Hourdakis et al, 2003; Park and Qi, 2005; Kim et al, 2005; Cunto and Saccomanno, 2008; Cicu et al, 2011; Vaiana and Gallelli, 2011. Table 1 provides a brief description of the main features of these single-criterion calibrations for several traffic platforms (VISSIM, PARAMICS, and AIMSUM).

Table 1 Single-criteria parameter calibration studies

Study	Type of Optimization	Model	Network	Measures of Performance	Results	Note
Ma and Abdulhai (2002)	genetic algorithm	PARAMICS	Arterial Network	flows	46.09 % (GRE)	Global relative error
Hourdakis et al (2003)	heuristic search	AIMSUM	Freeway	volume	8.84 % (RMSPE)	Root mean squate percentage error
Park and Qi (2005)	genetic algorithm	VISSIM	Freeway interchance	travel time	12.6 % (RMSPE)	Root mean squate percentage error
Kim et al (2005)	genetic algorithm	VISSIM	Freeway network	travel time	1 % (MAER)	Mean absoluted error ratio
Cunto and Saccomanno (2008)	genetic algorithm	VISSIM	Intersection	Crash Potential Index (CPI)	0.026 % (RMSPE)	Root mean squate percentage error
Cicu et al (2011)	Experimental	VISSIM	Roundabouts	Capacity	Visual Inspection (graphically)	Authors did not estimate errors
Vaiana and Gallelli (2011)	Experimental	VISSIM	Roundabouts	Speed	5 % (MAER)	Mean absoluted error ratio

In principle it can be argued that the single criterion approach fails to recognize that traffic is a multi-faceted entity. Accuracy in one attribute (e.g. travel time or speed) does not ensure accuracy in another attribute (e.g. acceleration or spacing), or even accuracy in the overall model outputs. Lack of accuracy in any important traffic attribute can significantly bias the simulation results, and in general weaken the acceptance of simulation as a logical tool for analyzing safety performance, traffic operations, or vehicle emissions. In this paper, it is argued that the calibration process needs to take into account trade-offs in traffic attribute error as well as minimizing overall model goodness of fit.

A number of recent calibration studies have acknowledged the multi-faceted nature of traffic by adopting a multi-criteria approach to calibration. A number of these multi-criteria calibration studies are summarized in Table 2 for several traffic model platforms (MITSIMLab, PARAMICS, AIMSUM and VISSIM). It should be noted that many of these studies have considered a maximum of two related traffic attributes - mostly speed and volume.

Table 2 'Multi-criteria' parameter calibration

Study	Type of Optimization	Model	Network	Measures of Performance	Results	Note
Toledo et. al. (2004)	iterative averaging	MITSIMLab	Freeway	Speed and Density	4.6 % (MAE for speed)	Only speed data shown; does not apply multi-criteria framework
Balakrishna et. al. (2007)	Simultaneous Perturbation Stochastic Approximation (SPSA)	MITSIMLab	Freeway	Volume (Counts)	22 to 65 % (RMSPE)	Introduces a multi-criteria framework but does not apply it
Ma et. al. (2007)	SPSA	PARAMICS	Freeway	Link capacity and critical occupancy	0.70 % (Sum of GEH)	Two-criteria calibration
Ciuffo et. al. (2008)	OptQuest/Multistart Heuristic) OQMS	AIMSUM	Freeway	Volume (Counts) and Speed	11 % (RMSPE speed); 17% (RMSPE Volume)	Two-criteria calibration
Duong et. al. (2010)	Genetic Algorithm	VISSIM	Freeway	Volume and Speed	1.9 % (RMSPE Speed); 10.5 % (RMSPE Volume)	Introduces the concept of Pareto optimality (non dominance) to the traffic calibration problem
Huang and Sun (2009)	NSGA II	VISSIM	Freeway	Volume and Speed	1.0 (Volume Fitness) and 0.97 (Speed Fitness)	Applies the NSGA II without looking at the resultant non dominance set

Any thorough calibration exercise must be able to identify the trade offs in error for different attributes, as well as the effect on overall model goodness-of-fit. For instance, how does error in one criterion such as speed affect error in another error such as, flow; and if speed and flow are inputs in safety performance, how do errors in these traffic attributes affect error in safety performance? A basic shortcoming in current multi-criteria calibrations is that trade-offs in error between different traffic attributes has not been considered explicitly with respect to errors in safety performance or in overall model goodness-of-fit. While several studies have recognized this issue, their attempts to resolve it has focused on subjective weighting procedures (Ma et al 2007; Ciuffo et al 2008); ie. expressing model goodness-of-fit in terms of a weighted combination of fitness criteria errors. The problem with this approach is that the weights themselves are treated externally to the calibration exercise, such that their values are determined subjectively or require calibration themselves.

In their formulation, Fonseca and Fleming (1993) combined Pareto optimality with a Genetic Algorithm to solve the multi-criteria calibration problem with trade-offs in constituent traffic attribute errors. They refer to this approach as Multi-objective Genetic Algorithm (MOGA). In

MOGA instead of converting the multi-criteria calibration into a combined weighted fitness function, trade-offs in different fitness errors are considered explicitly using dominance/non-dominance selection. The result of the MOGA calibration is a set of points that are Pareto non-dominated, such that any solution in this set cannot be improved upon without incurring a corresponding degradation in error in at least one other constituent criterion. For example, improvement in speed error cannot be achieved without increased error in either volume or safety performance or both.

The MOGA approach has been applied to solve multi-criteria calibration problems (Koski, 1994; Yapo, 1998; Madsen, 2000; Cheng et al, 2002; Shea et al, 2006) in a wide range of engineering disciplines. A number of these studies are summarized in Table 3.

Table 3 MOGA problems outside of transportation

Study	Field	Type of Optimization	Problem	Measures of Performance
Shea et al (2006)	Structural/ Construction	Ant Colony	Building Envelope Design	11 criteria, including costs, lighting, thermal conduction, veiw of the Eiffel Tower
Koski (1994)	Structural/ Construction	Heuristic	Design of a Flexural Plate	2 criteria, weight and deflection
Madsen (2000)	Hydrology	Shuffled Complex Evolution Algorithm	MIKE 11/NAM rainfall-runoff model	4 criteria, overall volume, overall error, peak flow, low flow
Yapo et al (1998)	Hydrology	Multi-objective complex evolution global optimization algorithm	Sacramento Soil Moisture Accounting Model and National Weather Service River Forcasting System	2 criteria, two fitting functions for flows
Cheng et al (2002)	Hydrology	Fuzzy Optimal Genetic Algorithm	Conceptual rainfall–runoff models (CRRS)	3 criteria, rainfall, runoff and evaporation

The studies in Table 3 found that MOGA yielded better calibration results for multi-criteria problems than the conventional weighted method. A major shortcoming of the MOGA approach, however, is that previous generations of non-dominated solutions in GA are not retained following the parent selection process. While this might enhance the speed of the search process, and reduce the number of computations involved, there is a possibility that the optimum (solution) may actually reside in the parent set. Hence, to ensure that an acceptable solution is obtained there needs to be a way of retaining or archiving solutions on the non-dominated set in the parent pool of solutions.

Knowles and Corne (2000) improved the MOGA approach by formulating a new class of algorithms, called the Pareto Archive Evolutionary Strategies (PAES), where the Pareto set of surviving solutions is recorded or archived throughout the iterative process. In the GA, the next generation of ‘offspring’ characteristics (traffic attributes) are created from mutation and/or crossover of ‘parent’ set (current non-dominated Pareto set), and these off-spring solutions then replace the ‘parent’ solution set if and only if any one of these solutions is dominated. Otherwise the parent solution is retained.

In this paper, a Pareto Archived Dynamically Dimensioned Search Algorithm (PA-DDS) is described and subsequently applied in calibrating the VISSIM traffic simulation model. This paper has two specific objectives:

1) Present the results of a PA-DDS multi-criteria calibration based on the three RMSPE fitness functions for speed, volume, and safety performance. Safety performance is measured in terms of a Crash Potential Index (CPI) per vehicle (after Cunto and Saccomanno, 2009).

2) Compare the results of the PA-DDS calibration with several conventional calibration methodologies for the same dataset. These include a single-criterion calibration based on either speed, volume or safety performance, and a combined weighted summation measure for the two traffic and one safety performance criteria.

For the calibration presented in this paper, observed vehicle tracking data from the FHWA (2007) NG-SIM program are used.

