# A General Framework for Calibrating and Comparing Car-Following Models

Dr. Ir. C.P.IJ. van Hinsbergen[1,2] (corresponding author)

Phone: +31 6 50810351, Email: chris@fileradar.nl

Ir. W.J. Schakel[2,3]

Phone: +31 15 2784030, Email: w.j.schakel@tudelft.nl

Dr. V.L. Knoop[2]

Phone: +31 15 2788413, Email: v.l.knoop@tudelft.nl

Dr. Ir. J.W.C. van Lint[2]

Phone: +31 15 2783179, Fax: +31 15 2783179, Email: j.w.c.vanlint@tudelft.nl

Prof. Dr. Ir. S.P. Hoogendoorn[2]

Phone: +31 15 2783179, Email: s.p.hoogendoorn@tudelft.nl

[1] Fileradar B.V., Justus van Schoonhovenstraat 19, 2613 NX, Delft, The Netherlands

[2] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, P.O. Box 5048, 2600 GA, Delft, The Netherlands, Fax: +31 15 2783179

[3] TRAIL Research School

## ABSTRACT

Recent research has shown that there exists large heterogeneity in car-following behavior such that different car-following models best describe different drivers' behavior. A literature review reveals that current approaches to calibrate and compare different models for one driver do not take the complexity of the models into account or are only able to compare a specific set of models. This contribution applies Bayesian techniques to the calibration of car-following models. The Bayesian framework promotes models that fit the data well but punishes models with a high complexity, resulting in a measure called the *evidence*. This evidence quantifies how probable each model is to be the model that best describes the car-following behavior of a single driver. It can be computed for any car-following model. When considered over multiple drivers the evidences can be used to describe the heterogeneity of the driving population. In an experiment seven different car-following models are calibrated and compared using a data set that was collected with a helicopter. The results indicate that multi-leader models better describe the car-following models even if their higher complexity is accounted for, and that for the description of microscopic driving behavior the reaction time is essential; models without a reaction time perform significantly worse.

## KEYWORDS

Calibration, car-following model, longitudinal driver behavior, Bayesian evidence, inter-driver differences

## 1. INTRODUCTION

The longitudinal driver behavior in a traffic stream determines for a large part the dynamics of the flow of traffic. This important role of this behavior has resulted in a multitude of mathematical models to predict the longitudinal driving behavior of individuals, such as the CHM model [1], the IDM model [2], the OVM model [3] and the models proposed by Helly [4], Bexelius [5], Gipps [6], Addison & Low [7], Lenz [8] and Tampère [9]. Brackstone and McDonald [10] present a historical review of these and other car-following models.

In recent microscopic traffic modeling research, a number of studies have revealed that there are large *inter-driver differences* in car-following behavior, such that different car-following models may apply to different drivers [11-13]. Additionally, *intra-driver differences* (the fact that individual drivers may change their behavior over the data collection period) can cause some car-following models to produce erroneous predictions during certain episodes of the driver's car-following behavior [14,15]. The effects of such heterogeneity of car-following behavior on the macroscopic properties of traffic are significant [16]. Therefore, when simulating individual vehicles in a software package (a microsimulation), it is an interesting idea to use multiple car-following models. To achieve such a heterogeneous microscopic simulation model, a selection needs to be made from all available car-following models that are available in literature. Ideally, such an identification procedure should take place based on data that is also used for calibration of the models because it is generally expensive to obtain data on microscopic driving behavior. This contribution describes both the calibration and identification process for car-following models.

## 1.1. State of the art in model calibration and comparison

First, an extensive literature study has been carried out to investigate current calibration and model selection methods. Four methods have been identified: using default parameters, using the calibration error, using the validation error or using the Likelihood-Ratio Test. Below, each of these methods is described.

