https://dx.doi.org/10.1016/j.trc.2015.12.006

A cross-entropy method and probabilistic sensitivity analysis framework for calibrating microscopic traffic models

R. X. ZHONG[a,b], K. Y. FU[a,b], A. SUMALEE[c], D. NGODUY[d,*], W.H.K. LAM[e]

[a]*School of Engineering, Sun Yat-Sen University, Guangzhou, China.*
[b]*Guangdong Provincial Key Laboratory of Intelligent Transportation Systems, Guangzhou, China.*
[c]*Department of Civil Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*
[d]*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom.*
[e]*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China.*

Abstract

Car following modeling framework seeks for a more realistic representation of car following behavior in complex driving situations to improve traffic safety and to better understand several puzzling traffic flow phenomena, such as stop-and-go oscillations. Calibration and validation techniques pave the way towards the descriptive power of car-following models and their applicability for analyzing traffic flow. However, calibrating these models is never a trivial task. This is caused by the fact that some parameters, such as reaction time, are generally not directly observable from traffic data. On the other hand, traffic data might be subject to various errors and noises. This contribution puts forward a Cross-Entropy Method (CEM) based approach to identify parameters of deterministic car-following models under noisy data by formulating it as a stochastic optimization problem. This approach allows for statistical analysis of the parameter estimations. Another challenge arising in the calibration of car following models concerns the selection of the most important parameters. This paper introduces a relative entropy based Probabilistic Sensitivity Analysis (PSA) algorithm to identify the important parameters so as to reduce the complexity, data requirement and computational effort of the calibration process. Since the CEM and the PSA are based on the Kullback-Leibler (K-L) distance, they can be simultaneously integrated into a unified framework to further reduce the computational burden. The proposed framework is applied to calibrate the intelligent driving model using vehicle trajectories data from the NGSIM project. Results confirm the great potential of this approach.

Keywords: Car-following model, Model calibration, Cross-entropy method, Probabilistic sensitivity analysis, Relative entropy, Kullback-Leibler distance

1. Introduction

For most planning and operational applications, accurate representation of realistic driving behaviors offers a great help to transportation analysts. Along this stream, microscopic traffic models especially Car Following (CF) models are widely adopted to simulate complex traffic scenarios such as traffic incident,

*Corresponding author. Tel: +44 (0)113 34 35345.
Email addresses: zhrenxin@mail.sysu.edu.cn (R. X. ZHONG), asumalee@gmail.com (A. SUMALEE), D.Ngoduy@leeds.ac.uk (D. NGODUY), cehklam@polyu.edu.hk (W.H.K. LAM)

signal control, public transport priority wherein analytical methods are unlikely to work due to the complexity. A large number of CF models have been developed to describe CF behavior under a wide range of traffic conditions in the past decades, see e.g. Barceló (2010); Saifuzzaman and Zheng (2014) for a review. Model calibration would heavily affect the reliability of the results achieved by using the underlying model as well as its applicability in traffic engineering practice. Calibration is vital for microscopic traffic models, yet it can be rather difficult since these models often contain a wide range of variables. Moreover, some parameters, such as reaction time, desired spacing, desired time headway, desired speed, are generally not directly observable from traffic data, and this makes them hard to be identified. Nonetheless, the parameters are scenario specific, i.e. they are not transferable to other situations (different locations, periods of the day such as morning rush hours and evening rush hours, etc.). On the other hand, driving behavior and local traffic rules, which the microscopic traffic flow models intend to describe, are variable in time and space, etc. (Hoogendoorn and Hoogendoorn, 2010; Treiber et al., 2014). Therefore, many of the microscopic traffic models were neither empirically calibrated nor validated using real traffic data until recently, a considerable amount of research have been devoted to the calibration and validation of microscopic models, see e.g., Ossen and Hoogendoorn (2008); Kesting and Treiber (2008); Hoogendoorn and Hoogendoorn (2010); Treiber et al. (2014). However, guidance on the systematic and rigorous calibration and validation of traffic flow models is still lacking (Ngoduy and Maher, 2012; Saifuzzaman and Zheng, 2014). Nevertheless, traffic data might be subject to various errors and noises. The calibration of stochastic microscopic models is more complicated as they will require multiple runs to reduce the noise in the objective function (Ngoduy and Maher, 2012). To tackle these difficulties, this paper aims to develop a new framework for calibrating microscopic traffic models with various uncertainties to maximize the model's descriptive power based on representative traffic data.

Conventionally, deterministic search methods which aim to minimize the discrepancy between the model prediction and observed data are common approaches to access model calibration for both microscopic and macroscopic traffic models. As a consensus in the literature, such kind of methods will result in a large number of local optima due to the complex structure of the optimization problem and different combinations of the set of parameters, see e.g., Ngoduy and Maher (2012); Ciuffo and Punzo (2014); Zhong et al. (2015). For this reason, gradient-based solution algorithms are not considered as a popular option (Kontorinaki et al., 2015). Therefore, a popular and convenient approach to compensate this is to use random search techniques (Ciuffo and Punzo, 2014; Hale et al., 2015). The basic idea behind such methods is to systematically partition the feasible region into smaller subregions and then to move from one subregion to another based on information obtained by random search (Ciuffo and Punzo, 2014; Ngoduy and Maher, 2012). Well-known examples include simulated annealing, genetic algorithms, tabu search, and ant colony methods. All these methods are reported to find a good local optimal solution (while some also claimed global optimal solution can be obtained) but there is not as yet a fully accepted method. Noticing the optimization nature of the calibration problem, remarkably, Ciuffo and Punzo (2014) applied the No Free Lunch (NFL) theorems to the calibration problem of microscopic traffic models to access the performance of various algorithms ranging from heuristic optimization methods to meta heuristic searching methods. To be more specific, the simultaneous perturbation stochastic approximation method, simulated annealing, Genetic Algorithm (GA), and OptQuest/Multistart heuristic methods are evaluated and compared in Ciuffo and Punzo (2014). It is found that the performance of different algorithms over the 42 experiments considerably differ, which confirms the validity of NFL theorems in the calibration problem, i.e., the dependence of the performance from the parameters to be calibrated, from the GoF measure, and from the quality of the data. Remarkably, the analysis reveals that the GA, which is probably the most widely used algorithm type for the calibration of microscopic traffic models, outperforms the others globally while the OptQuest results

the best algorithm in the sense of convergent time and optimization performance indicator. As a contrast, the simulated annealing results the worst algorithm for the tested cases in Ciuffo and Punzo (2014).

It is known that the GA is generally computationally intensive and of no convergence proof. Moreover, the presence of random noise would affect the optimization procedures. Ngoduy and Maher (2012) applied the Cross Entropy Method (CEM), which is a generic monte carlo technique with importance sampling for reducing the computational burden, to the calibration purposes of a second order macroscopic model. The empirical results have verified several merits of such method, e.g. attaining a set of parameters that are close to the global known optimal set, computational efficiency and convergence. Maher et al. (2013) extended this CEM framework for signal optimization to consider the effect of noise in the evaluation process.

Noting that single data source may be not sufficient for calibrating a microscopic traffic model, Hollander et al. (2008) summarized an exhaustive list of both Goodness-of-Fit (GoF) measures and optimization algorithms used in the calibration of microscopic traffic models with some tips for their usage. However, a thorough investigation still lacks while this has been further pursued by Ciuffo and Punzo (2014). To compensate the drawbacks of a single data source, Hoogendoorn and Hoogendoorn (2010) proposed a data fusing framework to calibrate the parameters of CF models based on several data sources. Treiber et al. (2014) also applied several (GoF) measures to assess descriptive quality of CF models and provided comparative calibration results using different optimization methods and data sources under the umbrella of Maximum-Likelihood and least squared errors.

Because of the unobservable parameters of the CF models, inherent noise of traffic data, complicated human factors to be modeled and traffic scenarios to be simulated, one of the main challenges arising in the calibration process concerns the selection of the most important parameters and the identification of their probabilistic characteristics. Selecting the most important parameters also helps guiding data collection by limiting the number of input parameters to be observed rather than the whole parameter set (Saifuzzaman and Zheng, 2014). However, there is usually no formal procedure for selecting these parameters other than the experiences of the participating engineers. Such a prior selection of relative parameters based on a priori knowledge or common sense could be harmful because many interactions among parameters remain hidden. To avoid the potential negative effects caused by subjectively chosen set of calibration parameters, a Sensitivity Analysis (SA) of the parameters is essentially required. SA is a procedure to explore the relationship between the simulation output and input parameters considering the potential uncertainty (including noise, un-modeled characteristics and parameter variations etc.) on model response. SA is also a crucial procedure to individuate the most important sources of modeling uncertainties. Despite of its importance, a proper SA for traffic models is barely performed in common practice (Ciuffo and Azevedo, 2014; Ge et al., 2014).

There are several methods to carry out SA in the literature. Among them, the One-at-A-Time (OAT) approach and the variance based approach are the two most commonly used approaches for SA in microscopic traffic simulation. The OAT measures are based on the estimation of partial derivatives to assess how uncertainty in one factor would affect the model output by fixing the other factors to their nominal values. The OAT approach has been applied to identify the important parameters in VISSIM and to get insight on the meaning of the values of parameters resulting from the calibration of the intelligent driving and the velocity difference models (Lownes and Machemehl, 2006; Kesting and Treiber, 2008; Mathew and Radhakrishnan, 2010; Asamer et al., 2013). The drawbacks of this method include the locality, i.e. fixing some factors to their nominal values rather than exploring the whole input space, and thus the interactions among factors cannot be detected. Another commonly adopted SA approach is the ANalysis Of VAriance (ANOVA) which is able to quantify the uncertainty of a model and to capture the interactions among different factors. A more efficient SA method based on the decomposition of the variance (see Saltelli et al. (2008) for technical details) has been adopted to the SA of car-following models (Punzo and Ciuffo, 2009).

The major drawback of such approach is its computational burden which often requires a large number of model evaluations.

To combine the calibration and SA into a unified framework, this paper puts forth a Cross-Entropy Method (CEM) (Rubinstein, 1997; De Boer et al., 2005; Kroese et al., 2013) and the Probabilistic Sensitivity Analysis (PSA) based approach for simultaneous model calibration and identification of important parameters. The CEM is able to find a set of parameters that are close to the global optimal solution set for the calibration problem. Meanwhile, the PSA could reveal the probabilistic behavior of a model's response with respect to its parameter uncertainties. Results from PSA can be used to assist engineering design from various aspects, e.g. to help reducing the dimension of a design problem by identifying the significant factors (Liu et al., 2006). Since this method is based on the relative entropy, it better fits into the framework of the CEM.

The rest of this paper is organized as follows. Section 2 presents a brief description of the intelligent driving model. Section 3 formulates the calibration problem through the lens of CEM. Section 4 further introduces the probabilistic sensitivity analysis and puts the optimization problem and the SA into a unified framework. Empirical studies are reported in Section 5. Finally, conclusions and future works are given in Section 6.

2. The intelligent driving model

The CF models aim to describe the movements of individual vehicle-driver combinations based on the stimulus-response framework. Conventional engineering car following models mainly consider the longitudinal interactions of vehicles on the road, e.g., vehicle acceleration towards a desired velocity (in the free-flow region), a braking strategy when approaching other vehicles or obstacles, maintaining speed, distance-keeping relative to leading vehicles. Thus, the simulation of car following behaviors is mainly determined by simple physical laws in conjunction with several parameters such as headway, gap, speed and acceleration of the leading and surrounding vehicles.

The Intelligent Driving Model (IDM) proposed by Treiber et al. (2000) is one of the widely-applied CF models. This model considers both the desired velocity and the desired space headway to model the acceleration/deceleration strategy of subject vehicle with respect to the leading vehicle,

$$\dot{v}_f\left(v_f, s, s^*(v_f, \Delta v)\right) = a\left[1 - \left(\frac{v_f}{v_0}\right)^4 - \left(\frac{s^*(v_f, \Delta v)}{s}\right)^2\right], \tag{1}$$

where a is the the maximum acceleration/deceleration of the subject vehicle, v_0 the desired velocity, s the spacing between two vehicles measured from the front edge of the subject vehicle to the rear end of the leading vehicle, and $s^*(\cdot)$ the desired space headway. When leading vehicle is far away, the third term in this equation becomes negligible small and the model performs as a free flow model where the desired velocity of the driver governs the acceleration, i.e., $\dot{v}_{free} = a\left[1 - (v_f/v_0)^4\right]$. When the subject vehicle approaches the leading vehicle, the braking strategy $\dot{v}_{brake}(s, v_f, \Delta v) = -a(s^*/s)^2$ will be dominant to ensure s to approach the desired space headway s^*, which depends on several factors: speed, speed difference (Δv), the maximum acceleration a, a comfortable deceleration (or the desired deceleration which will be active in non-stationary traffic) b, the minimum distance at the standstill situation s_0, and the desired time gap T. To be specific,

$$s^*(v_f, \Delta v) = s_0 + v_f T - \frac{v_f \Delta v}{2\sqrt{ab}}. \tag{2}$$

Remarkably, the deceleration of the IDM will be quite strong if the current gap becomes too small to ensure collision-free. Reaction time is ignored in this model.

It is worth noticing that parameters of IDM including the maximum acceleration a, the desired deceleration b, the desired velocity v_0, the desired time gap T, and the minimum distance s_0, are related to different traffic conditions. For example, the maximum acceleration a and the desired deceleration b are related to stop-and-go traffic flow while the desired time gap T is mainly involved in the steady-state car-following period, the desired velocity v_0 is observed in free-flow traffic condition while the creeping and standing traffic situation is crucial for the identification of the minimum distance s_0. In other words, the data sources used to calibrate IDM are suggested to contain all traffic regimes mentioned above.