CALIBRATION APPROACH

The calibration procedure that is illustrated in Figure 1 involves three basic steps:

1. Establish statistical significance of model parameters (select parameters to be calibrated)
2. Select appropriate fitness criteria(s) with corresponding fitness function(s)
3. Obtaining best estimate parameter values that results in acceptable fitness errors with corresponding trade-offs

The primary focus of the paper is on the second and third steps.

While the ultimate purpose of the simulation model (when used as a road safety assessment tool) is to generate measure of safety performance, these measures are themselves a complex function of traffic inputs/outputs. Hence, the simulation must capture individual driver/vehicle responses and actions that will vary over time along a given road segment. A number of different measures of safety performance have been documented in the literature, including: “time to collision” (TTC), “deceleration rate to avoid the crash” (DRAC), “post encroachment time” (PET), “crash potential index” (CPI). In this paper, we have adopted CPI/vehicle from Cunto and Saccomanno (2009).

For a given criteria, fitness can be expressed in a number of different forms (Ciuffo and Punzo, 2009). In this study, fitness is expressed in terms of the root mean square percentage error:

$$\text{Root Means Squared Percentage Error}_k = \sqrt{\frac{1}{n}\sum\left(\frac{S_t^k - O_t^k}{O_t^k}\right)^2} \qquad (1)$$

Where, S_t = simulated value for traffic factor k (e.g. speed) at time increment t
O_t = observed value for traffic factor k at time increment t
n = number of time increments in simulation

Screening parameter inputs for statistical significance can have a major effect on reducing the number of parameters in need of calibration, and hence improve the efficiency of the calibration exercise. VISSIM for example requires the specification of more than 30 parameters (Table 4) to simulate three different driving regimes, car-following, lane-changing, gap-acceptance. Only those having a significant effect on traffic attributes and corresponding safety performance need to be considered. Presumably the nature of the driver behaviour parameters and their significance depends on the geometric attributes of the study area being considered. For example freeway driver behaviour is expected to differ from arterial driver behaviour, etc.

To obtain the VISSIM parameters in need of calibration, a heuristic procedure for establishing the significance of model parameters as introduced by Cunto and Saccomanno (2008) has been applied. A total of seven significant parameters were determined to affect the three fitness criteria (speed, volume, CPI/veh), and these are summarized in Table 4.

Table 4 VISSIM parameters affecting speed, volume, and CPI (PTV 2008)

Number	Parameter	Description	Lower Bound	Upper Bound
1	(max) Look ahead Distance (m)	Defines the distance that vehicles can see forward to react to other vehicles in front or beside it on the same link	50.00	300.00
2	CC0	Standstill distance (m), which defines the desired distance between stopped vehicles	0.50	3.00
3	CC1	Headway time, is the time in seconds that a driver wants to keep. Setting a high value will make drivers more cautious	0.50	1.75
4	CC3	Threshold for entering Following, controls the start of the deceleration process. By setting this higher, a driver will wait longer before decelerating to the safe distance.	-15.00	-4.00
5	CC5	For positive speed differences; following thresholds control the speed differences during the following state. Smaller values result in a more sensitive reaction of drivers to accelerations or decelerations of the preceding car	0.10	2.00
6	Accepted deceleration	For the trailing vehicle.	-2.50	-0.25
7	Safety distance reduction factor	takes effect for; a) the safety distance of the trailing vehicle in the new lane for the decision whether to change lanes or not, b) the own safety distance during a lane change and c) the distance to the leading (slower) lane changing vehicle.	0.20	0.80

The screening process for establish statistical significance as applied to the VISSIM was found to significantly reduce the number of parameter requiring calibration, from 30 to 7, five for car-following and two for lane-changing models. It should be noted that these results apply to freeway driving regimes only.

PA-DDS Multi-criteria procedure

The Pareto Archived Dynamically Dimensioned Search Algorithm (PA-DDS) used in this study was developed by Asadzadeh and Tolson (2009) (see Appendix) to include non-dominance and crowding distance. As illustrated in Figure 1, the PA-DDS calibration consists of several steps:

1. Perform single-criterion calibration separately for each criteria and retain all solutions (set of parameter values).
2. Combine all these solution sets and sort the resultant solution into a non-dominated set using an algorithm suggested by Deb et al (2000)
3. For each solution in the non-dominated set, a crowding distance is calculated to enhance the search spectrum to include maximum variation in parameter values (Deb et al, 2000)
4. Select one solution based on order of dominance and highest crowding distance, and mutate the parameter values to generate next generation solution
5. Sort and determine the new non-dominated sets and recalculate crowding distances. If the mutated solution is found to be non-dominated it is retained, otherwise we return to step 4 and select another solution.
6. The process is terminated if either of two conditions are satisfied, ie. no further changes are observed in the membership of the highest non-dominated set, or the maximum number of iterations is reached as specified by the user.

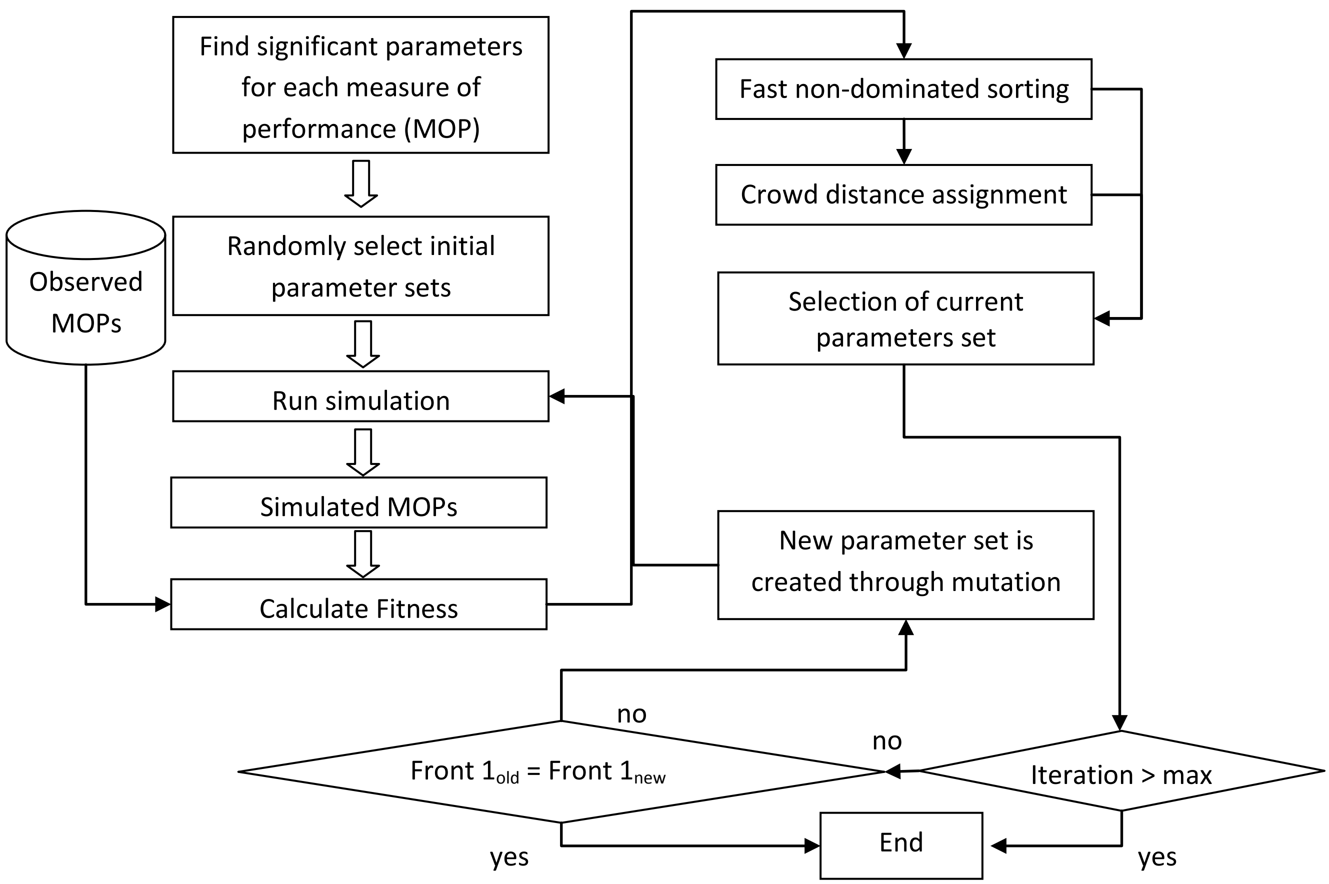

Figure 1 Flow chart of calibration procedure

The solutions in the combined sets (Step 2) include both dominated and non dominated solutions, with the optimum set of parameter values occurring in the non-dominated region. The Pareto (non-dominance and dominance) concepts are illustrated in Figure 2 for the minimization problem. When solution j = (j_1, j_2) is compared to k = (k_1, k_2), for all n=1 and 2, $j_n < k_n$; hence solution k is dominated by solution j. A solution is considered dominated, if there exist another solution in the set with lower errors in all criteria. Comparing solutions j and l, $j_1 < l_1$, but $j_2 > l_2$, and therefore neither solution dominates the other. There are no other solutions in this set that is better in all criteria for j and l and therefore these solutions are non-dominated. Simulation runs can then be ranked into a series of non-dominated classes, c_n, where lower values of n correspond to higher non-dominated sets (i.e. solution sets with lower overall model goodness-of-fit).