### 1.1.1. Default parameter settings

One study was found where the default manufacturer parameters are used to evaluate the performance of different models [17], even though the authors recognize the importance of the parameter values on the performance of the model. From the literature study that has been performed, it is found that the parameters of the car-following models are not constant for all drivers; both inter- and intra-driver differences cause large differences in the parameter values for individual drivers. Default parameters are therefore not useful for comparing different car-following models.

### 1.1.2. Calibration error approach

The second and most commonly used approach is to select models based on the outcomes of the calibration procedure [12,18-26]. This approach is however criticized here: more complex models (i.e. with a larger parameter space) are promoted by this approach, while more complex models do not necessarily make better predictions. For example, Abbas et al. [26] conclude that the Wiedemann model is the best model for the description of truck driver behavior. However, in the paper the Wiedemann process is modeled using no less than 15 parameters and compared to other models with 5, 6 and 11 parameters. Their conclusion is that “the flexibility available in the Wiedemann model undoubtedly contributes to its performance in mimicking the behavior of real drivers". This is a dubious conclusion. The Wiedemann model is better in mimicking the data, including its noise, but not necessarily the behavior of the drivers that have been measured. In many cases more complex models will *not* make better predictions. This is the concept of overfitting [27]. The use of calibration error as a basis to select models should therefore be rejected.

### *1.1.3. Validation error approach*

Instead of using the same data set to calibrate and compare the models, also a separate data set can be used to make a selection from a set of models [11,28-30]. The validation set approach is a theoretically sound way to compare models in case the validation data set is representative for the phenomenon that one is trying to model. Interestingly, different studies confirm that more complex models do not always perform better [11,28-30]. For example, in a paper by Brockfeld et al. [30] the model with 20 parameters performs significantly worse on the validation set than simpler models, a confirmation of the overfitting problem of overly complex car-following models.

Unfortunately, this approach requires two data sets to be available for one single driver. Such data can usually only be collected under controlled conditions, where the same drivers are asked in an experimental setup to perform the car-following task. This has two major drawbacks: the results of experiments in controlled conditions may not always be portable to a 'real-life' situation, and usually only a small data set is available because of the expenses that have to be made to equip vehicles and to attract participants. More data may be collected monitoring the 'regular' traffic system (using for example cameras or a helicopter), but in those cases usually the data set of each single driver is too small to be split in two. Therefore, a method that allows for all data to be used for calibration, while still preventing overfitted models to be selected, should be preferred.

### *1.1.4. Likelihood-ratio test*

The Likelihood-Ratio-Test (LRT) is a method that allows all data to be used for calibration and model selection in parallel, while still preventing overfitted models to be selected [13,16,31]. More complex models receive a penalty, while models that fit well on data are promoted. This balances the goodness-of-fit to the data with the model complexity. A second benefit of the LRT is that *prior* information can be included. A common problem in calibrating car-following models is that they contain parameters for both the free flow regime as well as the congested regime, but for example only data collected under congested conditions is available. The parameters describing free flow behavior cannot be calibrated using congested data. A 'normal' calibration procedure would cause the 'free-flow-parameters' to be calibrated on coincidences (noise) in the data. Using prior estimates can prevent these parameters to take on unrealistic values due to the presence of noise.

However, the LRT is only valid when used to compare *hierarchically nested models*. This means that the simple model must be a special case of the more complex model by setting one or multiple parameters to zero. Therefore, Hoogendoorn et al. [31] first formulate a general equation in which several car-following models can be fit. However, as there is a large variation in types of car-following models that have been developed over the years, and these models may not all be fit into one general equation, the LRT is not a generic approach.

Therefore, in this paper a novel method is proposed to calibrate and compare car-following models. The method is based on Bayes' theorem. It is an extension to the LRT method, and allows for the comparison of *any* car-following model.

### *1.2. Structure of this paper*

This paper is structured as follows. In the Methodology section the Bayesian approach to calibrate and compare car-following models is described. In the Experiment section a description is given of how the framework can be applied to a variety of car-following models. The Results section then shows the results of this application. Finally a discussion, a conclusion and recommendations are presented.