Some parameters of the CF models, such as the desired time gap, desired space headway, desired deceleration etc., are subject to human factors, which mainly features in the inter-driver and intra-driver heterogeneity. Generally, the inter-driver heterogeneity implies that different drivers behave in different ways even if they follow the same vehicle due to individual driving habits, while the intra-driver heterogeneity indicates that a driver shows differential response to the same change of driving situation at different time or under different conditions (Treiber et al., 2013). Inter-driver heterogeneity is easily explained by the traffic oscillations caused by the aggressive and timid driver behaviors (Laval et al., 2010). The intra-driver variability accounts for a large part of the deviations between simulations and empirical observations (Kesting et al., 2009). As indicated in Wagner (2012) that most fluctuations in time headways can be attributed to individual drivers adopting new preferred headways. Kim et al. (2008) found how heterogeneity in the driver population can be used to explain the observed scatter in flow-density plots. Several recent studies, e.g., Ossen et al. (2005); Duret et al. (2008); Chiabaut et al. (2010); Saifuzzaman and Zheng (2014), argued that the parameters of CF models differ from drivers because drivers are different so as their driving styles and risk-taking capabilities. All these suggest that the randomness in the CF models can be interpreted as uncertainty in the parameters (Chiabaut et al., 2010). Furthermore, the desired parameters such as desired time headway, desired velocity are generally unobservable in nature, which renders the parameter estimation problem more challenging (Saifuzzaman and Zheng, 2014). Therefore, it is necessary to consider such randomness in the calibration process.

3. Problem formulation

3.1. Mathematical Setting for calibration of microscopic traffic flow models

From the IDM example, we can see that the car following models are usually denoted by ordinary equations. Typically, a system of differential equations of the following form is used to model microscopic traffic behavior.

$$\begin{aligned} \dot{v}_f &= \tilde{g}(h, v_f, v_l, X), \\ \dot{h} &= v_l - v_f, \end{aligned} \tag{3}$$

where we denote the velocities of the following and the leading vehicle as v_f and v_l, the headway between them as h, respectively. The symbol $X = (X_1, X_2, \cdots, X_n)$ denotes a set of parameters of the model under consideration, and n denotes the dimension of parameter set. If we regard the above differential equations as a dynamical system and define the system state as $y = col(h, v_f, v_l)$, we can represent the above system in a compact form as

$$\dot{y} = g(y, X, t), \tag{4}$$

with initial condition $y(t_0) = y_0$. Then the complete set of parameters Θ of the model is defined by the vector $\Theta = col(X, y_0)$. Let $\bar{Y}_i,\ i = 1, \cdots, N$, denote a set of observed data vectors, where N represents the

total amount of data. Let Y_i denotes the "true" solution of the state at time t_i from Equation (4) with exact parameters Θ. We assume that each component of the data $\bar{Y}_{ij}$ satisfies the following relation

$$\bar{Y}_{ij} = Y_{ij} + \sigma_j \epsilon_{ij}, \tag{5}$$

where $\sigma_j > 0$ measures the standard deviation of the noise associated with the j^{th} component of Y which is related to the scale of the expected magnitude of the j^{th} component, i.e., $\|y_j(t)\|$. Like conventional settings, all ϵ_{ij} are independent and standard Gaussian distributed random variables. The objective is to estimate the initial state y_0 and parameters X from the measurements. The principle of maximum-likelihood yields an appropriate cost function which measures the degeneracy between the model generated data and the measured data.

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{d} \frac{\left(Y_{ij}(\Theta) - \bar{Y}_{ij}\right)^2}{2\sigma_j^2}, \tag{6}$$

and we seek a Θ^* that satisfies:

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}(\Theta), \tag{7}$$

where d denotes the dimension of the state. Such kind of method is termed as global least-squared errors calibration in Treiber et al. (2014). Minimizing the objective function reduces the calibration problem to a multivariate nonlinear optimization problem whose solution may be not unique.

3.2. The cross-entropy method

As discussed in the introduction, a cross-entropy method based approach is proposed to solve the calibration problem. The CEM approach can be broken down into two key steps:

1. Generate a number of trial parameter sets randomly according to the chosen distributions.
2. Based on the values of the objective function associated with each trial parameter set, update the probability distribution used to generate the random trial sets according to the principle of "importance sampling".

For a general optimization problem, obtaining an (global) optimum solution can be regarded as a rare event. The CEM is a general monte carlo approach to solve rare event probability estimation problems. In this sense, we can reformulate an optimization problem in terms of cross entropy method as follows. Without loss of generality, consider the following minimization problem:

$$\gamma^* = \min_{x \in \chi} S(x), \tag{8}$$

where γ^* represents the minimum of $S(x)$ and x is defined in a function space χ. By the above analogy, the cross-entropy method may first formulate a family of Probability Density Functions (PDF) distributed in χ, denoted by $f(x; v)$, parameterized by v. For the minimization of $S(x)$, by defining the minimum γ^* as a threshold (or some $\gamma \geq \gamma^*$ but sufficiently close to γ^*), we can define a rare event as $S(x) \leq \gamma$. To this end, we can define

$$\ell(\gamma) = P_u\left(S(X) \leq \gamma\right) = E_u\left(I_{\{S(X) \leq \gamma\}}\right), \tag{9}$$

where γ is a threshold and $X = (X_1, X_2, \cdots, X_n)$ is a random vector generated by PDF with parameter v set to u (i.e., a realization) in $f(x; v)$. P_u denotes the probability, E_u denotes the expectation, and $I(\cdot)$ is the

indicator function, i.e., $I_{\{S(X)\leq\gamma\}} = 1$ if and only if $S(X) \leq \gamma$ is true, 0 otherwise. By this, we convert the original optimization problem into a rare event probability estimation problem. Detailed routine for solving this estimation problem by the CEM is outlined in Appendix A. Here we would present the CEM algorithm directly as follows:

1. Set $\hat{v}_0$ to u, and initialize the iteration counter l to 1.
2. Generate a set of random samples $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}$ from $p(x; \hat{v}_{l-1})$. Calculate the objective function $S(\mathbf{x_i})$ for each sample, and then determine $\hat{\gamma}_l$ by the following equality,
$$\hat{\gamma}_l = S_{(\lceil(1-\rho)N\rceil)}, \tag{10}$$
where ρ is a small real number.
3. Using the same samples $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}$ to estimate parameter vector $\hat{v}_l^{new}$ by solving the equality below.
$$\hat{v}_l^{new} = \arg\max_v \frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\hat{\gamma}_l\}} W(\mathbf{x_i}; u, \hat{v}_{l-1}) \ln f(\mathbf{x_i}; v). \tag{11}$$
4. Let $0 \leq \beta \leq 1$ be a smoothing parameter, the parameter vector is updated according to the smoothed updating law:
$$\hat{v}_l = \beta\hat{v}_l^{new} + (1-\beta)\hat{v}_l$$
5. If the standard deviation of samples generated by $p(x; \hat{v}_l)$ is lower than ϵ, stop the algorithm and determine the minimum of $S(x)$:
$$\gamma^* = \hat{\gamma}_l. \tag{12}$$
Otherwise, increase the iteration counter l by 1, and return to Step 2.

4. Probabilistic sensitivity analysis

As previously discussed, the estimation of parameters of the car following models is subject to various issues such as observability, perturbations, uncertainty or errors. Hence their sensitivity analysis is a crucial question in determining which parameter directions are the most/least sensitive to perturbations, uncertainty and data errors. This is essential for the selection of the most important parameters and the identification of their probabilistic characteristics. One of the preferred methods for sensitivity analysis is based on functional ANalysis Of VAriance (ANOVA) decompositions (Saltelli et al., 2008). The uniqueness of the functional variance decomposition imposes a critical prerequisite for the conventional ANOVA that the inputs should not be correlated. Recently, this prerequisite was eliminated by a generalization of the variance-based sensitivity indices for the case of dependent variables (Kucherenko et al., 2012). While variance adequately quantifies the variability of distributions that are symmetrical and unimodal, it is well known that the variance based SA methods are not suitable to measure the dispersion of a variable with a heavy-tail or a multimodal distribution, or which contains some outliers. On the other hand, it is often the case (as it will be shown later in Section 5) that the Probability Density Functions (PDF) of interest is in general nonlinear and/or non-Gaussian, and not only in a few moments, due to the significance of rare/tail events. Noting that entropy is calculated directly from the probability distribution function and thus provides a more general measure of output variability. There is a broad recent literature relying on information theory tools, wherein the sensitivity analysis is conducted using the relative entropy and the Fisher information between PDFs, see e.g. Liu et al. (2006); Lüdtke et al. (2008); Majda and Gershgorin (2010, 2011); Komorowski et al. (2011); Pantazis and Katsoulakis (2013). Adopting the relative entropy approach for sensitivity analysis better fuses together the CEM into a unified framework.

4.1. Relative entropy and mutual information

The relative entropy (also known as Kullback-Leibler (K-L) distance) is often used to measure the dissimilarity of two probability distributions, e.g., $p(x)$ and $q(x)$, in information theory and machine learning communities. In line with the importance sampling adopted in the cross entropy method, we consider an unknown distribution $p(x)$, which is modelled using an approximating distribution $q(x)$. To be specific, the K-L distance of q from p, denoted $D_{KL}(p\|q)$, measures loss/change of information when $q(x)$ is used to approximate $p(x)$. In information theory, as explained in Bishop (2006), "if we use $q(x)$ to construct a coding scheme for the purpose of transmitting values of x to a receiver, then the average additional amount of information required to specify the value of x (assuming we choose an efficient coding scheme) as a result of using $q(x)$ instead of the true distribution $p(x)$ is given by the K-L distance." The relative entropy quantifies the role of initial conditions in the forecast of a future state of a system and measures the sensitivity of a system to external perturbations (Majda and Gershgorin, 2011). For discrete probability distributions p_k and q_k, the K-L distance of q from p is defined to be

$$D_{\mathrm{KL}}(p\|q) = \sum_{k=1}^{K} p_k \ln \frac{p_k}{q_k}.$$

For continuous distributions $p(x)$ and $q(x)$, the K-L distance is defined to be:

$$D_{\mathrm{KL}}(p\|q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} \,\mathrm{d}x.$$

The K-L distance has two attractive features as a metric for model fidelity (Majda and Wang, 2006):

1. It is always non-negative,
$$D_{\mathrm{KL}}(p\|q) \geq 0,$$
with $D_{\mathrm{KL}}(p\|q) = 0$ if and only if $p(x) = q(x)$ almost everywhere.
2. It is invariant under general (invertible) nonlinear changes of variables.

It can be shown that the K-L distance is connected to the cross entropy $H(p, q)$ (for the distributions p and q) through the following relationship:

$$H(p, q) = H(p) + D_{\mathrm{KL}}(p\|q),$$

where $H(p)$ is the entropy of p, which is defined as,

$$H(p) = -E(\ln(p)) = -\int p(x) \ln(p(x)) dx. \tag{13}$$

In the information theory, the mutual information of two random variables is a measure of the variables' mutual dependence. To be specific, it determines how similar the joint distribution $p(x, y)$ is to the product of factored marginal distribution $p(x)p(y)$. The K-L distance between the joint distribution and the product of the marginals can be used to measure how 'close' these two random variables being independent as:

$$I(X; Y) = D_{\mathrm{KL}}(p(x, y)\|p(x)\, p(y)) = -\int\int p(x, y) \ln\left(\frac{p(x)\, p(y)}{p(x, y)}\right) dx\, dy,$$

From the properties of the K-L distance, we see that $I(X, Y) \geq 0$ with equality if, and only if, X and Y are independent. Mutual information can be equivalently expressed as

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

By this, the mutual information can be regarded as the reduction in the uncertainty about X by virtue of being told the value of Y (or vice versa). From a Bayesian perspective, the mutual information represents the reduction in uncertainty about X as a consequence of obtaining a new observation of Y.

4.2. Relative entropy based probabilistic sensitivity analysis

Based on the relative entropy, Liu et al. (2006) proposed a new Probabilistic Sensitivity Analysis (PSA) approach to evaluate the impact of a random variable on the performance index function by measuring the K-L distance between two probability density functions of the performance index function, obtained before and after the variation reduction of the chosen random variable. This approach can be applied to access both global SA and local SA, i.e. in any interested partial range of a response distribution. This flexibility facilitates the usage of the relative entropy based PSA under various scenarios of design under uncertainty, e.g. reliability-based design, robust design, and utility optimization. It is claimed that the relative entropy based PSA can be applied to both the prior-design stage for variable screening when a design solution is yet identified and the post-design stage for uncertainty reduction after an optimal design has been determined (Liu et al., 2006). Lüdtke et al. (2008) conducted a more rigorous analysis on the relative entropy based sensitivity analysis approach and termed such approach as information-theoretic sensitivity analysis. The relative entropy based approach can also consider the case when input correlations exist (Lüdtke et al., 2008). This method has been recently extended to study the SA of complex stochastic processes, see e.g., Majda and Gershgorin (2010, 2011); Komorowski et al. (2011); Pantazis and Katsoulakis (2013).