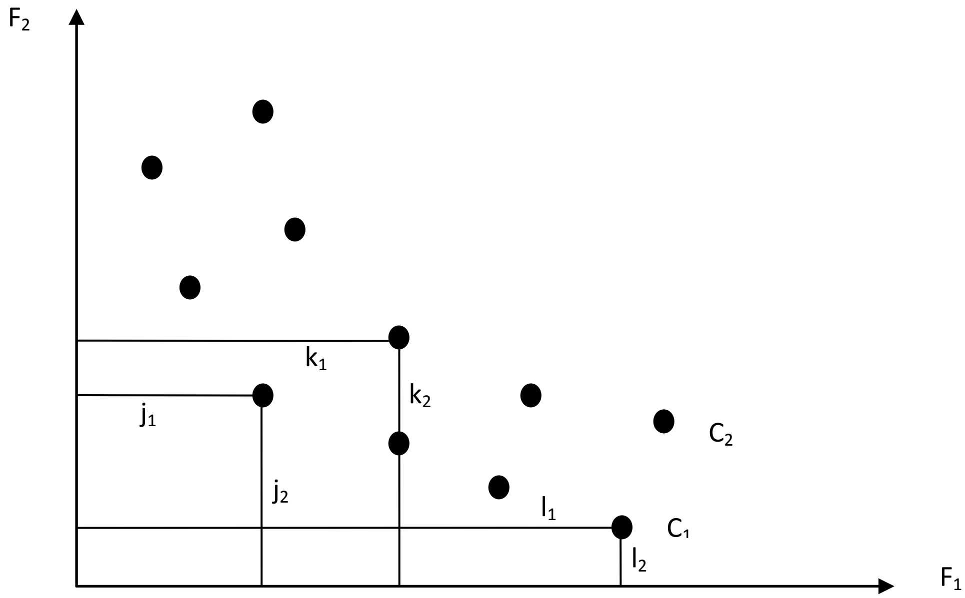

FIGURE 2 Graphical illustration of non-dominated sets.

Crowding distance was introduced into PA-DDS to ensure that solutions found in more crowded regions of each non-dominance curve are deleted from the surviving solution set. Solutions in crowded regions of the solution space often share common attributes (e.g. parameter values), and hence, in order to capture the full spectrum of parameter attributes in the surviving set it is preferred to mutate these survivors from solutions found in less crowded regions. Thus we ensure that the maximum variation in parent attributes is passed onto the off-spring set. As illustrated in Figure 3, for each point on the same non-dominated set a cuboid is established with respect to its two neighbouring points, and a crowding distance, $I_{i,distance}$, is estimated in terms of the average of the cuboid lengths.

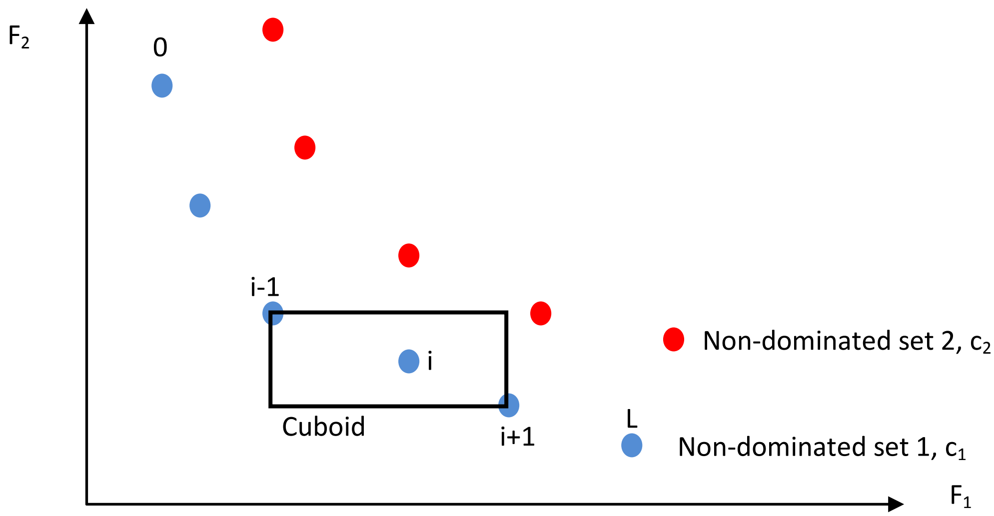

FIGURE 3 - Graphical depiction of crowd distance calculations.

The crowding distance has been expressed by Deb et al (2000) as:

$$I_{i,\ distance} = \sum(I_{(i+1),\ Fi} - I_{(i-1),Fi}) \qquad (2)$$

The application of crowding distance requires a minimum threshold wherein neighbouring solutions are deleted from parent pool in favour of nearest solution set.

CALIBRATION CASE STUDY

In this paper the observed vehicle tracking data from the FHWA (2007) NG-SIM Interstate Highway 101 dataset is used to calibrate the VISSIM platform parameters using the PA-DDS approach. The NG-SIM data were extracted from a video taping of Highway 101, California on June 15, 2005 for a 15-minute time period (7:50 am to 8:05 am).

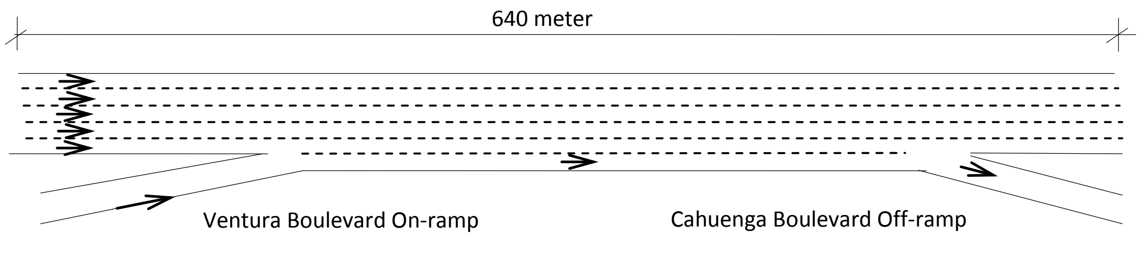

Figure 4 NG-SIM highway 101 study area

Single criteria calibration exercises were carried out and the results are summarized in Tables 5 – 7. The calibration is based on minimizing each of the three criteria for speed, volume, and CPI/vehicle separately. Table 8 summarizes the results of the calibration based on the total weighed sum of the three criteria fitness errors. For this exercise all criteria are given equal weighting. A total of 20 parameter sets were obtained from each calibration for the single and weighted summation criteria used in the calibration.