### *1.3. List of Symbols*

Throughout this contribution the following symbols will be used:

| | |
|---|---|
| $\boldsymbol{\theta}$ | parameter vector |
| $\bar{\boldsymbol{\theta}}$ | mean of prior distribution of parameters |
| $\boldsymbol{\theta}^{MP}$ | mean of posterior distribution of parameters |
| $\Sigma$ | covariance matrix of prior distribution of parameters |
| $\Theta$ | covariance matrix of posterior distribution of parameters |
| $K$ | total number of time steps in data set D |
| $k$ | one specific time step |
| $N$ | number of parameters |
| $M$ | number of models to be compared |
| $D$ | data set used for calibration and comparison |
| $\sigma_l$ | standard deviation of data |
| $Z_p$ | normalizing constant in posterior distribution of parameters |
| $H_q$ | total set of assumptions made for model q |

## 2. METHODOLOGY

The Bayesian approach is a generalization of the Likelihood-Ratio Test [13]. This approach has several advantages over existing mechanisms:

1. The most important feature is that it leads to a probabilistic approach to *compare* different models on the basis of posterior distributions of their parameters, without the need to split the available dataset in two. This allows a modeler to select the model that most probably best describes a certain driver's behavior, taking into account both the *calibration error* as well as the *model complexity*. The main contribution of the Bayesian approach compared to the LRT approach is that any model can be used;
2. Just as with the LRT approach *prior information* can be included when calibrating the parameters of car-following models to rule out unrealistic estimation results due to the fact that too little information is present on certain parameters within data;
3. Prediction intervals can be constructed on the predictions of the car-following models.
4. The Bayesian framework can be used to combine the predictions of several models in a so-called *committee* or *ensemble* of models. In a committee several models are all used to make predictions about one single driver. Empirical evidence shows that if the (weighted) average of these predictions is used, the prediction accuracy can be improved [32-36]. Furthermore, the accuracy of the prediction intervals can be improved too.

Below, the framework will be described. Note that the framework is closely related to the Bayesian Information Criterion (BIC) [37], which is a similar approach that also penalizes overly complex models. However, in BIC the benefits 2-4 mentioned above are not automatically included; in order to include prior information, to construct a committee or to find prediction intervals separate efforts would need to be taken. The framework presented below is ‘all-in-one’. In this paper, committees and prediction intervals are not constructed because the main goal is the calibration and comparison of car-following models. For more information on how these would be constructed, see [33,38].

## *2.1. The Bayesian Framework for Calibration and Comparison*

In this section the general framework for calibrating and comparing car-following models will be unveiled. In section 2.2 the solutions to the equations will be presented.

Define a parameter vector $\boldsymbol{\theta}_q = (\theta_{q,1},\ldots,\theta_{q,N})$ which contains all $N$ parameters of a certain car-following model $q$, and define the data set that will be used for the calibration procedure $D$. This data set contains for example positions (lateral and longitudinal) and speeds of different vehicles, from which car-following models can be calibrated. Finally, define $H_q$ to be the total set of assumptions that model $q$ makes, i.e. the mathematical structure of the model, the number and type of parameters the model contains, the underlying theoretical assumptions, et cetera.

Considering the variations of car-following behavior due to inter- and intra-driver differences and the corresponding uncertainty of parameter values, it makes sense to consider the parameters of car-following models as *distributions* rather than as singular values. In probabilistic terms, the result of the calibration procedure can then be written as $p(\boldsymbol{\theta}_q|D,H_q)$, e.g. the probability density function of the parameters $\boldsymbol{\theta}_q$ of model $q$ given the data set $D$ and given the assumptions made for model $q$. Bayes’ rule can be applied to find an expression for this posterior:

$$p(\boldsymbol{\theta}_q \mid D, H_q) = \frac{p(D \mid \boldsymbol{\theta}_q, H_q)\, p(\boldsymbol{\theta}_q \mid H_q)}{p(D \mid H_q)} \tag{1}$$

where $p(D|\boldsymbol{\theta},H_q)$ represents the distribution of noise on the data (which could be measurement noise due to imprecise measurement equipment, or system noise, i.e. behavior of drivers that results in variations but which is not measured and therefore not represented in the data) and corresponds to the likelihood function, $p(\boldsymbol{\theta}|H_q)$ is a prior probability of the parameters, which represents our prior knowledge of possible values for each parameter in our model, and $p(D|H_q)$ is a normalization factor. Later, it will be shown how to solve equation (1).