Assume the model of interest admits a mapping of the the following form

$$Y = h(X_1, X_2, \ldots, X_n),$$

where $X_1, \cdots, X_n$ are the random inputs that refer to intrinsic model parameters of the car following models in this paper. We assume that the model output has a PDF of $f(y)$. Global sensitivity analysis aims to rank the inputs $X_1, \cdots, X_n$ according to the degree to which they influence the output, individually and conjointly. If one particular parameter X_i is fixed to its nominal value (or its mean value or one of its realizations/observations) say replacing it with $\bar{x}_i$, then the yielded PDF of Y is denoted as $g(y|\bar{x}_i)$. From the previous section, the K-L distance $D\left(g(y|\bar{x}_i)\|f(y)\right)$ measures the difference between two probability density functions of the output obtained before and after the variation reduction of the random variable X_i. Specifically,

$$D_{KL_{X_i}}\left(g(y|\bar{x}_i)\|f(y)\right) = \int_{-\infty}^{\infty} g(y|\bar{x}_i) \times \ln \frac{g(y|\bar{x}_i)}{f(y)} dy. \tag{14}$$

Equation (14) quantifies the change of the PDF of the model output after eliminating the variability in X_i. Therefore, the K-L distance $D_{KL_{X_i}}\left(g(y|\bar{x}_i)\|f(y)\right)$ is regarded as the total sensitivity index of X_i. The larger value $D_{KL_{X_i}}\left(g(y|\bar{x}_i)\|f(y)\right)$ is, the more important X_i is (which implies that Y is more sensitive to X_i). Similarly, the combined effect of all random variables except X_i can be obtained by fixing the complementary set of X_i to get $g(y|\bar{x}_i^c) \triangleq g(y|\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{i-1}, \bar{x}_{i+1}, \cdots, \bar{x}_n)$ and

$$D_{KL_{X_i^c}}\left(g(y|\bar{x}_i^c)\|f(y)\right) = \int_{-\infty}^{\infty} g(y|\bar{x}_i^c) \times \ln \frac{g(y|\bar{x}_i^c)}{f(y)} dy. \tag{15}$$

$D_{KL_{X_i^c}}\left(g(y|\bar{x}_i^c)\|f(y)\right)$ can be viewed as the reverse of the main effect of X_i, i.e., the main effect of X_i is:

$$\text{Main effect of } X_i = -D_{KL_{X_i^c}}\left(g(y|\bar{x}_i^c)\|f(y)\right). \tag{16}$$

The negativity of the main effect ensures that the larger the value of a main effect, the more important X_i. The PSA can be also applied to partial probability sensitivity analysis and reliability-based design. However, discussing these issues in detail are out of the scope of this paper, and the interested reader can refer to Appendix B for a brief discussion.

By the notion of mutual information, one can define the K-L distance based information-theoretic sensitivity approaches which quantify the amount of output uncertainty removed by the knowledge of individual inputs and combinations thereof (Lüdtke et al., 2008). Sensitivity analysis for the joint impact of all higher order terms can be achieved by conducting an analysis of the total mutual information $I(X_1, \cdots, X_n; Y)$, which subsumes all input-output associations including interactions. The interested reader is referred to Lüdtke et al. (2008) for a detailed discussion on this aspect.

To apply the CEM optimization algorithm to the calibration problem, it is necessary to apply the time-discretization scheme to convert the infinite-dimensional functional optimization problem into a finite-dimensional parametric optimization problem. A simple way to achieve this is to divide the simulation time span into $\mathcal{N} - 1$ subintervals. The parametric optimization problem yielded belongs to continuous multi-extremal optimization problems, i.e. each decision variable is real valued and that the decision vector belongs to a subset of Euclidean space. Ngoduy and Maher (2012) modified the CEM by Kernel density estimation method to solve such optimization problems. The pseudo code for Kernel-based CEM for such continuous multi-extremal optimization is summarized as:

Step 1. Discretize the simulation time span into $\mathcal{N} - 1$ equal subintervals $[\tau_1, \tau_2], \cdots, [\tau_{\mathcal{N}-1}, \tau_{\mathcal{N}}]$.

Step 2. Set $l = 0$, choose $\hat{\mu}_0$ and $\hat{\sigma}_0^2$. Identify the initial estimation $\hat{f}_0(S(x)) = \hat{g}_0(S(x))$ using Kernel density estimation method depicted as follows:

Step 3. Increase l by 1. Generate a set of random samples $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}$ from $\mathbb{N}(\hat{\mu}_{l-1}, \hat{\sigma}_{l-1}^2)$ distribution, which is an $n-$dimensional normal distribution with independent components.

Step 3.1. For a given sample of the parameters, solve the IDM governed by the ODE system (1-2) using the Runge-Kutta method, then calculate the objective function $S(\mathbf{x_i})$ for each sample.

Step 3.2. Ordering $S(\mathbf{x_i})$ from smallest to the largest as $S_{(1)} \leq \cdots \leq S_{(N)}$ and finally evaluating the $(1-\rho)100\%$ sample percentile $\hat{\gamma}_l$ of the sample scores,

$$\hat{\gamma}_l = S_{(\lceil (1-\rho)N \rceil)}, \tag{17}$$

where ρ is a small real number.

Step 3.3. Let $\mathcal{I}$ be the indices of the N^{elite} best performing samples.

Step 3.4. For all $i \in \mathcal{I}$, determine $\hat{g}_l(S(x))$ using Kernel density estimation method:

$$\hat{g}_l(S(x)) = \frac{1}{N\varrho_l} \sum_{i \in \mathcal{I}} \frac{1}{2\pi} \exp\left(-\frac{1}{2} \left(\frac{S(x) - S(\mathbf{x_i})}{\varrho_l} \right)^2 \right). \tag{18}$$

to obtain $\hat{g}_l(x)$, where ϱ_l is the bandwidth at iteration l (Ngoduy and Maher, 2012).

Step 4. Smooth: $\hat{f}_l(x) = \beta \hat{g}_l(x) + (1-\beta)\hat{f}_{l-1}(x)$, where β is the smoothing parameter.

Step 5. The algorithm will stop if the standard deviation of random samples generated by $\hat{f}_l(x)$ is less than ϵ. Otherwise, increase l by 1 and generate random samples from $\hat{f}_l(x)$, and return to Step 3.

One of the drawbacks of the Kernel based CEM is its high computational demand. On the other hand, noting that the sampling distribution of the CEM can be quite arbitrary, and does not need to be related to the function that is being optimized. Adopting Gaussian (mixture) distributions is convenient and can reduce

the computational burden (Kroese et al., 2006). Note also that the PSA requires PDF estimations which can be achieved by the Kernel density estimation method. To this end, we combine the CEM for continuous optimization in terms of Normal updating by Kroese et al. (2006) and the PSA based on the Kernel-based estimation method by Ngoduy and Maher (2012) into a unified framework, which is summarized as follows.

Step 1. Discretize the simulation time span into $\mathcal{N}-1$ equal subintervals $[\tau_1, \tau_2], \cdots, [\tau_{\mathcal{N}-1}, \tau_{\mathcal{N}}]$.

Step 2. Set $l = 0$, choose $\hat{\mu}_0$, and $\hat{\sigma}_0^2$. In the empirical study we choose $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ with each element corresponds to the maximum acceleration a, desired deceleration b, desired speed v_0, desired time gap T, and minimum gap s_0 of IDM, respectively.

Step 3. Increase l by 1. Generate a set of random samples $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}$ from $\mathbb{N}(\hat{\mu}_{l-1}, \hat{\sigma}_{l-1}^2)$ distribution, which is an $n-$dimensional normal distribution with independent components. By default, $\mathbf{N} = 1000$.

Step 3.1. For a given sample of the parameters, solve the IDM governed by the ODE system (1-2) using the Runge-Kutta method, then calculate the objective function $S(\mathbf{x_i})$ for each sample.

Step 3.2. Ordering $S(\mathbf{x_i})$ from smallest to the largest as $S_{(1)} \leq \cdots \leq S_{(N)}$ and finally evaluating the $(1-\rho)100\%$ sample percentile $\hat{\gamma}_l$ of the sample scores,

$$\hat{\gamma}_l = S_{(\lceil (1-\rho)N \rceil)}, \tag{19}$$

where ρ is a small real number and set to 0.01.

Step 3.3. Let $\mathcal{I}$ be the indices of the N^{elite} best performing samples.

Step 3.4. For all $j = 1, \cdots, n$, let

$$\tilde{\mu}_{l,j} = \sum_{i \in \mathcal{I}} \mathbf{x}_{i,j} / N^{elite},$$

and

$$\tilde{\sigma}_{l,j}^2 = \sum_{i \in \mathcal{I}} (\mathbf{x}_{i,j} - \tilde{\mu}_{l,j})^2 / N^{elite}.$$

Step 4. Smooth: $\hat{\mu}_l = \beta\tilde{\mu}_l + (1-\beta)\hat{\mu}_{l-1}$, $\hat{\sigma}_l = \beta\tilde{\sigma}_l + (1-\beta)\hat{\sigma}_{l-1}$. Here β is set to 0.7.

Step 5. If $\max_j\{\hat{\sigma}_{l,j}\} < \epsilon$ **stop** and return $\mu \triangleq \hat{\mu}_l$ (or the overall best solution generated by the algorithm) as the approximate solution to the optimization, where $\epsilon \in R_+$ is a preset tolerance and set to 10^{-6} here. Otherwise, return to Step 3.

Step 6. PSA: Generate a new set of random samples $\mathbf{X}$ from $\mathbb{N}(\mu, \hat{\sigma}_0^2)$ distribution and regard it as original set of random samples.

Step 6.1. Fix one parameter x_i to its optimal solution μ_i to produce conditional samples, i.e. $\mathbf{X}|x_i$, $i = 1, 2, ..., 5$ for the IDM.

Step 6.2. For original and conditional sets of samples $\mathbf{X}, \mathbf{X}|x_i, i = 1, 2, ...5$, simulate the IDM, then calculate the objective function $S(\mathbf{x_i})$ accordingly. Then determine the PDFs $f(y)$ and $g(y|\bar{x}_i)$ in Section 4.2 using Kernel density estimation method be of Equation (18).

Step 6.3. Calculate the corresponding sensitivities by solving Equations (14-16) and return the PSA result.

The CEM with Normal updating could reduce computational burden of the Kernel based CEM proposed by Ngoduy and Maher (2012).

Remark 4.1. The default settings of parameters recommended in the CEM literature are capable for the proposed CEM algorithm to find plausible parameters for the IDM as it will be shown in the empirical studies. In line with the literature, see e.g. De Boer et al. (2005); Kroese et al. (2006), the meanings and selections of each parameter are summarized as follows:

- β is the smoothing parameter of the CEM which is a weighting coefficient ranging from 0 to 1. The main reason why the heuristic smoothed updating procedure performs better is that it prevents the occurrences of 0s and 1s in the parameter vectors. Once such an entry is 0 or 1, it often will remain so forever, which is undesirable (De Boer et al., 2005; Kroese et al., 2006). According to the CEM literature, the empirical studies suggest that a value of β between $0.4 \leq \beta \leq 0.9$ gives the best results. Generally, β is set to 0.7.

- ρ is the percentage of elite samples on all samples, ranging from 0 to 1. In each iteration, $\rho * N$ is the number of samples regarded as elite samples which are of good scores (i.e. small values of objective function). $\rho = 0.01$ is the recommended value in the CEM literature.

- ϵ is the prescribed threshold to terminate the CEM. The CEM stops when the smoothed standard deviation of current iteration is less than ϵ. Choosing a smaller value of ϵ can ensure a more accurate optimum at the price of more intensive computational burden. Generally, ϵ is set to 10^{-6}.

- N is the number of samples generated in each iteration. Generally, a number ranging from 1000 to 2000 is recommended in the literature.

Of course, it would be better to determine these parameters through a trial-and-error procedure for a specific problem so as to improve the performance of CEM. ∎

Remark 4.2. The PSA result can be used as the preconditioning to improve the search process of optimization problems which is a popular technique in numerical optimization especially in multiobjective applications. The dependency among parameters (captured by the traffic models) would benefit the PSA guided CEM for reducing the computational burden in the optimization process. Roughly speaking the process can be summarized as:

1. Perform a sensitivity analysis using the PSA method to calculate the sensitivity indices of all decision variables regarding the performance measure;
2. Define a simplified problem that considers only the most sensitive decision variables by imposing a user-specified threshold (or classification) of sensitivity;
3. Solve the simplified problem (that is ignoring the unimportant parameters by fixing them with initial guesses) using the CEM algorithm. Denote the optimal solution thus obtained as the Pareto optimal solutions;
4. Solve the original problem using the CEM algorithm with the Pareto optimal solutions from the simplified problem as the initial guess. Denote the optimal solutions obtained as the calibrated parameters.

The dependency among parameters would benefit the fourth step for finding the global optimal solution from the Pareto optimal solutions. When the dimension of the optimization is large and the structure is complex, this PSA guided CEM can reduce the diversity of search by exploiting preconditioning with an initial small subset of decision variables while maintaining search diversity with the injection of random solutions in the last CEM process through the correlated simulations through traffic models. ∎

5. Empirical study

In this section, several empirical tests are conducted to show the merits of the proposed framework. The test with synthetic data highlights the proposed CEM algorithm can find an optimal solution that is sufficiently close to the ground truth. Then calibration with single trajectory and multi-trajectory cases are performed to emphasize the importance of data completeness. Finally, the proposed method is compared against the benchmark GA algorithm in terms of convergence and computational efficiency.

5.1. Data processing

The data adopted in this empirical study is extracted from the trajectory dataset of Next Generation SIMulation (NGSIM) project. The data was collected on April 13, 2005 which is a 15 minutes time frame observation (4:00 pm to 4:15 pm) on a stretch of Interstate 80 in San Francisco, California. The freeway segment consists of six lanes with one on-ramp. In this empirical study, we focus on the car-following behaviors of the middle lane, which is termed lane 3 in Figure 1 and hereafter, in order to maximize the fraction of pure car-following situations.