Table 5 DDS results using speed RMSPE

Trials	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reduction factor	Speed (km/h)	Volume	CPI	**RMSPE Speed**	RMSPE Volume	RMSPE CPI	SUM
1	151.44	1.81	1.08	-10.37	0.48	-2.19	0.65	95.9	2065	610,042	0.645	0.041	0.311	0.997
2	219.56	2.12	1.08	-10.37	0.48	-2.50	0.37	101.0	2066	203,774	0.732	0.040	0.770	1.543
3	80.50	2.14	1.08	-7.34	0.48	-1.87	0.49	102.0	2066	2,839,001	0.749	0.040	2.206	2.996
4	163.71	1.81	1.21	-11.44	0.48	-1.72	0.62	95.4	2065	373,982	0.636	0.041	0.578	1.255
5	209.47	2.21	1.34	-11.44	0.60	-1.72	0.61	92.4	2064	669,973	0.585	0.041	0.243	0.869
6	209.47	2.21	1.47	-6.11	0.63	-1.72	0.53	84.6	2062	649,650	0.451	0.042	0.266	0.760
7	209.47	2.51	1.55	-7.31	0.63	-1.77	0.53	78.3	2043	235,803	0.343	0.051	0.734	1.128
8	271.44	2.51	1.55	-7.31	0.10	-1.76	0.53	79.2	2047	327,290	0.358	0.049	0.630	1.038
9	223.57	2.32	1.55	-7.31	0.63	-1.77	0.53	80.4	2051	874,467	0.379	0.047	0.012	0.439
10	209.47	2.47	1.55	-10.74	0.63	-1.77	0.53	77.7	2032	384,533	0.333	0.056	0.566	0.955
11	**209.47**	**2.47**	**1.72**	**-10.74**	**0.59**	**-1.91**	**0.53**	**70.6**	**1963**	**66,869**	**0.211**	**0.088**	**0.924**	**1.224**
12	209.47	1.97	1.72	-14.03	0.59	-2.03	0.60	72.6	1962	402,531	0.245	0.089	0.545	0.879
13	209.47	2.47	1.72	-9.46	0.74	-1.86	0.53	72.7	1937	235,891	0.247	0.100	0.734	1.081
14	209.47	3.29	1.72	-9.36	0.59	-1.91	0.53	74.0	1913	371,697	0.269	0.111	0.580	0.961
15	209.47	2.47	1.50	-11.09	0.59	-1.91	0.61	84.0	2050	823,323	0.441	0.048	0.070	0.559
16	236.92	2.47	1.55	-13.98	0.59	-1.91	0.51	81.7	2051	365,084	0.401	0.047	0.588	1.036
17	196.64	2.62	1.72	-10.74	0.38	-1.91	0.53	70.7	1954	222,870	0.213	0.092	0.748	1.053
18	190.11	2.55	1.72	-10.74	0.59	-1.91	0.53	71.1	1958	302,332	0.219	0.091	0.659	0.969
19	222.33	2.47	1.72	-10.74	0.59	-1.91	0.53	72.6	1962	389,248	0.245	0.089	0.560	0.894
20	209.47	2.47	1.34	-10.23	0.59	-1.91	0.53	94.1	2064	189,842	0.614	0.041	0.786	1.441
Defaults	250.00	1.50	0.90	-8.00	0.35	-0.50	0.60	104	1992	539,547	0.787	0.0748	0.391	1.253
Observed								**58**	**2153**	**885,402**				

Table 6 DDS results using volume RMSPE

Trial	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reduction factor	Speed (km/h)	Volume	CPI	RMSPE Speed	**RMSPE Volume**	RMSPE CPI	SUM
1	59.76	1.12	0.93	-4.63	1.24	-1.11	0.57	103.6	2064	207,594	0.777	0.0413	0.766	1.584
2	135.18	1.12	0.93	-4.63	1.05	-1.28	0.55	99.6	2066	1,588,737	0.708	0.0404	0.794	1.543
3	83.94	0.85	0.52	-4.63	0.93	-1.28	0.62	104.6	2067	1,851,319	0.794	0.0399	1.091	1.925
4	83.94	1.52	0.50	-5.21	0.48	-1.28	0.56	104.5	2068	1,492,440	0.792	0.0395	0.686	1.517
5	**129.97**	**1.52**	**0.50**	**-7.62**	**0.48**	**-1.42**	**0.56**	**103.7**	**2069**	**727,334**	**0.779**	**0.0390**	**0.179**	**0.996**
6	118.28	1.55	0.65	-5.72	1.16	-1.42	0.56	104.4	2067	222,437	0.791	0.0399	0.749	1.579
7	129.97	1.52	0.50	-6.99	0.96	-0.56	0.48	104.8	2068	15,661	0.797	0.0395	0.982	1.819
8	170.42	0.61	0.52	-7.05	0.48	-1.42	0.48	103.6	2069	395,737	0.777	0.0390	0.553	1.369
9	129.97	0.76	0.86	-7.24	0.56	-0.90	0.56	103.1	2066	400,221	0.768	0.0404	0.548	1.357
10	143.50	1.52	0.50	-10.14	0.37	-1.42	0.65	103.5	2068	523,447	0.775	0.0395	0.409	1.223
11	182.06	1.79	0.50	-7.62	0.48	-1.42	0.54	102.5	2069	1,604,647	0.758	0.039	0.812	1.609
12	129.97	1.96	0.50	-7.62	0.62	-1.40	0.56	102.9	2069	958,357	0.765	0.039	0.082	0.886
13	129.97	1.52	0.72	-4.00	0.48	-1.42	0.76	101.1	2067	1,822,120	0.734	0.040	1.058	1.832
14	68.82	2.14	0.50	-9.86	0.27	-1.42	0.56	104.4	2058	4,977,715	0.791	0.044	4.622	5.457
15	112.98	1.52	0.50	-12.86	0.71	-1.42	0.56	104.6	2066	444,511	0.794	0.040	0.498	1.332
16	166.31	1.52	0.50	-7.62	0.48	-1.42	0.70	102.2	2069	1,158,228	0.753	0.039	0.308	1.100
17	129.97	0.94	0.65	-8.10	0.48	-1.42	0.71	102.1	2067	824,496	0.751	0.040	0.069	0.860
18	129.97	0.94	0.50	-7.62	0.48	-1.42	0.56	103.7	2069	709,174	0.779	0.039	0.199	1.017
19	129.97	1.52	0.50	-8.10	0.48	-1.42	0.56	103.2	2066	521,645	0.770	0.040	0.411	1.221
20	129.97	1.52	0.65	-7.62	0.48	-1.42	0.56	102.9	2066	624,740	0.765	0.040	0.294	1.100
Defaults	250.00	1.50	0.90	-8.00	0.35	-0.50	0.60	104	1992	539,547	0.787	0.0748	0.391	1.253
Observed								**58**	**2153**	**885,402**				

Table 7 DDS results using CPI RMSPE

Trials	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reduction factor	Speed (km/h)	Volume	CPI	RMSPE Speed	RMSPE Volume	**RMSPE CPI**	SUM
1	185.64	1.22	0.54	-5.05	0.15	-1.39	0.80	103.5	2068	783,598	0.775	0.0395	0.115	0.930
2	185.64	1.22	0.85	-4.18	0.15	-1.75	0.80	97.1	2066	1,946,692	0.665	0.0404	1.199	1.904
3	138.90	1.70	0.54	-5.05	0.15	-1.69	0.80	102.6	2068	1,027,223	0.760	0.0395	0.160	0.959
4	185.64	1.22	0.77	-4.00	0.81	-1.39	0.72	100.2	2066	1,801,185	0.719	0.0404	1.034	1.793
5	185.64	1.22	0.54	-5.05	0.10	-1.95	0.80	102.8	2068	1,868,404	0.763	0.0395	1.110	1.913
6	**185.64**	**1.22**	**0.50**	**-4.20**	**0.15**	**-1.35**	**0.72**	**103.9**	**2068**	**826,564**	**0.782**	**0.0395**	**0.066**	**0.888**
7	119.65	1.66	0.50	-4.20	0.15	-2.82	0.72	102.2	2069	2,373,310	0.753	0.0390	1.680	2.472
8	185.64	1.50	0.50	-4.20	0.15	-1.35	0.72	101.4	2068	3,934,349	0.739	0.0395	3.444	4.222
9	162.27	0.78	0.61	-4.20	0.15	-1.35	0.72	102.6	2067	1,444,319	0.760	0.0399	0.631	1.431
10	244.45	2.27	0.50	-4.20	0.87	-1.12	0.80	95.1	2068	8,999,845	0.631	0.0395	9.165	9.835
11	191.39	1.22	0.50	-4.20	0.10	-1.35	0.54	103.0	2069	1,691,052	0.767	0.0390	0.910	1.716
12	201.57	1.42	0.50	-4.20	0.15	-1.35	0.80	102.3	2067	1,759,941	0.755	0.0399	0.988	1.782
13	237.64	1.04	0.55	-4.52	0.15	-0.99	0.72	101.8	2069	2,867,942	0.746	0.0390	2.239	3.024
14	133.77	0.63	0.50	-4.88	0.15	-1.53	0.72	102.5	2069	1,749,924	0.758	0.0390	0.976	1.773
15	185.64	1.22	0.77	-4.80	0.15	-0.94	0.72	99.9	2069	1,851,269	0.713	0.0390	1.091	1.843
16	140.05	1.22	0.77	-4.20	0.19	-1.34	0.58	102.5	2067	950,139	0.758	0.0399	0.073	0.871
17	248.46	1.35	0.50	-6.40	-0.14	-1.77	0.72	99.7	2069	4,737,069	0.710	0.0390	4.350	5.099
18	223.59	1.22	0.86	-4.20	0.15	-1.35	0.67	99.8	2069	1,176,434	0.712	0.0390	0.329	1.079
19	185.64	1.22	0.50	-4.20	0.15	-1.35	0.63	103.3	2068	1,811,671	0.772	0.0395	1.046	1.857
20	185.64	1.22	0.50	-4.20	0.37	-1.35	0.72	102.0	2071	2,506,836	0.749	0.0381	1.831	2.619
Defaults	250.00	1.50	0.90	-8.00	0.35	-0.50	0.60	104	1992	539,547	0.787	0.0748	0.391	1.253
Observed								**58**	**2153**	**885,402**				