Consider now a set of $M$ different car-following models each with its own set of assumptions $H_q$, where $q \in (q_1, q_2, ..., q_M)$ These models can be compared in how well they describe the car-following behavior of a certain driver using the outcomes of the calibration procedure. For this, consider the posterior probability of an entire model $q$ denoted by $P(H_q|D)$. This probability expresses a belief in model $q$ after we have applied dataset $D$ for its calibration. An expression for this probability can be found by applying Bayes' rule once more:

$$P(H_q \mid D) = \frac{p(D \mid H_q) P(H_q)}{p(D)} \tag{2}$$

The term $P(H_q)$ represents the prior probability of the model $q$ which is chosen by the user before the calibration and comparison of models starts. As the denominator of (2) is independent of the models $H_q$, the only unknown in (2) is the term $p(D|H_q)$, which is called the *evidence* for model $q$ [38]. This evidence can be recognized as the denominator of (1). The expressions derived during the calibration procedure can therefore be used to derive expressions for the evidence for the entire model. From (1) the evidence can be written in the form

$$p(D|H_q) = \int p(D|\boldsymbol{\theta}_q, H_q)\, p(\boldsymbol{\theta}_q|H_q)\, d\boldsymbol{\theta}_q \qquad (3)$$

where the conditional dependence on the model assumptions $H_q$ is made explicit and where $\boldsymbol{\theta}_q$ is the parameter vector for model $q$. Since (3) would require integration (marginalization) over the entire parameter space, calculating it analytically is only possible in case of very simple models, and even then requires elaborate calculations. Therefore, an approximation needs to be made for which a technique is proposed by MacKay [38]. The required assumptions may or may not be valid, but the solution presented by MacKay at least has lead to many successful applications of the Bayesian framework in various fields of application such as travel time prediction, prognosis after breast cancer surgery, fraud detection and economics (see for example [39-44]). If it is assumed that both the likelihood and the prior of the parameters are Gaussian distributed (in which case the posterior also becomes Gaussian), and if it is assumed that the posterior distribution of the parameters is sharply peaked around its maximum, the evidence can be approximated as the value at this maximum times the width of the peak, which in the multivariate case leads to the expression

$$p(D|H_q) \simeq \underbrace{p\left(D|\boldsymbol{\theta}_q^{MP}, H_q\right)}_{\text{Best fit likelihood}} \underbrace{p\left(\boldsymbol{\theta}_q^{MP}|H_q\right) \times \det^{-1/2}\left(\mathbf{A}(\boldsymbol{\theta}_q^{MP})/2\pi\right)}_{\text{Occam factor}} \qquad (4)$$

where $\boldsymbol{\theta}_q^{MP}$ equals the *most probable* parameter vector for model $q$, i.e. the parameter vector that describes the maximum of the posterior of $p(\boldsymbol{\theta}_q|D,H_q)$, which in this case equals the mean of the distribution because of the Gaussian assumptions. $\mathbf{A}(\boldsymbol{\theta}_q^{MP})$ is the Hessian matrix which is used to estimate the width of the distribution and which is evaluated around the maximum $\boldsymbol{\theta}_q^{MP}$. The evidence of (4) can be interpreted as consisting of two elements, the best fit likelihood and the *Occam factor*, as indicated in (4) below the equation. A higher best fit likelihood favors models that can explain the data well, i.e. that have a low calibration error. However, if only this would be investigated the overfitting problem would occur as when the calibration error is used for model selection (see 1.1.2). Therefore, the model's performance is penalized by the Occam factor, which is always smaller than 1 and is named after Occam's Razor [45]. A complex model with many parameters, each of which is free to vary over a large range (which is reflected by the prior) will typically be penalized by a larger Occam factor than a simpler model. The Occam factor also penalizes models which have to he finely tuned to fit the data: models for which the required precision of the parameters is coarse are favored [38,46]. The evidence thus naturally reflects the trade-off between a good fit and overfitting. Extensive literature is available on the importance of this trade-off and other features of the evidence [33,38,39,47-50].