Recent research has found that the trajectory data derived from the NGSIM project contains kinds of errors ranging from measurement to traffic flow characteristics extraction from the image treatment such as motorcycle misidentification, errors due to vehicles' lane ID misidentification and lane changing/overtaking behaviours, etc. (Punzo et al., 2011). Spikes (or outliers) can be observed in the acceleration and velocity distributions. Filtering techniques are developed to reduce the damaging consequences of those noises (leading to unrealistic values as well as the random disturbances) on the accuracy of calibration results (Punzo et al., 2011; Marzack et al., 2012; Montanino et al., 2013). Specifically, Montanino et al. (2013) proposed a multi-step procedure to eliminate possible biased trajectories in the I80 dataset, wherein a spline interpolation is adopted to remove the main outliers, a low-pass filter is applied to cut the high frequency noises, a polynomial interpolation is introduced to remove the residual non reasonable accelerations. Meanwhile, several important characteristics are preserved, e.g. the dynamics of actual vehicle driving, the internal vehicle trajectories consistency, and the consistency in platoon level. Later, it is claimed in Monteil et al. (2014) that a low pass filter would be enough to remove the main inconsistencies in the NGSIM dataset.

5.2. Calibration through synthetic data

As suggested in Ossen and Hoogendoorn (2008), to assess how well parameters can be identified by the calibration procedure, the 'ground truths' of these parameters must be known since parameters minimizing the objective function do not necessarily capture following dynamics best. To show the proposed method can actually find the global optimum, a set of synthetic data is used. The synthetic data also helps verifying whether the data contains enough information to estimate the parameters of interest which helps in analyzing the cause for deviations from the real parameters when performing calibration with real data. Calibrating the synthetic data can also provide guidance on choosing suitable variables in the objective function for better model calibration.

The synthetic data is obtained from simulating the IDM with trajectory of the leading vehicle and a set of parameters. To ensure that the synthetic data resembles real data as much as possible, the data of leading vehicle is derived from NGSIM project. Here a parameter set $x^* = [1.5, 0.8, 20, 1.25, 4.5]^T$, with each element corresponds to the maximum acceleration, desired deceleration, desired speed, desired time gap, and minimum gap, respectively, is given as the ground truth. With this set of synthetic data and the same profile for the leading vehicle, calibration is carried out to find out the ability of CEM for searching the optimal parameters from this synthetic data set. Motivated by the empirical studies by Ossen and

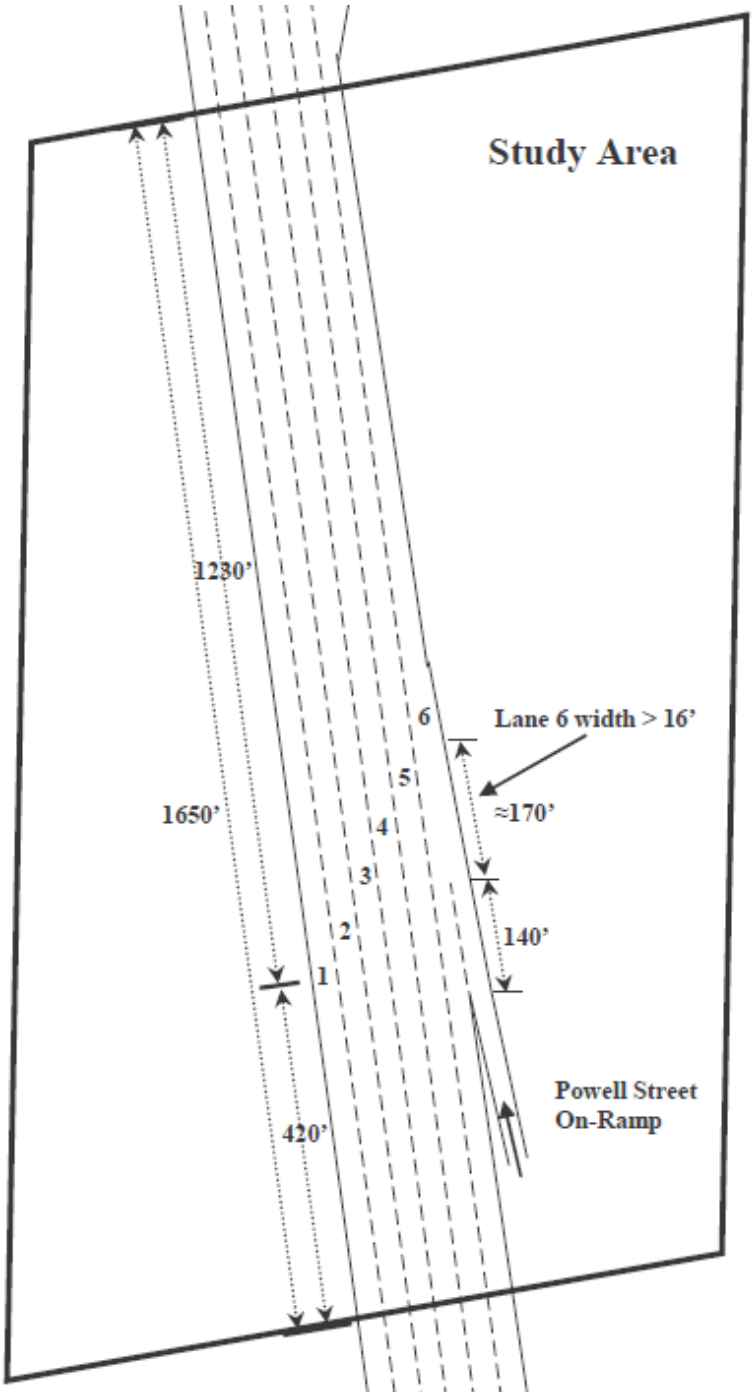

Figure 1: The study area schematic of I80, including six lanes and one on-ramp, and 1650 feet in length. Source: NGSIM project

Table 1: The searching range of each parameter of IDM.

Parameters	a, m/s^2 Maximum acceleration	b, m/s^2 Desired deceleration	v_0, m/s Desired velocity	T, s Desired time gap	s_0, m Minimum gap
Range	(0.1,6)	(0.1,6)	(1,35)	(0.1,5)	(0.1,8)

Hoogendoorn (2008); Paz et al. (2015) that including gap and speed in objective function would improve the calibration result. Thus, we adopted the following Combined Objective Function (COF):

$$\hat{x}^* = \arg\min_{x\in\mathcal{X}} \lambda \frac{\sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(s_t^{cali}(x) - s_t^{data})^2}}{\sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(s_t^{data})^2} + \sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(s_t^{cali}(x))^2}} + (1-\lambda)\frac{\sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(v_t^{cali}(x) - v_t^{data})^2}}{\sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(v_t^{data})^2} + \sqrt{\frac{1}{\mathbf{T}}\sum_{t=1}^{\mathbf{T}}(v_t^{cali}(x))^2}}, \quad (20)$$

where $s_t^{cali}(x)$ and s_t^{data} denote the calibrated gap and observed gap of following vehicle towards leading vehicle at time t respectively, and $v_t^{cali}(x)$ and v_t^{data} denote the calibrated speed derived from simulation of IDM and observed speed, respectively. $0 \leq \lambda \leq 1$ is the weighting coefficient. Constraints on each parameters of IDM, shown in Table 1, are adopted into CEM to ensure a plausible optimum solution. The initial guess of the searching ranges of parameters was based on some similar works on the calibration of the IDM, e.g. Hoogendoorn and Hoogendoorn (2010); Ciuffo et al. (2014); Rahman et al. (2015). Furthermore, Ciuffo et al. (2014) pointed out that the searching ranges should be adapted to the underlying data sets, otherwise it may lead to unmeaningful or inappropriate results of sensitivity analysis. Therefore, we defined a reasonable searching range by combining the information on the actual characteristics of the trajectory and going through trail-and-error to ensure plausible results of optimum parameters and probability sensitivity analysis.

By varying λ from 1 to 0, the difference between calibrated parameters and their "ground truth" counterparts are presented in Figure 2 and Table 2. As observed, increasing the weighting of the speed fit in objective function, i.e. decreasing λ, yields more accurate parameter estimation. Particularly, when the objective function contains only speed fit, the calibrated parameters are almost identical to their ground truth. The reason may be that when the solution is approaching closely to the optimum, the space gap is sufficiently small and/or stationary, i.e. the effect of the gap fit part is very small. Under this situation, looking into its first order derivative, i.e. the speed gap, would help moving the solution towards to the optimum. Moreover, the speed evolution of the IDM depends on the following vehicle only while the derivative of the space gap is relative to the difference between the speeds of the leading and the following vehicles, i.e.,

$$\hat{s}_t = (v_t^l - \hat{v}_t) * \triangle t + \hat{s}_{t-1},$$

where $\triangle t = 0.1$ is the sample interval of NGSIM data. As the speed profile of the leading vehicle is given, the space gap will decrease by reducing the speed gap given that there is no offset which can be achieved easily by initializing the IDM with the corresponding measured location.

5.3. Calibration with actual single trajectory

The goal of this test is to reproduce the chosen actual single trajectory data by identifying the five parameters of the IDM using cross entropy method whilst to evaluate the impact of each parameter on objective function by probabilistic sensitivity analysis. To this end, we randomly select a pair of following-leading vehicle from I80 dataset and set λ to be 1, 0.5, 0.01, 0 respectively in the COF to assess the performance of the proposed calibration method.

Table 2: The calibrated parameters and the corresponding Relative Errors (RE) with respect to different weighting coefficients. Here $RE = \frac{|\hat{x}^* - x^*|}{x^*}$ and OV means the Value of the Objective function.

	Parameters	$a, m/s^2$	$b, m/s^2$	v_0, m/s	T, s	s_0, m	OV
$\lambda = 1$	Calibration	1.43	0.89	18.55	1.29	4.31	1.34E-5
	RE	4.67%	11.25%	7.25%	3.20%	4.22%	
$\lambda = 0.8$	Calibration	1.43	0.86	19.14	1.29	4.31	1.41E-5
	RE	4.67%	7.50%	4.30%	3.20%	4.22%	
$\lambda = 0.5$	Calibration	1.45	0.84	19.97	1.29	4.33	1.28E-5
	RE	3.33%	5.00%	0.15%	3.20%	3.78%	
$\lambda = 0.2$	Calibration	1.47	0.81	20.47	1.28	4.38	7.81E-6
	RE	2.00%	1.25%	2.35%	2.40%	2.67%	
$\lambda = 0.1$	Calibration	1.48	0.81	20.39	1.27	4.42	4.82E-6
	RE	1.33%	1.25%	1.95%	1.60%	1.78%	
$\lambda = 0.01$	Calibration	1.50	0.80	20.06	1.25	4.49	6.53E-7
	RE	0.00%	0.00%	0.00%	0.00%	0.00%	
$\lambda = 0.001$	Calibration	1.50	0.80	20.00	1.25	4.50	6.82E-8
	RE	0.00%	0.00%	0.00%	0.00%	0.00%	

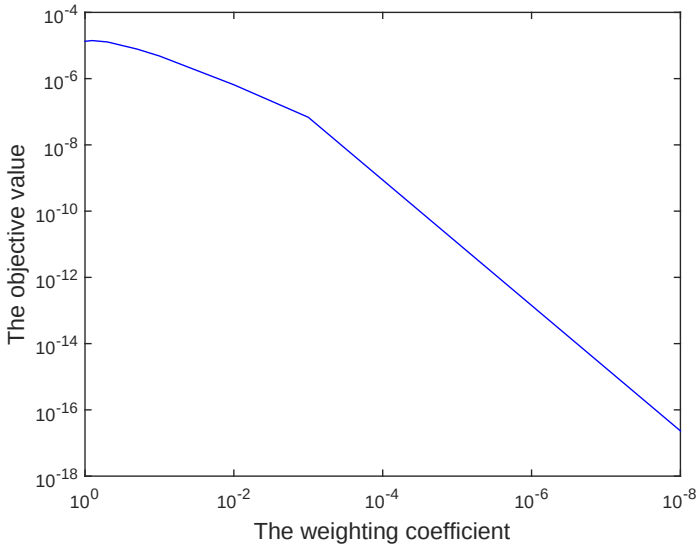

Figure 2: Optimal value of the objective function with respect to λ. The smaller the weighting coefficient is, the closer to the global optimum 0 the objective function is.

Table 3: Solutions of the IDM calibrated with single trajectory with respect to different of λ.

Parameters	$a, m/s^2$	$b, m/s^2$	v_0, m/s	T, s	s_0, m	OV
$\lambda = 1$	1.20	6.00	35.00	1.25	3.28	0.0170
$\lambda = 0.5$	1.19	6.00	35.00	1.25	3.25	0.0105
$\lambda = 0.01$	1.07	6.00	35.00	1.29	2.31	0.0042
$\lambda = 0$	1.04	6.00	35.00	1.31	1.62	0.0040

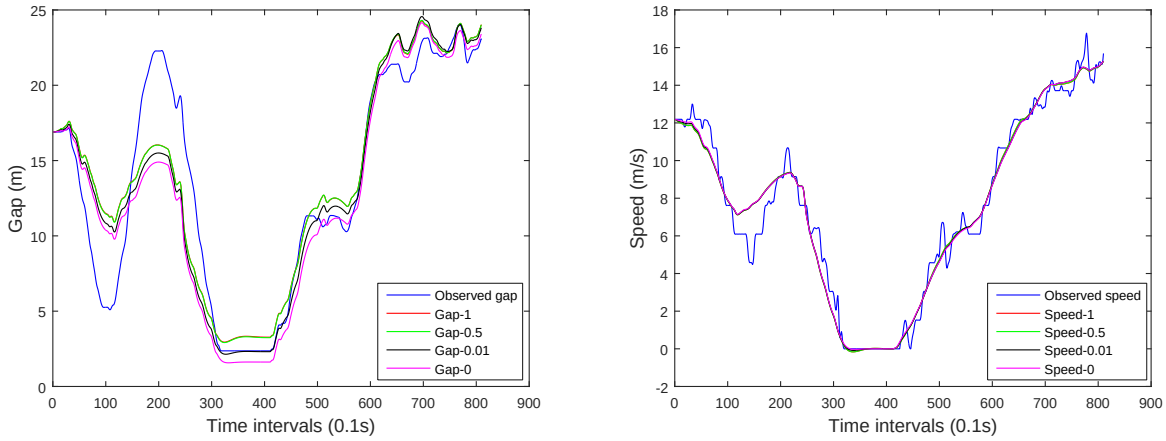

Figure 3: Calibration results of single trajectory with respect to Gap (left), Speed (right) measures, the legend 'Gap-1' means the case of $\lambda = 1$.