From Table 5, the speed criteria calibration, Solution 11 yielded the minimum fitness error for speed (21.1%). However, the safety performance metric, CPI/vehicle, resulted in an unacceptably high error of 92.4%. This is indicative of the limitation of a single criteria calibration, as only one objective function is minimized without regard to the other criteria. Table 6 summarizes the results from a single criterion calibration based on volume. Solution 6 produced the minimum volume error (4.0%) and an acceptable CPI/vehicle error of 17.9%. However, for this solution the speed error was unacceptably high at 77.9%. For the CPI-based calibration given in Table 7, a minimum error of 6.6% was obtained. The volume error was acceptable at 4.0%, but the speed error proved to be unacceptably high at 78.2%.

None of these parameter sets were found to be acceptable for use in a road safety study. While some solutions sets yielded acceptable CPI/vehicle error this was achieved at the expense of unacceptably high traffic related errors. Since CPI/vehicle is a function of these traffic attributes this creates validity issues related to the application of this type of simulation platform in road safety studies.

A number of researchers have attempted to overcome the single-criteria calibration problem by adopting a 'multi-criteria' approach, whereby all individual attribute errors are combined into a into a single fitness criterion, which is then minimized. Table 8 summarizes the results of the calibration based on a weighted summation approach.

Table 8 DDS results using RMSPE summation

Trial	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reduction factor	Speed (km/h)	Volume	CPI	RMSPE Speed	RMSPE Volume	RMSPE CPI	**SUM of RMSPEs**
1	143.31	1.01	0.56	-5.26	0.14	-0.42	0.35	102.9	2069	931,768	0.765	0.0390	0.052	0.856
2	143.31	1.01	0.67	-5.68	-0.38	-0.42	0.33	103.5	2066	598,078	0.775	0.0404	0.325	1.140
3	143.31	1.01	0.56	-8.53	0.14	-0.25	0.20	104.8	2069	151,817	0.797	0.0390	0.829	1.665
4	100.31	1.10	0.56	-5.26	0.17	-0.95	0.35	104.7	2068	309,076	0.796	0.0395	0.651	1.486
5	176.98	0.86	0.56	-4.87	0.69	-0.53	0.43	103.0	2068	1,264,603	0.767	0.0395	0.428	1.234
6	164.55	1.05	0.56	-4.00	0.14	-0.42	0.37	104.1	2068	318,775	0.785	0.0395	0.640	1.465
7	189.32	1.11	0.59	-5.26	0.22	-0.42	0.59	102.4	2067	1,697,217	0.756	0.0399	0.917	1.713
8	197.39	0.65	0.70	-5.26	0.14	-0.42	0.35	102.9	2066	866,013	0.765	0.0404	0.022	0.827
9	197.39	1.36	0.70	-5.26	0.28	-0.25	0.35	101.9	2068	1,093,930	0.748	0.0395	0.236	1.023
10	211.12	0.65	0.70	-4.64	0.32	-0.42	0.35	104.1	2067	28,302	0.785	0.0399	0.968	1.793
11	260.95	1.35	0.70	-7.45	0.14	-0.42	0.35	103.0	2064	654,838	0.767	0.0413	0.260	1.068
12	**187.97**	**1.08**	**0.70**	**-9.03**	**0.14**	**-0.42**	**0.35**	**102.4**	**2067**	**870,483**	**0.756**	**0.0399**	**0.017**	**0.813**
13	225.70	1.23	0.70	-9.03	0.14	-0.64	0.47	102.6	2066	864,907	0.760	0.0404	0.023	0.823
14	239.38	1.08	0.61	-9.03	0.14	-0.42	0.35	102.0	2068	1,445,721	0.749	0.0395	0.633	1.422
15	246.77	1.08	0.70	-9.03	0.14	-0.25	0.37	101.8	2068	806,006	0.746	0.0395	0.090	0.875
16	185.56	1.91	0.70	-9.03	0.14	-1.61	0.35	101.5	2068	1,015,698	0.741	0.0395	0.147	0.928
17	91.70	1.08	0.70	-9.03	0.14	-0.25	0.35	104.3	2065	942,723	0.789	0.0409	0.065	0.895
18	254.36	1.08	0.70	-6.77	0.14	-0.25	0.30	101.0	2068	1,781,189	0.732	0.0395	1.012	1.784
19	239.30	1.08	0.70	-11.23	0.50	-0.42	0.23	102.5	2068	711,718	0.758	0.0395	0.196	0.994
20	234.31	1.40	1.24	-11.66	0.14	-1.16	0.35	100.0	2067	169,056	0.715	0.0399	0.809	1.564
Defaults	250.00	1.50	0.90	-8.00	0.35	-0.50	0.60	104	1992	539,547	0.787	0.0748	0.391	1.253
Observed								**58**	**2153**	**885,402**				

The weighted summation calibration also fails to provide parameter sets that are acceptable for use in road safety analysis. The CPI/vehicle error produces a disproportionate impact on the calibration exercise. This means that the parameter search acts similar to a single criteria CPI/vehicle calibration and there is little sensitivity to both speed and volume error. In practice, different weights have to be attributed to the various criteria in order overcome this issue. However, there is no conclusive evidence as to what these weights should be. Extra data would be needed to calibrate these weights accurately and objectively. Another problem that arises is that the weighted summation calibration can become trapped in the local minima that can differ from the true minima, as is the case in Table 8 where default parameter values actually result in errors that are lower than the other solutions. This is because the weighted summation method, using GA or DDS, archives only one or two best solutions from the previous iteration.

The multi-criteria procedure suggested in this paper makes use of non-dominance to replace a rather arbitrary summation error function. One of these multi-criteria procedures is the Pareto Archive Dynamically Dimensioned Search algorithm (PA-DDS). The PA-DDS procedure discussed in this paper makes use of three RMSPE fitness measures based on speed, volume and CPI/vehicle. Table 9 shows the non-dominated solutions obtained from applying the PA-DDS algorithm to the NG-SIM dataset for VISSIM.

Table 9 Pareto set of solutions (non-dominated solutions)