Using the evidence, the posterior probability of the entire model can also be expressed. An important assumptions then needs to be made: a *closed world* is assumed, i.e. the models that are to be compared are considered to be the *only possible* models for explaining car-following behavior, such that $P(\emptyset) = 0$. The fact that the 'correct' model for the description of the car-following behavior may not be in the list of models at all is ignored. As MacKay notes, inference is normally open ended: in the scientific process new models will be tested or developed to account for the data that have been gathered [51]. Nevertheless, the closed world assumption here aids to express posterior probabilities for each of the considered models, which are meaningful in comparison to each other. Under this assumptions the normalization factor of (2) can be written as

$$P(D)=\sum_q P(H_q)P(D\,|\,H_q) \tag{5}$$

so that the posterior probability of one model $q_m$ equals

$$P(H_{q_m}\,|\,D)=\frac{P(H_{q_m})P(D\,|\,H_{q_m})}{\sum_q P(H_q)P(D\,|\,H_q)} \tag{6}$$

Aggregated over all drivers, this mechanism can be used to see how well a model performs relative to the other models for a group of drivers. By taking the mean of (6) over all individual drivers, an expression is found for the probability of a model compared to the probability of all used models for an entire driver population.

## *2.2. Computing the evidence*

In (4) an expression was found for the evidence, the main component in model comparison. In this section, equations will be introduced in order to compute this evidence.

As stated before, the prior and the likelihood are assumed to be Gaussian in order to be able to approximate the evidence. Define the prior probability as a multivariate Gaussian with mean $\overline{\boldsymbol{\theta}}$ and covariance matrix Σ, both of which are user-defined:

$$p(\boldsymbol{\theta})=\frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right)^T\Sigma^{-1}\left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right)\right)}{(2\pi)^{N/2}\left|\Sigma\right|^{1/2}} \tag{7}$$

where $N$ equals the number of parameters of the model. If it is assumed that the noise of the data is Gaussian distributed as well with mean zero and standard deviation $\sigma_l$, the likelihood function $p(D|\boldsymbol{\theta})$ can be defined as a univariate Gaussian [13]:

$$p(D\,|\,\boldsymbol{\theta})=\frac{\exp\left(-\frac{1}{2\sigma_l^2}\sum_{k=1}^{K}\left(v_{pred}(k,\boldsymbol{\theta})-v_{obs}(k)\right)^2\right)}{\left(\sigma_l^2 2\pi\right)^{K/2}} \tag{8}$$

Where it is chosen to calibrate the models on the difference between the predicted speed $v_{pred}(k,\boldsymbol{\theta})$ and the measured speed $v_{obs}(k)$ over a series of time instants $k = (1,\ldots K)$. Substituting (7) and (8) into (1) results in an expression for the posterior distribution of the parameters:

$$\begin{aligned} p(\boldsymbol{\theta} \mid D) = & \frac{1}{Z_p \sigma_l^K |\Sigma|^{1/2}} \exp\left( -\frac{1}{2\sigma_l^2} \sum_{k=1}^{K} \left( v_{pred}(k,\boldsymbol{\theta}) - v_{obs}(k) \right)^2 \right) \\ & \cdot \exp\left( -\frac{1}{2} \left( \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \right)^T \Sigma^{-1} \left( \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \right) \right) \end{aligned} \tag{9}$$

where $Z_p$ is a constant.