With default settings of the CEM algorithm depicted in Section 4.2, the calibration results are presented in Table 3. The optimum solutions of four cases seem plausible except the same high values of desired speed v_0 and desired deceleration b, which indeed hit the specified boundaries. This may be because that this actual single trajectory is not sufficient for calibrating all parameters which are relevant to different traffic regimes rather than the detailed structure of the objective function. To be specific, this single trajectory contains neither free flow nor approaching traffic regime which is necessary to calibrate desired speed and desired deceleration, respectively. Treiber et al. (2014) regarded this phenomenon as data incompleteness. As a consensus in the literature of calibration of microscopic models, the choice of suitable inputs (inputs have to sufficiently excite the system in order to be informative) is critical. Montanino et al. (2013) pointed out that the leading drivers with large standard deviation on velocity would lead to a better identification of parameters. Later in Section 5.4, we can see that such phenomenon can be alleviated if more vehicle trajectories covering more traffic operation regimes are used in the calibration.

For a more visible comparison, the simulated trajectory is plotted against the measured one in terms of gap and speed measures in Figure 3. The high consistency between the calibrated and observed measured of four cases indicates that the calibration results by the CEM are plausible. Remarkably, a better identification of standing traffic situation during time intervals from 300 to 400 is observed when $\lambda = 0.01$ rather than $\lambda = 0$. Although the global optimum can be found with synthetic data when there is only speed fit in objective function, i.e. $\lambda = 0$, a small weighting of the space gap is more suitable for calibration with actual trajectory data in this case study. This is may be due to the reason that, in contrast with the synthetic data, actual data such as NGSIM contains kinds of errors which renders considering speed fit only in the objective function is unsuitable for calibration (Treiber et al., 2014; Ossen and Hoogendoorn, 2008).

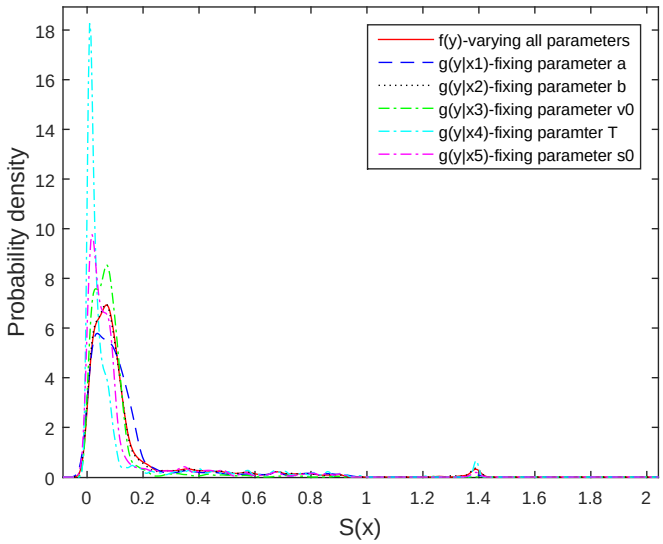

Figure 4: The probability distribution function of objective function for the case of COF. $g(y|\bar{x}_i)$ from 1 to 5 are derived from fixing a, b, v_0, T, and s_0 to its solution shown in Table 3 respectively while $f(y)$ is defined as the original PDF by varying all parameters. The horizontal axis denotes objective function value and vertical axis represents relevant probability density.

Probabilistic sensitivity analysis is conducted to provide an intuitive view of data incompleteness and identify the importance of certain parameters. For the case of $\lambda = 0.01$, the PDF by fixing certain parameter x_i to its optimal value while the other four parameters vary in pre-specified ranges, which is denoted as $g(y|\bar{x}_i)$ in Section 4.2, is presented in Figure 4. Note that $g(y|x_2)$, where x_2 is the desired deceleration, is almost the same as $f(y)$ which implies that the variation of other parameters except the desired deceleration contributes almost the total uncertainty to the objective function. As a result, the desired deceleration is identified as an insignificant parameter with respect to the objective function under this single trajectory case. In contrast, the PDF of the objective function by fixing the desired time gap T presents a sharp spike. This distinct difference between the original PDF and $g(y|x_4)$ indicates that the variance of objective function concentrates around the range of 0 when desired time gap T is fixing to its optimum. That is to say the variance of desired time gap contributes a lot to the uncertainty of the objective function under this single trajectory. Moreover, the low impact of certain parameters on objective function may be regarded as an index to data incompleteness such as the desired deceleration case.

Note that if the calibrated parameters (the converged mean values found by the CEM) are the optimal ones, the objective function should be close to 0. Negative values of the objective function can be observed with very small probability density in Figure 4. This is induced by the smoothing technique of the Kernel density estimation method that a jump from zero probability (left-hand side of the origin) to a positive probability (right-hand side of the origin that is very close to 0) is approximated by a smooth Kernel function.

The discrepancy of each $g(y|\bar{x}_i)$ to $f(y)$ is estimated by K-L distance shown in Table 4. On the one hand, small values of K-L distance of the desired deceleration and desired speed indicate that these two parameters are relevant insignificant inputs to objective function under this single trajectory, which may also correspond to the conclusion of data incompleteness of this single trajectory. On the other hand, the desired time gap T and minimum gap s_0 , which are related to safe distance of individual driver in the IDM, are regarded as the

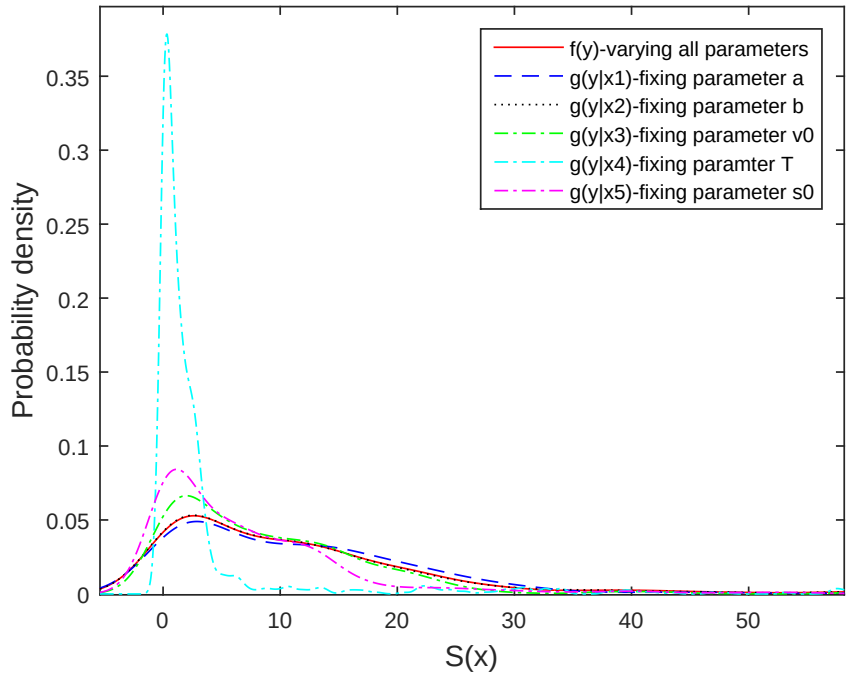

Figure 5: The probability distribution function of objective function for the case of AOF.

Table 4: The K-L distance between $g(y|\bar{x}_i)$ and $f(y)$ shown in Figure 4-5.

Fixing parameter	a	b	v_0	T	s_0
$D(P_i\|\|P_0)$ COF	28	2	21	84	38
$D(P_i\|\|P_0)$ AOF	0.59	0.05	0.51	4.81	1.97

most significant parameters out beyond other parameters as indicated by their large values of K-L distance. Their variation contributes a lot to the fluctuation in the difference between calibrated and observed gap, which is consistent with the finding of Kesting et al. (2009).

In line with the literature, the absolute measure (which will be denoted as AOF hereafter) that minimizes the difference between simulated gap and observed gap is adopted for the comparison purpose, i.e.,

$$x^* = \arg\min_{x\in\chi} \frac{\sum_{t=1}^{\mathbf{T}}(s_t^{cali}(x) - s_t^{data})^2}{\sum_{t=1}^{\mathbf{T}}(s_t^{data})^2}. \tag{21}$$

Similarly, the PSA is conducted for this objective function which is reported in Figure 5 and Table 4. Comparing the PSA results, we have the following observations:

- The PSA method introduced can handle non-Gaussian PDFs, which is always the case in traffic engineering as claimed in Section 4.
- Different objective functions can yield very different PDFs for accessing the PSA as shown in Figures 4-5. As a consequence, the values of K-L distance are also very different.

Table 5: Calibrated parameters with multi-trajectories and $\lambda = 0.05$.

a, m/s^2	b, m/s^2	v_0, m/s	T, s	s_0, m	OV
0.82	0.81	19.80	1.47	1.17	0.0084

- Regardless the differences in the PDFs and the values of K-L distance, the order of important parameters remains unchanged. This is important in the sense that the PSA result will not vary with respect to the choice of objective function.

5.4. Calibration with multiple vehicle trajectories

In this case, the application of multi-trajectories aims at extending relevant traffic regimes for calibrating parameters of the IDM especially the desired deceleration. Exacted from the same freeway segment of the NGSIM project, 244 vehicle trajectories collected from lane 3, including 111 effective leading-following vehicle pairs, are chosen to build the multi-trajectories dataset. However, the approach to compare dynamic position of single trajectory as shown in Figure 3 is unsuitable for representing the case of multi-trajectories. Instead, a comparison and analysis of the space mean speed at a certain location is adopted to evaluate the correctness and effectiveness of calibration with multi-trajectories (Tian et al., 2015). The chosen site for space mean speed evaluation is located in the middle of lane 3 ranging from 220m to 222m. The reason for choosing this site is that the following driving behavior was the dominating one. A time interval of 30 seconds is used to aggregate traffic data so as to evaluate the space mean speed of the test site.

Under the same constraints on each parameter of the IDM (i.e. Table 1) **and COF with** $\lambda = 0.05$, the space mean speed of all 244 vehicle, 111 observed and calibrated leading-following pairs are plotted in Figure 6. The high consistency among the observed and calibrated space mean speed indicates the capability of the CEM for calibrating the IDM with multi-trajectories. Since there are only two vehicle pairs passing by the test site during the 13th and 14th counting time intervals, the calibrated speed admits larger error. This is because the two-vehicle pair is not adequate to reflect the macroscopic traffic characteristics.

The calibrated parameters are presented in Table 5, in particular the desired deceleration seems to be much more plausible compared with the one calibrated through single trajectory. This may be because that multi-trajectories, enriching the approaching traffic regime which is validated by the apparent sharp drop of space mean speed shown in Figure 6, are efficient to calibrate desired deceleration. Although the desired speed seems to be plausible when compared with the one calibrated through single trajectory, it is worthwhile to point out that lacking free flow traffic regime in the multi-trajectories, as reflected by Figure 6 that the maximum space mean speed is less than 11 m/s, the desired velocity may be still unreliable. While other parameters are consistent with those obtained in the literature, the desired velocity is heavily related to the chosen data. In Rahman et al. (2015), very small desired velocity was calibrated.

To show the convergence of the CEM, the evolution of the PDF of the desired time gap under an arbitrarily guessed initial gaussian distribution (with large variance) is shown in Figure 7. The change on the horizontal axis values of the PDF over iterations indicates that the PDF of the desired time gap becomes more and more concentrated around the optimal solution over iterations, and the optimal parameter is found when the standard deviations of this distribution is approximately zero (i.e. the density function is spiked).

The PSA is conducted to validate the effective identification of the desired deceleration and provide interpretation of each parameter in case of multi-trajectories. The results can be found in Figure 8 and relevant K-L distance is shown in Table 6. The larger value of K-L distance of desired deceleration, in contrast to single trajectory, indicates that desired deceleration plays a relevant more significant role on objective function when calibrating with multi-trajectories in which approaching traffic regime is sufficient.