Pareto Solution Number	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reduction factor	Speed (km/h)	Volume	CPI	**RMSPE Speed**	**RMSPE Volume**	**RMSPE CPI**	SUM
1	240.15	3.00	1.50	-4.00	2.00	-0.25	0.80	76.9	1996	876,037	0.319	0.073	0.011	0.402
2	223.57	2.32	1.55	-7.31	0.63	-1.77	0.53	80.4	2051	874,467	0.3790	0.0474	0.0124	0.4387
3	239.86	3.00	1.49	-11.11	1.01	-1.38	0.80	78.2	2016	691,229	0.3412	0.0636	0.2193	0.6242
4	206.80	3.00	1.62	-13.54	1.62	-1.43	0.79	74.2	1956	618,912	0.2726	0.0915	0.3010	0.6651
5	223.57	2.32	1.55	-7.31	0.78	-1.77	0.53	79.8	2036	661,708	0.3687	0.0543	0.2526	0.6757
6	209.47	2.21	1.47	-6.11	0.63	-1.72	0.53	84.6	2062	649,650	0.4510	0.0423	0.2663	0.7595
7	209.07	2.19	1.65	-6.77	0.62	-1.63	0.30	73.1	2035	477,015	0.2538	0.0548	0.4612	0.7698
8	129.97	0.94	0.65	-8.10	0.48	-1.42	0.71	102.1	2067	824,496	0.7512	0.0399	0.0688	0.8599
9	209.47	2.21	1.34	-11.44	0.60	-1.72	0.61	92.4	2064	669,973	0.5848	0.0413	0.2433	0.8694
10	140.05	1.22	0.77	-4.20	0.19	-1.34	0.58	102.5	2067	950,139	0.7580	0.0399	0.0731	0.8711
11	209.47	1.97	1.72	-14.03	0.59	-2.03	0.60	72.6	1962	402,531	0.2452	0.0887	0.5454	0.8793
12	129.97	1.96	0.50	-7.62	0.62	-1.40	0.56	102.9	2069	958,357	0.7649	0.0390	0.0824	0.8863
13	185.64	1.22	0.50	-4.20	0.15	-1.35	0.72	103.9	2068	826,564	0.7820	0.0395	0.0665	0.8880
14	196.64	2.44	1.72	-7.05	0.38	-1.89	0.53	72.9	1960	360,669	0.2503	0.0896	0.5926	0.9326
15	138.90	1.70	0.54	-5.05	0.15	-1.69	0.80	102.6	2068	1,027,223	0.7597	0.0395	0.1602	0.9594
16	190.11	2.55	1.72	-10.74	0.59	-1.91	0.53	71.1	1958	302,332	0.2195	0.0906	0.6585	0.9686
17	151.44	1.81	1.08	-10.37	0.48	-2.19	0.65	95.9	2065	610,042	0.6448	0.0409	0.3110	0.9967
18	271.44	2.51	1.55	-7.31	0.10	-1.76	0.53	79.2	2047	327,290	0.3584	0.0492	0.6303	1.0380
19	196.64	2.62	1.72	-10.74	0.38	-1.91	0.53	70.7	1954	222,870	0.2126	0.0924	0.7483	1.0533
20	223.59	1.22	0.86	-4.20	0.15	-1.35	0.67	99.8	2069	1,176,434	0.7117	0.0390	0.3287	1.0794
21	166.31	1.52	0.50	-7.62	0.48	-1.42	0.70	102.2	2069	1,158,228	0.7529	0.0390	0.3081	1.1000
22	209.47	2.51	1.55	-7.31	0.63	-1.77	0.53	78.3	2043	235,803	0.3430	0.0511	0.7337	1.1277
23	196.64	2.44	1.72	-7.05	1.16	-1.89	0.34	70.9	1962	91,572	0.2160	0.0887	0.8966	1.2013
24	209.47	2.47	1.72	-10.74	0.59	-1.91	0.53	70.6	1963	66,869	0.2109	0.0882	0.9245	1.2236
25	239.86	3.00	1.49	-12.94	0.80	-1.71	0.53	84.6	2054	240,117	0.4510	0.0460	0.7288	1.2258
26	163.71	1.81	1.21	-11.44	0.48	-1.72	0.62	95.4	2065	373,982	0.6363	0.0409	0.5776	1.2547
27	135.18	1.12	0.93	-4.63	1.05	-1.28	0.55	99.6	2066	1,588,737	0.7083	0.0404	0.7944	1.5431
28	185.64	1.22	0.85	-4.18	0.15	-1.75	0.80	97.1	2066	1,946,692	0.6654	0.0404	1.1987	1.9045
29	185.64	1.22	0.50	-4.20	0.37	-1.35	0.72	102.0	2071	2,506,836	0.7495	0.0381	1.8313	2.6188
30	248.46	1.35	0.50	-6.40	-0.14	-1.77	0.72	99.7	2069	4,737,069	0.7100	0.0390	4.3502	5.0992
31	244.45	2.27	0.50	-4.20	0.87	-1.12	0.80	95.1	2068	8,999,845	0.6311	0.0395	9.1647	9.8353

The PA-DDS seems to overcome many of the issue associated with the weighted summation approach through the use of 'trade-offs' or non-dominance. There are no weights needed since the algorithm will allow some criteria to become worse in order to improve other criteria. Since more solutions are retained the chance of being stuck in local minima is reduced. The set of non-dominated solutions is retained to allow for the random sampling within this set.

The results of the PA-DDS application are summarized in Table 9. This application yield two possible solution that minimize the sum of the errors while obtaining acceptable trade-offs for most constituent criteria (Solutions 1 and 2). For Solutions 1 and 2 the CPI/vehicle errors are minimized with very low speed errors (although not minimum) and reasonably low volume errors. Since volume does not vary significantly between simulations runs the results obtained from PA-DDS are acceptable from both a traffic and road safety analysis perspective.

A number of solutions in Table 9 were found to have unacceptably high fitness errors; in some cases exceeding 40%. These were subsequently removed from our set of non-dominate solutions. Five non-dominated solutions with acceptable fitness errors are summarized in Table 10. These solutions reflect the best estimate parameter values as determined by the PA-DDS application. While all these solutions are non-dominated and reflect good balance in fitness errors between speed, volume and CPI/vehicle, they do not necessarily result in the lowest overall model goodness-of-fit. Overall model goodness-of-fit can be expressed in terms of the sum of all the fitness criteria errors, with regard to tradeoffs. This defines a point with which we compare the errors associated with the five solutions shown in Table 10.

In the weighted summation method all the criteria must be in the same form or they cannot be summed. For the PA-DDS method the criteria do not have to be of the same form. This requires a standardization of the fitness errors when estimating the overall model goodness of fit, such that fitness errors have the same scale. These were expressed as the RMSPE error and summarised in Table 10 for the five non-dominated solutions.

Table 10: Acceptable non-dominated solutions

Pareto Solution Number	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reductio n factor	Speed (km/h)	Volume	CPI	**RMSPE Speed**	**RMSPE Volume**	**RMSPE CPI**	SUM
4	206.80	3.00	1.62	-13.54	1.62	-1.43	0.79	74.2	1956	618,912	0.2726	0.0915	0.3010	0.6651
1	**240.15**	**3.00**	**1.50**	**-4.00**	**2.00**	**-0.25**	**0.80**	**76.9**	**1996**	**876,037**	**0.319**	**0.073**	**0.011**	**0.402**
3	239.86	3.00	1.49	-11.11	1.01	-1.38	0.80	78.2	2016	691,229	0.3412	0.0636	0.2193	0.6242
5	223.57	2.32	1.55	-7.31	0.78	-1.77	0.53	79.8	2036	661,708	0.3687	0.0543	0.2526	0.6757
2	223.57	2.32	1.55	-7.31	0.63	-1.77	0.53	80.4	2051	874,467	0.3790	0.0474	0.0124	0.4387

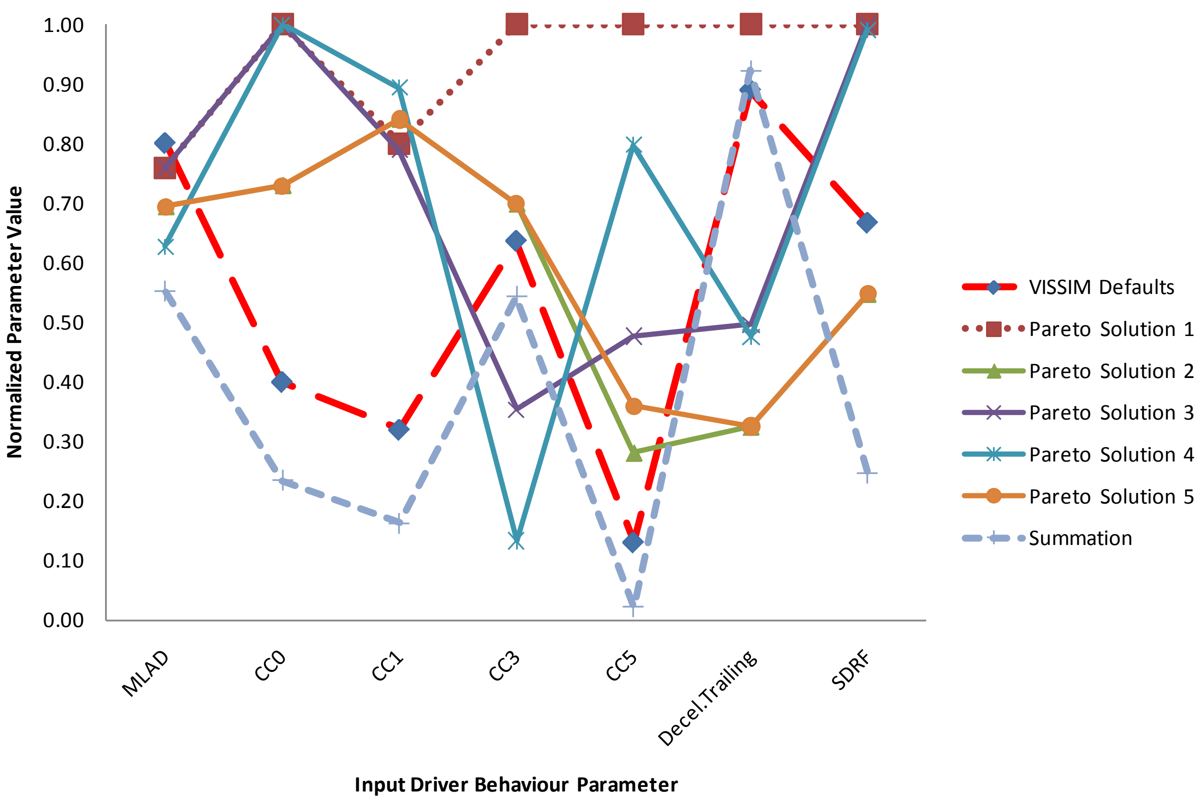

Figure 5: Sensitivity of VISSIM parameter values to calibration methods

The solution sets (parameter values) from Table 10 were plotted in Figure 5 along with default values and values obtained from the weighted summation method for speed, volume, and CPI/vehicle. It should be noted that individual parameter values need to be consider as constituents of the whole solution set suggested by each method with associated errors. The parameter values in Figure 5 have been normalized for comparison purposes.