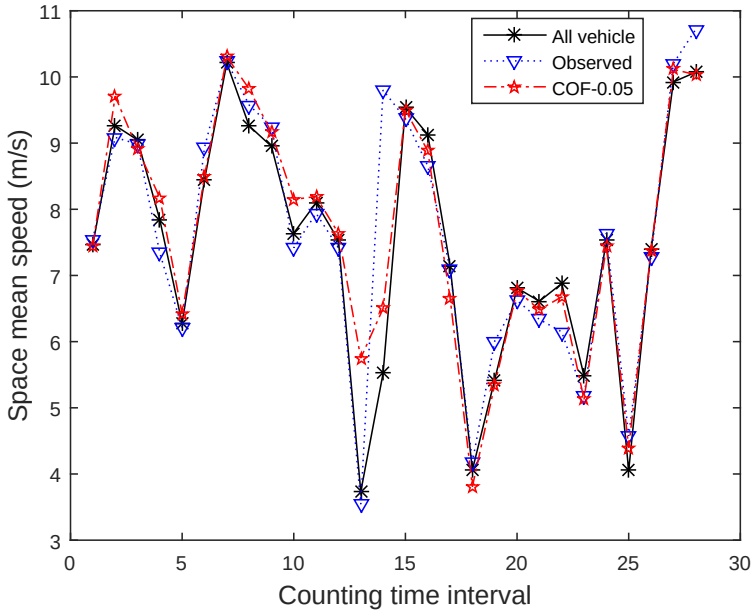

Figure 6: Space mean speed at location from 220m to 222m.

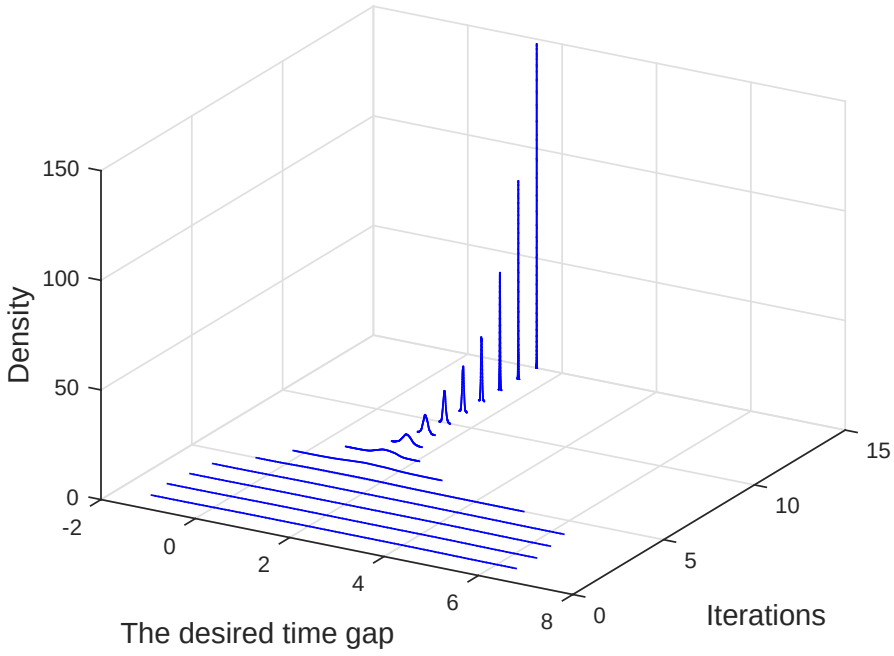

Figure 7: PDF estimation of T over iterations under multi-trajectories.

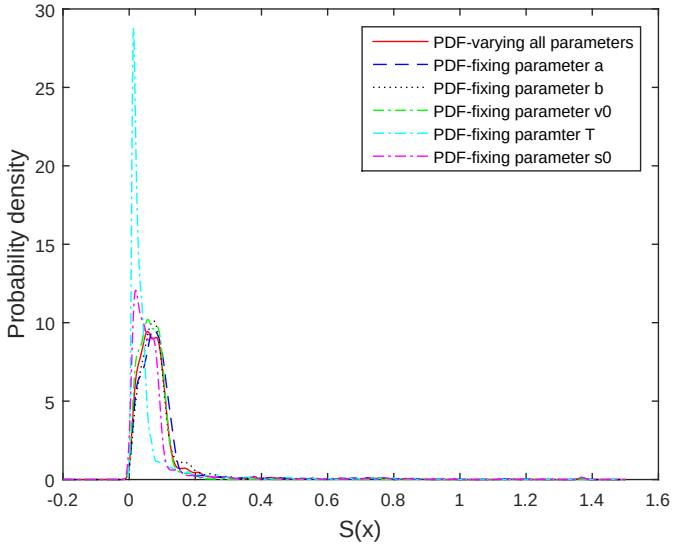

Figure 8: The probability distribution function of the objective function for the case of multi-trajectories.

Table 6: The K-L distance between $g(y|\bar{x}_i)$ and $f(y)$ shown in Figure 8.

Fixing parameters	a	b	v_0	T	s_0
$D(P_i\|\|P_0)$	82	46	50	400	145

However, the desired velocity remains an insignificant factor in both cases. Since desired speed (or free flow speed) in real traffic situation is a relevant stable parameter in macroscopic traffic flow (i.e. by aggregating lots of vehicles), its low impact on the objective function is observed. The largest K-L distance of T confirms that desired time gap remains the most significant input beyond other parameters. Specifically, if all vehicle are restricted to drive under a same safe distance, small fluctuation on safe distance causes great change on the difference between observed and calibrated data, indicating the existing of different preferred safe distance of different drivers, which is consistent with the observation of Laval et al. (2010).

5.5. Performance assessment of the CEM and GA for calibration

In microscopic calibration literature, Genetic Algorithm (GA) is a widely adopted algorithm, which is also regarded as the benchmark solution algorithm, to solve the optimization problem, see e.g. Ciuffo and Punzo (2014); Rahman et al. (2015). The numerical results in Ciuffo and Punzo (2014) revealed that the GA outperforms the simultaneous perturbation stochastic approximation, simulated annealing and OptQuest/Multistart heuristic methods in the sense of convergent time and optimization performance indicator for calibration of microscopic models with the same test site and data set. Paz et al. (2015) showed that a combination of the GA and the simulated annealing (i.e. Memetic algorithm) can further refine the calibration result. To this end, we compare the performance of the proposed algorithm with that of the GA.

The GA is a stochastic-based global optimization method which mimics the natural biological evolution mechanisms, including selection and reproduction, cross over, and mutation mechanisms. These three mechanisms supplement each other: the fittest individuals (parameters set) which contain good genetic material (parameters value) are preserved by the selection and reproduction mechanism; the cross over mechanism guarantees next generation to retain the good genetic material of the parent preserved by

Table 7: The solution to parameters of the IDM calibrated with single trajectory using GA.

Parameters	a, m/s^2	b, m/s^2	v_0, m/s	T, s	s_0, m
	1.07	6.00	35.00	1.29	2.31

last mechanism; the mutation mechanism helps maintain the population diversity and prevent precocious convergence. The GA is one of the most popular optimization algorithms because not only any derivative information for optimization but also differentiable property for objective function are not required during process.

The GATBX toolbox of MATLAB[1] (developed by The University of Sheffield) is adopted in this empirical study. The parameters of the GA are as follows:

- Population size: 1000;
- Maximum number of generations: not specified (from 100 to 1000);
- Precision of variables (i.e. number of binary bits used to represent a variable of an individual): 20 (which is the default setting);
- The crossover mechanism is single point crossover with a crossover rate of 0.7 (which is the default setting);
- The selection mechanism is stochastic universal sampling with a generation gap equal of 0.9 (which is the default setting);
- The mutation rate is resulted from the crossover rate divided by the precision of variables, i.e. 0.7/20.

The optimal solution to the calibration problem of the IDM with single trajectory with respect to the COF with $\lambda = 0.01$ using GA is shown in Table 7. As observed, the parameters calibrated by the CEM and the GA are almost the same when comparing Table 3 and Table 7. However, their convergence properties are quite different.

5.5.1. Conditions of convergence

To explore the convergence of the CEM and the GA, two commonly used termination criteria for GA are adopted:

- Standard deviation of objective function S of current iteration k is less than ε:

$$TermCond_1 : \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_{ki} - \bar{S}_k)^2} \leq \varepsilon.$$

- The difference between the current optimal value of the objective function S_k^{best} and the average of the optimal values of the objective function of last (several) iterations t_{last} is less than ε, which is also termed as running mean:

$$TermCond_2 : \left\| S_k^{best} - \frac{1}{t_{last}} \sum_{i=k-t_{last}+1}^{k} S_i^{best} \right\| \leq \varepsilon.$$

[1]The interested reader is referred to the online tutorial: https://www.shef.ac.uk/polopoly_fs/1.60188!/file/manual.pdf.

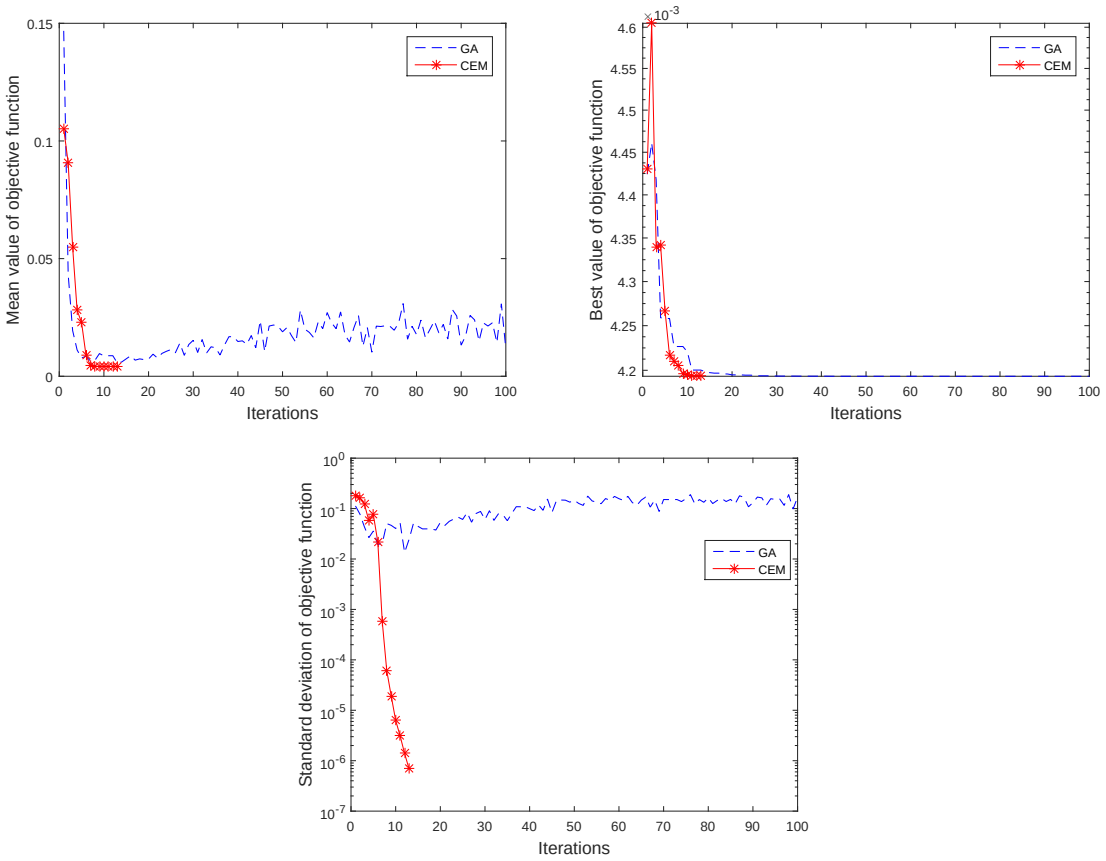

Figure 9: The convergence property of the CEM and the GA with respect to the termination criteria of 'standard deviation', 'mean value' and 'best value', respectively. The 'mean value' and 'best value' criteria correspond to the average value of the objective function of the whole 1000 samples and the minimum value of the objective function in each iteration, respectively. The maximum iteration number for GA is set to 100.

Here we set $t_{last} = 15$ and $\varepsilon = 10^{-6}$. In the following cases, the above two termination criteria are used to calibrate the single trajectory shown in Figure 3 to investigate different convergent properties of these two algorithms.

As indicated in Figure 9 that in terms of the standard deviation convergence criterion the CEM converges within 17 iterations while the GA seems to fail to converge. In terms of the mean value, the CEM also converges within several iterations. In contrast, the mean value of GA decreases in the first 20 iterations but trends to fluctuate in a large scale later on. This may be due to that the GA cannot decrease the mean value of the objective after several iterations (say 20 in this case). The mutation mechanism then tries to change certain property of several samples to explore new potential best samples but fails. This failure results in several extremely high values of the objective function. As a consequence, the standard deviation of objective function also oscillates and cannot meet the standard deviation convergence criterion.

As for the second termination criterion, Figure 10 depicts the convergence process for the case of calibration with multi-trajectories. Observe that both the running means of the GA and CEM are decreasing with respect to iterations in general and are less than ε after finite iterations. Therefore, the running mean

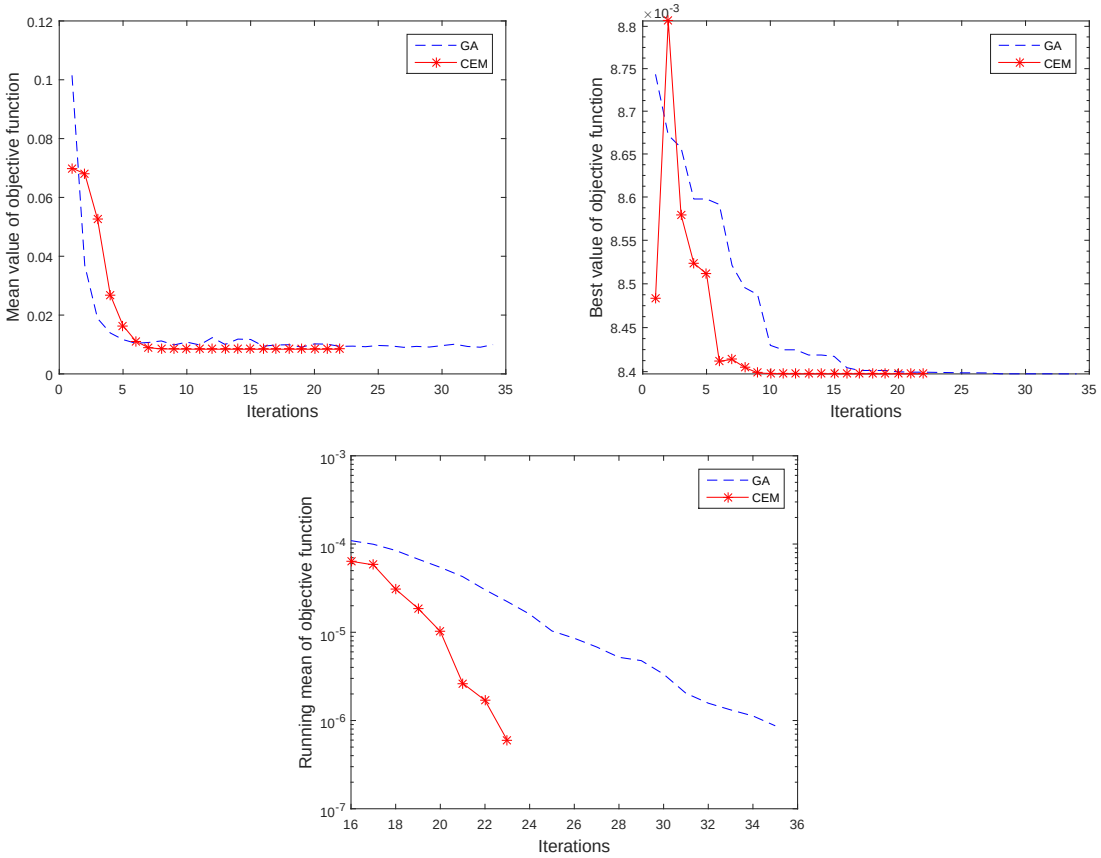

Figure 10: The convergence property of the CEM and GA with respect to running mean for multi-trajectories case with $\lambda = 0.05$, where $t_{last} = 15$.

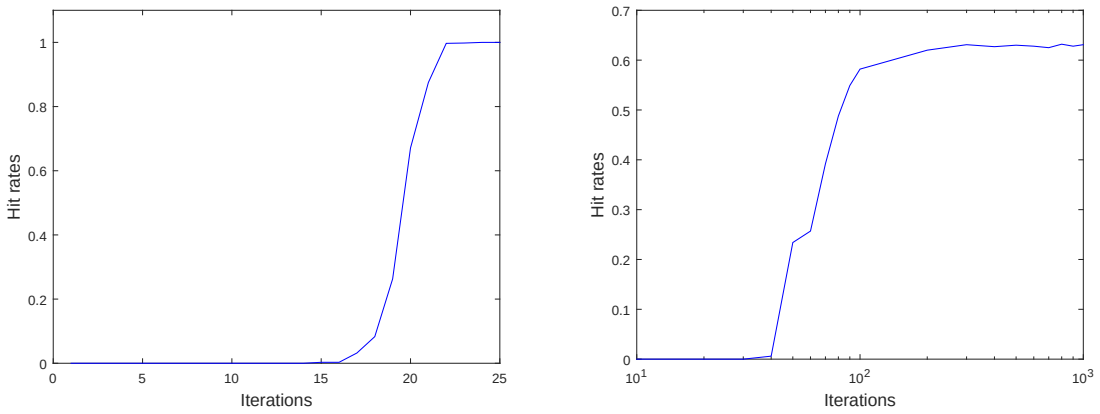

Figure 11: Comparison of hit rates of the CEM and GA.

termination criterion is fulfilled. The CEM takes less iterations to converge than the GA does. In terms of computational time, the CEM requires around 11 seconds to converge while the GA takes around 15 seconds for the single trajectory case. For the multi-trajectories case, the CEM takes around 700 seconds to converge while the GA needs around 1000 seconds, indicating a better performance of the CEM than the GA.

5.5.2. Hit rates towards optimum solution

The CEM and the GA are both stochastic optimization algorithms by random samples that aim to improve the generation of individual (or sample) towards optimal solution at each iteration. The hit rates is an indicator to quantify the percentage of individuals in a generation that attain the optimal solution in the GA literature. Ten independent trials with 1000 samples are carried out to ensure a stable estimation of the hit rates. As indicated in Figure 11 that the CEM is of 100% hit rates after 24 iterations while the GA is of less than 70% after 1000 iterations. Furthermore, the hit rates of the GA cannot improve further with more iterations since the performance of the GA is mostly dependent on initial settings, such as initial generation, sample size, crossover and mutation probabilities. There is no systematical method for choosing preferable initial settings for fast convergence for the GA. Generally, it is obtained through trial-and-error. In contrast, the CEM is with a rigorous proof of convergence under mild assumptions. The CEM outperforms the GA in this sense.

6. Conclusions

This paper proposed a new framework to calibrate car-following models with uncertainties such as noisy data and heterogeneous driver behaviors. Converting the calibration problem into a rare event detection problem, the CEM based optimization method is proposed to calibrate the parameters of car-following models. To identify important parameters, the probabilistic sensitivity analysis is applied to investigate the probabilistic behavior of the model output response with respect to its parameter uncertainties. The PSA can handle the case that the PDF of interest is nonlinear and/or non-Gaussian, and not only in a few moments, due to its heavy tail. Furthermore, since the CEM and PSA are both based on the K-L distance, they can be simultaneously integrated into a unified framework.

Several empirical studies were conducted to illustrate the performance of the proposed calibration framework. First, a set of synthetic data was used to show the proposed calibration algorithm is capa-

ble of finding the "true" global optimum. Then both single trajectory data and multi-trajectories data were adopted to show the effectiveness of the CEM to calibrate the IDM. Meanwhile, the unrealistic value of certain parameter, which may be caused by insufficient information, has been quantified by the low value of the relative entropy through the lens of PSA for the single trajectory case. This quantification is consistent with the mutual information concept. By doing this, it is interesting to find out that low impact of certain parameters on the objective function indicates the missing relevant trajectory data to calibrate these parameters. This data incompleteness is further confirmed by the high significance of the desired deceleration for the multi-trajectories case. The low impact of the desired velocity in multi-trajectories cases suggests the calibration program is less sensitive with respect to it. This may be because that most of the drivers would expect more or less the same desired velocity. The desired time gap and the minimum spacing are the two most significant factors by the PSA. This is consistent with the observation in huge heterogeneity in drivers' desired time gap and minimum spacing. Comparing with the benchmark solution algorithm, i.e. the GA, for the calibration of microscopic traffic models, the CEM shows its superiority in terms of accuracy, computational efficiency, and convergence property under different termination criteria. An interesting future work is to extend the framework to tackle the calibration of stochastic car-following models. The PSA guided CEM algorithm as outlined in Remark 4.2 is useful for large scale applications which also deserves further research effort.

Acknowledgments

The authors sincerely thank the referees for their insightful and constructive comments which have substantially improved the paper. Financial support from National Twelfth Five-year Plan Science & Technology Pillar Program under Grant No. 2014BAG01B05, the National Natural Science Foundation of China under Grant No. 51308559, NSFC-RS Exchange Program under Grant No. 513111163, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20130171120032, the Research Grants Council of the Hong Kong Special Administration Region under grant No. PolyU5242/12E, and Research Institute for Sustainable Urban Development (RISUD) of the Hong Kong Polytechnic University under grant No. 1-ZVBZ is gratefully acknowledged.. The authors are grateful to the NGSIM Program for data availability.

Appendix A: The cross entropy method

The objective of cross-entropy based optimization is to maximize γ such that $\ell(\gamma)$ approaches 0. By that, we may conclude that γ gives the largest value such that $S(X)$ is (with high probability) greater than γ, i.e., $S(x) > \gamma$. That is, γ is the maximum lower bound for the minimization of $S(x)$ in probabilistic sense (or $\ell(\gamma) \to 0$), and thus achieves the minimum. Two main issues needed to be addressed to proceed with the above idea. First, to compute $\ell(\gamma)$ for a given γ. Second, given current γ, to search for a better one, i.e., a descending search. Regarding the first one, given a γ, one would perform an exhaustive search on the solution space χ to estimate $\ell(\gamma)$ that is certainly not practical. The practical way is to generate some samples from χ and perform Monte Carlo simulations to estimate $\ell(\gamma)$. Specifically, generate N samples drawn from $f(x;u)$ and estimate $\ell(\gamma)$ as

$$\hat{\ell} = \frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\gamma\}}, \tag{A.1}$$

where $\mathbf{x_1}, \cdots, \mathbf{x_N}$ denotes a random sample. The problem with this crude monte carlo simulation idea is its low efficiency. One may require a large number of samples to accurately estimate $\ell(\gamma)$ when $S(X) \leq \gamma$ is

a rare event[2]. Note that, when searching the optimum, $S(X) \leq \gamma$ will eventually be a rare event when $\ell(\gamma)$ approaches 0 as γ approaches the minimum. To tackle this problem, the cross entropy method explores the power of importance sampling technique.

Roughly speaking, importance sampling uses a different probability density function $k(x;\varrho)$ on χ and computes the estimation of $\ell(\gamma)$ as $\hat{\ell}(\gamma)$ by

$$\hat{\ell}(\gamma) = \frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\gamma\}}\frac{f(\mathbf{x_i};u)}{k(\mathbf{x_i};\varrho)}. \tag{A.2}$$

Defining

$$k^*(x,\varrho) = \frac{I_{\{S(x)\leq\gamma\}}f(x;u)}{\ell(\gamma)},$$

and replacing k by k^*, we have

$$\hat{\ell}(\gamma) = \frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\gamma\}}\frac{f(\mathbf{x_i};u)}{k^*(\mathbf{x_i};\varrho)} = \ell(\gamma). \tag{A.3}$$

The problem now turns to the estimation of k^* since $\ell(\gamma)$ is unknown. Again, we try to find some probabilistic distribution close to k^* in computation. Formally, the cross-entropy technique defines the distance between two probability density functions $k(x;\varrho)$ and $f(x;v)$ using Kullback-Leibler (K-L) distance which is also known as cross-entropy. The K-L distance between these two Probability Distribution Functions (PDF) associated with a vector x and $k(x)$ is zero only when $f(x)$ is zero, is defined as follows:

$$\begin{aligned} D_{KL}(k\|f) &= E_k\left(\ln\frac{k(x;\varrho)}{f(x;v)}\right) \\ &= \int k(x;\varrho)\ln k(x;\varrho)dx - \int k(x;\varrho)\ln f(x;v)dx, \end{aligned}$$

where E_f means the expectation under the probability density f, and $D(k,f) \geq 0$ with and $D_{KL}(k\|f) = 0$ if and only if $f(x) = k(x)$. Minimizing K-L distance is equivalent to finding v to maximize

$$\max_v \int k^*(x,\varrho)\ln f(x;v)dx. \tag{A.4}$$

Substituting the definition of k^* we arrive

$$\max_v \int \frac{I_{\{S(x)\leq\gamma\}}f(x;u)}{\ell(\gamma)}\ln f(x;v)dx. \tag{A.5}$$

Remind the definition of $\ell(\gamma)$, we can obtain an equivalent optimization problem as follows:

$$v^* = \arg\max_v E_u I_{\{S(X)\leq\gamma\}}\ln f(X;v). \tag{A.6}$$

Note that the above computation needs to assume that the nominal PDF f can be parameterized by a finite-dimensional vector u; that is, $f(x) = f(x;u)$, and that the importance sampling PDF is $f(\cdot;v)$ for some

[2]Recall that if γ is close to the optimum and the probability $\ell(\gamma)$ would be quite small, then we regard $S(X) \leq \gamma$ a rare event.

parameter v. Using again importance sampling, with a change of measure $f(\cdot; w)$ we can rewrite the above optimization problem as

$$v^* = \arg\max_v E_w I_{\{S(X)\leq\gamma\}} W(X; u, w) \ln f(X; v), \tag{A.7}$$

for any reference parameter w, where

$$W(x; u, w) = \frac{f(x; u)}{f(x; w)},$$

is the likelihood ratio, at x, between $f(\cdot; u)$ and $f(\cdot; w)$ (Rubinstein et al., 2004). We may estimate v^* by solving the following stochastic program

$$\hat{v}^* = \arg\max_v \frac{1}{N} \sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\gamma\}} W(\mathbf{x_i}; u, w) \ln f(\mathbf{x_i}; v) \doteq \arg\max_v \hat{D}(v), \tag{A.8}$$

where $\mathbf{x_1}, \cdots, \mathbf{x_N}$ is a random sample from $f(\cdot; w)$. In typical applications the function $\hat{D}$ is convex and differentiable with respect to v. The solution may be readily obtained by solving the following system of equations (with respect to v):

$$\frac{1}{N} \sum_{i=1}^{N} I_{\{S(\mathbf{x_i})\leq\gamma\}} W(\mathbf{x_i}; u, w) \nabla_v \ln f(\mathbf{x_i}; v) = 0, \tag{A.9}$$

where ∇_v indicates the gradient is with respect to v.