Figure 5 suggests that there is fair amount of variation associated with the parameter values depending on which calibration method is used. This is especially true when comparing the weighted summation and default methods to the Pareto (non-dominated) methods. The default profile compare well with those suggested by the weighted summation method (for all seven parameters). The Pareto methods seem to yield some consistency in parameter values. However, the Pareto method that explicitly considers overall model goodness of fit (distance to the origin) seems to differ somewhat from the other Pareto solutions for a selected number of parameters. This result will require some investigation, however, it seems that the non-dominated solutions generated in this exercise while essentially on the same plane in terms of error they vary to a degree in parameter values. These results are consistent with observations made by Madsen (2000).

It should be note that Parameter Set 1 gives the best trade off in fitness error while at the same time minimizing the overall fitness error for the model. As such, the parameter values associated with this solution set suggest the best estimate parameter values for simulation. Solution Set 1 results in the lowest CPI/vehicle error, but yields higher speed and volume errors with respect to the minima for these solutions. The choice of whether to select Solution Set 1 depends on how tolerant we are on moderate errors in constituent traffic and safety attributes. In this case the compromise was not found to be particularly severe.

Figure 6 provides a three dimensional representation of criteria error for several calibration methods, the Pareto solutions in Table 10, as well as the single-criterion (speed, volume, and CPI/vehicle), weighted summation, and VISSIM defaults.

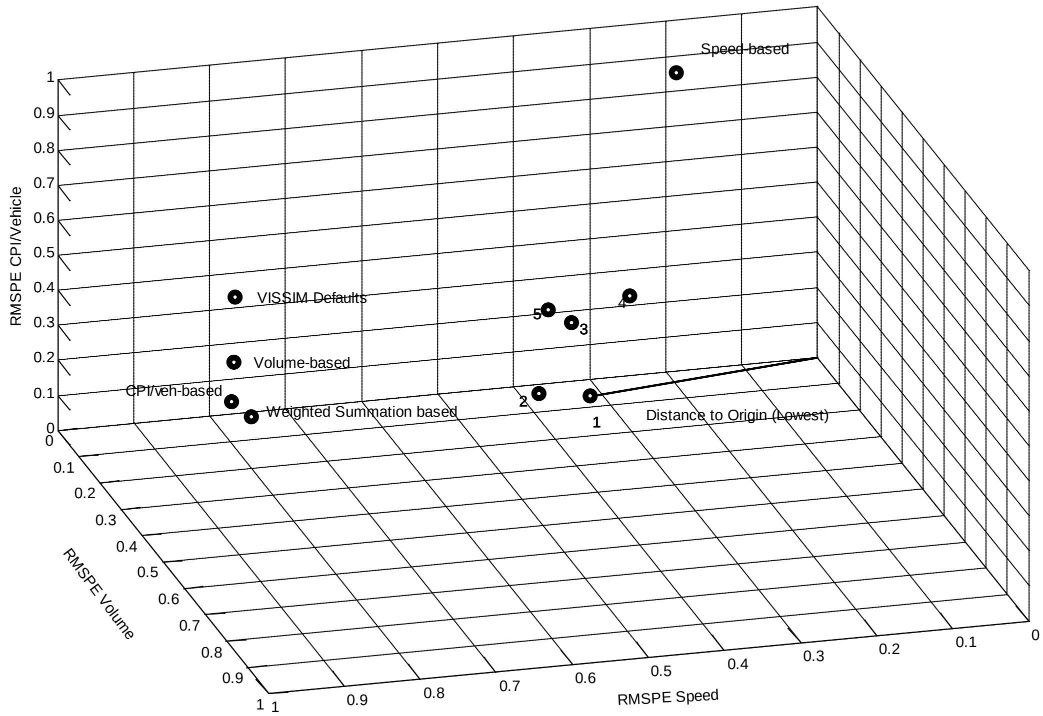

Figure 6: 3D Solution Space

The single-criteria methods are clearing in unacceptable region for other (non-calibrating) errors as well as overall model goodness of fit (distance to the origin). The same can be said for the weighted summation method and VISSIM defaults. The Pareto solutions tend to cluster nearer to each other close to the origin. Solution 1 in Figure 6 reflects the Pareto non-dominated solution set that minimizes the overall model goodness of fit.

VALIDATION RESULTS

The transferability of the results to a separate set of data needs to be ascertained if these results are generalized. A separate set of observed vehicle tracking data was extracted on which to validate the transferability of the parameter results obtained in Table 10. These data were also extracted from the FHWA (2007) NG-SIM program for Interstate Highway 101 and apply to a different time period (8:20 am to 8:35 am) than was used in the calibration exercise. For this paper we have assumed that the two NG-SIM datasets are independent.

Table 11 summaries the results of the validation test solution sets 1 and 2 from the PA-DDS calibration in the previous table.

Table 11 Validation errors versus errors from default parameters

Pareto Solution Number	(max) Look ahead Distance (m)	CC0	CC1	CC3	CC5	Accepted deceleration of trailing vehicle for lane change	Safety distance reductio n factor	Speed (km/h)	Volume (veh)	CPI	**RMSPE Speed**	**RMSPE Volume**	**RMSPE CPI**	SUM
1	240.15	3.00	1.50	-4.00	2.00	-0.25	0.80	64.7	1932	1,035,306	0.320	0.009	0.087	0.416
2	223.57	2.32	1.55	-7.31	0.63	-1.77	0.53	67.8	1968	808,162	0.384	0.028	0.152	0.563
Defaults	250.00	1.50	0.90	-8.00	0.35	-0.50	0.60	102.0	1891	793,907	1.082	0.013	0.167	1.261
Observed								**49.0**	**1915**	**952,591**				

The results suggest a good transferability of parameter values between the two datasets. The best estimate parameter values from Table 10 yielded errors shown in Table 11 for the validation dataset. We note that in both cases the best estimate parameter values produce errors that were acceptable for speed, volume and CPI/vehicle as well good overall model goodness of fit. For the validation dataset both PA-DDS solution sets indicate significant improvement in error when compared to default values.

CONCLUSIONS

This paper introduced the basic concepts of multi-criteria calibration based on non-dominance. The Pareto Archive Dynamically Dimensioned Search was demonstrated using three fitness criteria: speed, volume, and CPI/vehicle. The following observations were made from the case study analysis.

- Single-criterion calibration gives only good results for the criterion on which calibration is carried out, but other traffic and safety metrics may not be within acceptable ranges of error.
- The weighted summation method can become “stuck” in local minima since certain criteria may disproportionately govern the parameter search, and this suggests a need to calibrate the weights for the summation algorithm itself. The results presented in this paper for the weighted summation method assume a value of one for all the criteria. Using these weights, the weighted summation method yielded parameter values and fitness errors that differed significantly from the Pareto non-dominated methods.
- The PA-DDS algorithms consider trade-offs (non-dominance) and do not need the specification of weights in an overall error function. It provides a set of solutions (parameter values) that yield acceptable though not necessarily minimal fitness errors for each of the criteria. In this paper three criteria were used: speed, volume, and CPI/vehicle. This should not be viewed as a limitation of the approach since any number of criteria and different fitness functions can be selected.
- The application of the PA-DDS algorithm used a RMSPE fitness function for each criterion. This function produces dimensionless errors terms that can be comparable amongst the various criteria used. As such RMSPE provides an objective bases for comparing the different calibration approaches.
- The overall best parameter set was found to be the one with the lowest least squares summation of the errors from the Pareto non dominated set of solutions. The resultant parameter values were found to differ significantly from those suggested by defaults and the weighted summation method.