Appendix B: Partial probability sensitivity analysis & reliability-based design

One advantage of the K-L entropy approach is that the method can be extended fairly easily by simply changing the integration range of y from $[-\infty, \infty]$ to the partial region of interest $[y_L, y_U]$ so as to conduct regional probabilistic sensitivity analysis of the model output, which is useful for robust and reliability-based design. The modified total effect of X_i is defined as

$$D_{KL_{X_i}}(g(y|\bar{x}_i)\|f(y)) = \int_{y_L}^{y_U} f(y) \times \left| \ln \frac{g(y|\bar{x}_i)}{f(y)} \right| dy, \tag{B.1}$$

where y_L and y_U stand for the lower and upper bounds of a partial range of y. Similarly, the modified main effect of X_i is defined as the negative of the combined effect of the complementary set X_i^c, i.e.,

$$\begin{aligned} \text{Main effect of } X_i &= -D_{KL_{X_i^c}}\left(g(y|\bar{x}_i^c)\|f(y)\right) \\ &= -\int_{y_L}^{y_U} f(y) \times \left| \ln \frac{g(y|\bar{x}_i^c)}{f(y)} \right| dy. \end{aligned} \tag{B.2}$$

Different from Equations (14) and (16), Equations (B.1) and (B.2) measure the absolute divergence between two distributions over the region of interest. As discussed in Liu et al. (2006) that the absolute value is used because the log-likelihood can take both positive and negative values within the partial region of y. For the integral calculation, without taking the absolute values, the positive and negative log-likelihood values may cancel with each other. In that case, the relative entropy value could be very small regardless the huge difference between $g(y)$ and $f(y)$ in actuality. Integration of the absolute values better reflects the actual difference between two distributions in a specified region. Moreover, in Equations (B.1) and (B.2),

$f(y)$ is used as a weighting factor instead of $g(y)$ as in Equations (14) and (16). This is necessary when dealing with heavy tailed distributions. For a heavy tailed distribution, if the region of interest is its extreme tail, the value of $g(y)$ may be very small or even approaching to zero because $g(Y)$ stands for the distribution of Y after eliminating the uncertainty of input variables. When compared with $f(y)$, the amplitude may be even more small. The small values of $g(y)$ may diminish the actual divergence from $g(y)$ to $f(y)$. As demonstrated in Liu et al. (2006), the value of $f(y)$ is larger at the tail and better suited to be used as the weighting factor.

For reliability-based design, the sensitivity of the probability of failure P_f or its complementary-the probability of success $1 - P_f$ is of particular interest. For a given design setting, the system has two possible outcomes: safe or unsafe, with probability of $1 - P_f$ and P_f, respectively. If we view the system safety as a discrete event, a relative entropy based PSA for reliability can be calculated based on contributions of random variables to both the success and failure conditions, denoted $D_{KL_{x_i}}^{P_f}(\hat{P}_f \| P_f)$, as follows:

$$D_{KL_{x_i}}^{P_f}(\hat{P}_f | P_f) = \hat{P}_f \log \frac{\hat{P}_f}{P_f} + (1 - \hat{P}_f) \log \frac{(1 - \hat{P}_f)}{(1 - P_f)}. \tag{B.3}$$

where P_f is the original probability of failure of a performance. $\hat{P}_f$ is the probability of failure when fixing X_i at a specific value. Equation (B.3) is the total sensitivity (effect) index of X_i on the reliability. We view the formulation in Eq. (B.3) as an alternative approach to that in Equation (B.1). Similarly, by fixing all random variables except X_i in $\hat{P}_f$, $-D_{KL_{x_i^c}}^{P_f}(\hat{P}_f \| P_f)$ provides an equivalent measure of the main effect of X_i.

Appendix C: Notation list

References

Asamer, J., Zuylen, H., Heilmann, B., 2013. Calibrating car-following parameters for snowy road conditions in the microscopic traffic simulator VISSIM. IEEE Trans. Intell. Transport. Syst. 7(1), 114-121.

Barceló, J., 2010. Fundamentals of traffic simulation. Springer.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer-Verlag New York.

Chiabaut, N., Leclercq, L., Buisson, C., 2010. From heterogeneous drivers to macroscopic patterns in congestion. Transport. Res. Part B 44, 299-308.

Ciuffo, B., and Azevedo, C., 2014. A sensitivity-analysis-based approach for the calibration of traffic simulation models. IEEE Trans. Intell. Transport. Syst. 15, 1298-1309.

Ciuffo, B., Punzo, V., 2014. 'No Free Lunch' Theorems applied to the calibration of traffic simulation models. IEEE Trans. Intell. Transport. Syst. 15, 553-562.

Ciuffo, B., Punzo, V., Montanino, M., 2014. Global sensitivity analysis techniques to simplify the calibration of traffic simulation models. Methodology and application to the IDM car-following model. IET Intell. Transport Syst. 8 (5), 479-489.

De Boer, P., Kroese, D., Mannor, S., Rubinstein, R., 2005. A tutorial on the cross-entropy method. Ann. Oper. Res. 134, 19-67.

Duret, A., Buisson, C., Chiabaut, N., 2008. Estimating individual speed-spacing relationship and assessing ability of newells car-following model to reproduce trajectories. Transport. Res. Rec. 2088, 188-197.

Ge, Q., Ciuffo, B., Menendez, M., 2014. An exploratory study of two efficient approaches for the sensitivity analysis of computationally expensive traffic simulation models. IEEE Trans. Intell. Transport. Syst. 15, 1288-1297.

Hale, D., Antoniou, C., Brackstone, M., Michalaka, D., Moreno, A., Parikh, K., 2015. Optimization-based assisted calibration of traffic simulation models. Transport. Res. Part C 55, 100-115.

Hollander, Y., and Liu, R., 2008. The principles of calibrating traffic microsimulation models. Transport. 35: 347-362.

Hoogendoorn, S., and Hoogendoorn, R., 2010. Calibration of microscopic traffic-flow models using multiple data sources. Phil. Trans. R. Soc. A 368, 4497-4517.

Kesting, A., Treiber, M., 2008. Calibrating car-following models using trajectory data: methodological study. Transp. Res. Rec. 2088, 148-156.

Parameters	Description
a	the maximum acceleration of following vehicle
b	the desired deceleration
s	the instantaneous bumper-to-bumper gap between following and leading vehicle
s_0	the minimum gap
T	the desired time gap
v_0	the desired velocity of following vehicle
Y_{ij}	the 'true' data of i^{th} set with j^{th} component
$\tilde{Y}_{ij}$	the observed data of i^{th} set with j^{th} component
d	the dimension of the state
n	the dimension of parameters set
σ_j	the standard deviation of the noise of j^{th} component
ϵ_{ij}	the standard Gaussian distributed random variables
Θ	the complete set of parameters
$X = (X_1, X_1, ..., X_n)$	the parameters set
$f(x; v)$	the PDF parameterized by v with respect to parameters set x
$\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N})$	the set of N random samples
$I(\cdot)$	the indicator function
$D_{KL}(k\|\|f)$	the Kullback-Leibler distance between PDF k and f
$W(x; u, w)$	the likelihood ratio between $f(x; u)$ and $f(x; w)$
Y	the response of model
$f(y)$	the PDF of Y with all random inputs
$g(y\|\bar{x}_i)$	the PDF of Y while fixing X_i to $\bar{x}_i$
$s_t^{cali}(x)$	the calibrated gap of IDM with respect to x at time interval t
s_t^{data}	the observed gap at time interval t

Kesting, A., Treiber, M., 2009. Calibration of car-following models using floating car data. Springer, Berlin Heidelberg, pp. 117-127.

Kim, T., Zhang, H., 2008. A stochastic wave propagation model. Transport. Res. Part B 42 (7), 619-634.

Komorowski, M., Costa, M., Rand, D., Stumpf, D., 2011. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. Proc. Natl. Acad. Sci. U.S.A. 108, 8645-8650.

Kontorinaki, M., Spiliopoulou, A., Papamichail, I., Papageorgiou, M., Tyrinopoulos, Y., Chrysoulakis, J., 2015. Overview of nonlinear programming methods suitable for calibration of traffic flow models. Oper. Res. Int. J. inpress (DOI 10.1007/s12351-014-0146-9).

Kroese, D., Porotsky, S., Rubinstein, R., 2006. The cross-entropy method for continuous multi-extremal optimization. Methodol. Comput. Appl. Probab. 8, 383-407.

Kroese, D., Rubinstein, R., and Glynn, P., 2013. The cross-entropy method for estimation. Handbook of Statistics, Vol 31. Chennai: Elsevier B.V., pp. 19-34.

Kucherenko, S., Tarantola, S., Annoni, P., 2012. Estimation of global sensitivity indices for models with dependent variables. Comput. Phys. Commun., 183, 937-946.

Laval, J., Leclercq, L., 2010. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. Phil. Trans. R. Soc. A, 368, 4519-4541.

Liu, H., Chen, W., Sudjianto, A., 2006. Relative entropy based method for probabilistic sensitivity analysis in engineering design. Transactions of the ASME, 128, pp. 326-336.

Lownes, N., Machemehl, R., 2006. Sensitivity of simulated capacity to modification of VISSIM driver behavior parameters. Transport. Res. Rec. 1988, 102-110.

Lüdtke, N., Panzeri, S., Brown, M., Broomhead, D., Knowles, J., Montemurro, M., Kell, D., 2008. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. J. R. Soc. Interface 5, 223-235.

Maher, M., Liu, R., and Ngoduy, D. , 2013. Signal optimisation using the cross entropy method. Transport. Res. Part C 27, 76-88.

Majda, A., Wang, X., 2006. Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows. Cambridge Univ Press, Cambridge, UK, pp. 564.

Majda, A., Gershgorin, B., 2010. Quantifying uncertainty in climate change science through empirical information theory. Proc. Natl. Acad. Sci. U.S.A. 107, 14958-14963.

Majda, A., Gershgorin, B., 2011. Improving model fidelity and sensitivity for complex systems through empirical information theory. Proc. Natl. Acad. Sci. U.S.A. 108, 10044-10049.

Marzack, F., Buisson, C., 2012. New filtering method for trajectory measurement errors and its comparison with existing methods. Transport. Res. Rec. J. 2315, pp. 35-46.

Mathew, T., Radhakrishnan, P., 2010. Calibration of microsimulation models for nonlane-based heterogeneous traffic at signalized intersections. ASCE J. Urban Plann. Develop., 136 (1), 59-66.

Montanino, M., Punzo, V., 2013. Making NGSIM data usable for studies on traffic flow theory: a multistep method for one vehicle trajectory reconstruction. Transportation Research Board Annual Meeting.

Monteil, J., Billot, R., Sau, J., Buisson, C., Faouzi, N., 2014. Calibration, estimation and sampling issues of car-following parameters. Transport. Res. Rec., 2422, 131-140.

Ngoduy, D., Maher, M., 2012. Calibration of second order traffic models using continuous cross entropy method. Transport. Res. Part C 24, 102-121.

Ossen, S., Hoogendoorn, S., 2005. Car-following behavior analysis from microscopic trajectory data. Transport. Res. Rec. 1934, 13-21.

Ossen, S., Hoogendoorn, S., 2008. Calibrating car-following models using microscopic trajectory data: A critical analysis of both microscopic trajectory data collection methods, and calibration studies based on these data. Techical report, Delft University of Technology.

Ossen, S., Hoogendoorn, S., 2008. Validity of trajectory-based calibration approach of car-following models in presence of measurement errors. Transp. Res. Rec. 2088, 117-125.

Pantazis, Y., Katsoulakis, M., 2013. A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics. J. Chem. Phys. 138, 054115.

Paz, A., Molano, V., Martinez, E., Gaviria, C., Arteaga, C., 2015. Calibration of traffic flow models using a memetic algorithm. Transport. Res. Part C 55, 432-443.

Punzo, V., Ciuffo, B., 2009. How parameters ofmicroscopic traffic flow models relate to traffic dynamics in simulation. Transp. Res. Rec., 2124, 249-256.

Punzo, V., Borzacchiello, M., Ciuffo, B., 2011. On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (NGSIM) program data. Transport. Res. Part C 19, 1243-1262.

Rahman, M., Chowdhury, M., Khan, T., Bhavsar, P., 2015. Improving the efficacy of car-following models with a new stochastic parameter estimation and calibration method. IEEE Trans. Intell. Transport. Syst., inpress.

Rubinstein, R., 1997. Optimization of computer simulation models with rare events. Eur. J. Oper. Res., 99 (1), 89-112.
Rubinstein, R., and Kroese, D., 2004. The cross-entropy method: a unified approach to combinatorial optimization, monte-carlo simulation and machine learning. Springer.
Saifuzzaman, M., Zheng, Z., 2014. Incorporating human-factors in car-following models: A review of recent developments and research needs. Transport. Res. Part C 48, 379-403.
Saltelli, A., Ratto, M., Andres, T., 2008. Global sensitivity analysis: the primer, 1st ed. Hoboken, NJ, USA: Wiley.
Tian, J., Treiber, M., Ma, S., 2015. Microscopic driving theory with oscillatory congested states: model and empirical verification. Transport. Res. Part B 71, 138-157.
Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Phys. Rev. E 62, 1805.
Treiber, M., Kesting, A., 2013. Traffic flow dynamics: data, models and simulation. Springer.
Treiber, M., Kesting, A., 2014. Microscopic calibration and validation of car-following models - a systematic approach. Procedia-Social and Behavioral Sciences 80, 922-939.
Wagner, P., 2012. Analyzing fluctuations in car-following. Transport. Res. Part B 46, 1384-1392.
Wang, H., Wang, W., Chen, J., Jing, M., 2010. Using trajectory data to analyze intradriver heterogeneity in car-following. Transport. Res. Rec., 2188, 85-95.
Zhong, R., Chen, C., Andy, C., Pan, T., Yuan, F., He, Z., 2015. Automatic calibration of fundamental diagram for first-order macroscopic freeway traffic models. J. Adv. Transport., DOI: 10.1002/atr.1334.