This paper has demonstrated that microscopic simulation platforms need to be calibrated using a multi-criteria approach that considers tradeoffs explicitly. This is especially true when the simulation platform is applied as a road safety assessment tool. Without careful consideration of the multi-faceted nature of the transportation problem, the validity of safety performance outputs derived from simulation can be suspect. Other applications of simulation platforms, such as capacity estimation, traffic operations and vehicle emission estimation, suffer from the very same problem wherein they are affected by interactions of other traffic factors. Pareto archival approaches offer a method to calibrate simulation platforms that can consider accuracy in a number of fundamental traffic metrics (e.g. speed, volume, density) as well safety performance (in this study, CPI/vehicle) and other metrics (e.g. capacity and emissions).

APPENDIX - PA-DDS PSEUDO CODE

The following is the pseudo code for the PA-DDS Algorithm as described by Asadzadeh and Tolson (2009):

Step 0 – Define the measures of performances, n objectives
Step 1 – Optimize each measure of performance using a portion of the computational budget (e.g. in this case minimize each objective)

- Use DDS to optimize each objective using n trials
- Sort the resultant trials into a non-dominated set called the 'external set' using the 'fast non-dominated sort' algorithm developed by Deb et al (2000)

Step 2 – Select a 'current' solution, $x_{current}$, from the external set

- Calculate crowding distance as proposed by Deb et al (2000)
- Selection based on roulette wheel with emphasis on picking solutions from less crowded regions

Step 3 – Sample one new solution and evaluate

- Generate a new solution, x_{new}, by perturbing the current solution as defined in the original DDS algorithm developed by Tolson and Shoemaker (2007)
- Check the dominance of x_{new} against the external set
- If computation budget is not exceeded
  - If X_{new} is non-dominated then Set $X_{current} = X_{new}$
  - Else, go back to Step 3
- Else, Stop

The DDS pseudo code is thus (Tolson and Shoemaker, 2007):

Step 1 – Define the DDS inputs:

- Neighbourhood perturbation size, r (0.2 is the default value)
- Iteration size, m
- The lower and upper bounds of the D parameters, $\mathbf{x}^{min}$ and $\mathbf{x}^{max}$
- Initial solution, $x^{o}=[x_1, ...,x_D]$

Step 2 – Set the counter i = 1, evaluate measure of performance F, $F^{best} = F(x^{o})$ and $x^{best} = x^{o}$

Step 3 – Randomly choose J of D parameters for inclusion in the neighbourhood set {N}. If {N} is empty then select one random parameter

Step 4 – For j = 1...J parameters in {N}, perturb x_j^{best} using the standard normal variable, N(0,1):

- $x_j^{new} = x_j^{best} + r(x_j^{max} - x_j^{min})*N(0,1)$
- If $x_j^{new} < x_j^{min}$ then $x_j^{new} = {}_{x}^{jmin} + (x_j^{min} - x_j^{new)}$
  - If $x_j^{new} > X_j^{max}$, set $x_j^{new} = x_j^{min}$
- If $x_j^{new} > x_j^{max}$ then $x_j^{new} = x_j^{max} - (x_j^{new} - x_j^{max})$
  - If $x_j^{new} < x_j^{min}$, set $x_j^{new} = x_j^{max}$

Step 5 – Evaluate new $F(x^{new})$ and update best solution if $F(x^{new}) \leq F_{best}$ then $F^{best} = F(x^{new})$ and $x^{best} = x^{new}$

Step 6 – Update iteration counter i = i + 1, stop if i = m, else go to Step 3

The pseudo code for the crowing distance assignment is shown below (Deb et al, 2000):

$l = \|I\|$	number of solutions in the archive
for each i, set $I_{i,distance} = 0$	initialize distance
for each objective m	
I=sort(I, m)	
$I_{1,\ distance} = I_{l,\ distance} = \infty$	boundary points are always selected
For i = 2 to (l – 1)	for all other points
$I_{i,\ distance} = I_{i,\ distance} + (I_{(i+1),\ m} - I_{(i-1),m})$	

Higher $I_{i,\ distance}$ means solutions are on less crowded regions of the solution space.

REFERENCES

Asadzadeh, M., and Tolson, B.A. (2009). A new multi-objective algorithm, pareto archive DDS. *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, pp. 1963-1966.

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., and Wen, Y. (2007). Calibration of microscopic traffic simulation models. *Transportation Research Record: No.1999*, pp. 198–207.

Cheng, C.T., Ou, C., and Chau, K.W. (2002). Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology*, Vol 268, Issue 1-4, pp. 72 86.

Cicu, F., Illotta, P.F., Bared J., and Isebrands, H. (2011) VISSIM calibration of roundabout traffic performance. Presented at 90th Annual Meeting of the Transportation Research Board,Washington, D.C.

Ciuffo, B.F., and Punzo, V. (2010). Verification of traffic micro-simulation model calibration procedures: Analysis of goodness-of-fit measures. *Proceedings of the Annual Transportation Research Board Conference*, Washington, D.C., Paper # 10-3229.

Ciuffo, B.F., and Punzo, V., and Torrieri, V. (2008). Comparison of simulation-based and model-based calibrations of traffic-flow microsimulation models. *Transportation Research Record, No.2088,* pp. 36-44.

Cunto, F., Saccomanno, F.F. (2008). Calibration and validation of simulated vehicle safety performance at signalized intersections. *Accident Analysis and Prevention*. Vol. 40, pp 1171-1179.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Parallel Problem Solving from Nature, PPSN VI.*, pp. 849-858.

Duong, D., Saccomanno, F., and Hellinga, B. (2010). Calibration of microscopic traffic model for simulating safety performance. *Proceedings of the Annual Transportation Research Board Conference*, Washington, D.C., Paper # 10-0858.

Federal Highway Administration (FHWA), *US Highway 101 Dataset*, Next Generation SIMulation Fact Sheet, FHWA-HRT-07-030, 2007, http://www.tfhrc.gov/about/07030.htm, Website accessed July 2007.

Fonseca, C.M., and Fleming, P.J. (1993). Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In Genetic Algorithms: Proceedings of the Fifth International Conference, San Mateo, CA.

Hourdakis, J., Michalopoulos, P.G., and Kottommannil, J. (2003). A practical procedure for calibrating microscopic traffic simulation models. Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, DC.

Huang, W., and Sun, J. (2009). A NSGA-II based parameter calibration algorithm for traffic microsimulation model. Proceedings of the IEEE Computer Society: 2009 International Conference on Measuring Technology and Mechatronics Automation.

Kim, S., Kim, W., and R, L.R. (2005). Calibration of micro-simulation models using non-parametric statistical techniques. *Transportation Research Record, No.1935*, pp. 111–119.

Knowles, J.D., and Corne, D.W. (2000). Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. Evolutionary Computation, Vol 8, Issue 2, pp. 149-172.

Koski, J. (1994). Multicriterion structural optimization. *Advances in Design Optimization*, pp. 194-224.

Ma, J., Dong, H., and Zhang, H.M. (2007). Calibration of microscopic with heuristic optimization methods. *Transportation Research Record*, *No.1999*, pp. 208–217.

Ma, T., and Abdulhai, B. (2002). Genetic algorithm-based optimization approach and generic tool for calibrating traffic microscopic simulation parameters. *Transportation Research Record*, *No.1800*, pp. 6–15.

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, Vol 235, pp. 276-288.

Park, B., and Qi, H. (2005). Development and evaluation of a procedure for the calibration of simulation models. *Journal of the Transportation Research Board, No.1934*, pp. 208–217.

PTV. (2008). VISSIM© 4.3 User Manual. Planung Transport Verkehr AG.

Sayed, T. and Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning and Technology*, Vol 22 pp. 309–323.

Shea, K., Sedgwick, A., and Antonuntto, G. (2006). Multicritia optimization of paneled building envelopes using ant colony optimization, Intelligent Computing in Engineering and Architecture, pp. 627 – 636.

Toledo, T., Ben-Akiva, M.E., Darda, D., Jha, M., and Koutsopoulos, H.N. (2004). Calibration of microscopic traffic simulation models with aggregate data. *Transportation Research Record, No.1864*, pp. 10–19.

Tolson, B.A., and Shoemaker, C.A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, Vol. 43(1), W01413, pp. 1-16.

Vaiana, R., and Gallelli V. (2011)The calibration of traffic microscopic simulation models: Kinematical approach to the through movement on roundabouts. Presented at 90th Annual Meeting of the Transportation Research Board,Washington, D.C.

Yapo, P.O., Gupta, H.V., and Sorooshian, S. (1998). Multi-objective global optimization for hydrological models. *Journal of Hydrology*, Vol 204, Issue 1-4, pp. 83 97.