In order to evaluate (4) the maximum of this posterior is required ($\boldsymbol{\theta}^{MP}$), which can be found by maximizing the logarithm of (9):

$$\begin{aligned} \left( \boldsymbol{\theta}^{MP} \right) &= \arg\max_{\boldsymbol{\theta}} \ln\left( p(\boldsymbol{\theta} \mid D) \right) \\ &= \arg\max_{\boldsymbol{\theta}} \left( -E(\boldsymbol{\theta}) \right) \\ &= \arg\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \end{aligned} \tag{10}$$

where $E(\boldsymbol{\theta})$ is defined as

$$\begin{aligned} E(\boldsymbol{\theta}) &= K \ln\left( \sigma_l \right) + E_p(\boldsymbol{\theta}) + E_l(\boldsymbol{\theta}) \\ E_p(\boldsymbol{\theta}) &= \frac{1}{2} \left( \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \right)^T \Sigma^{-1} \left( \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \right) \\ E_l(\boldsymbol{\theta}) &= \frac{1}{2\sigma_l^2} \sum_{k=1}^{K} \left( v_{pred}(k,\boldsymbol{\theta}) - v_{obs}(k) \right)^2 \end{aligned} \tag{11}$$

Notice that in (11) the expressions resulting from $Z_p$ and $|\Sigma|^{1/2}$ have been omitted as these do not depend on $\boldsymbol{\theta}$ and therefore do not influence the minimization of the error function. For the minimization many different approaches can be chosen [52,53]. In this paper the Matlab toolbox *fminunc* is used, which uses the BFGS Quasi-Newton method [54]. Finally note that although a distribution of parameter values is found for one driver, during simulation the parameters are held constant for one driver.

One unnatural artifact of the Gaussian assumption for the parameters is that negative values are then considered possible while many parameters of the car-following models can never be negative, such as reaction times, stopping distances, et cetera. A heuristic solution for such parameters is not to assume a normal distribution but a *log-normal* distribution. In the optimization algorithm the natural logarithm of the parameter is optimized rather than the parameter itself. The resulting parameter value can then never be negative.

For the evaluation of (4) the Hessian **A** is also needed. This is approximated numerically using the Matlab DERIVEST toolbox. For a detailed description of this toolbox, see [55].

Finally, for the evaluation of the evidence a value for the standard deviation of the likelihood function $\sigma_l$ needs to be found, for which the derivative $\partial E(\boldsymbol{\theta})/\partial \sigma_l$ is set to zero. This leads to

$$\sigma_l^2(\boldsymbol{\theta}) = \frac{1}{K}\sum_{k=1}^{K}\left(v_{pred}(k,\boldsymbol{\theta}) - v_{obs}(k)\right)^2 \qquad (12)$$

## 3. EXPERIMENT

In the previous section a framework was presented that allows any car-following model $q$ to be calibrated and compared to other models. The first step of this procedure is the calibration for which any optimization method that the user chooses can be used, leading to a posterior distribution of the parameters of the model $q$ denoted by $p(\boldsymbol{\theta}|D,H_q)$. Using this distribution, an expression can be obtained for the so-called *evidence* for model $q$ denoted by $p(D|H_q)$. By comparing the evidences for models multiplied by the user-defined prior probability $P(H_q)$ of model $q$, the posterior probability for model $q$ denoted by $P(H_q|D)$ can be found.

In order to test the framework an experiment is conducted where seven different car-following models are calibrated and compared using the Bayesian framework. First, a brief description of the used models is given.

### *3.1. Car-following models to be compared*

In this paper a total of seven different car-following models of different complexity are compared: four single-leader models (CHM, Helly, OVM and IDM) and three multi-anticipation models (Generalized Helly, Lenz and HDM).

A data set of leading and following vehicles was obtained (see section 3.2). Each of the models is used to predict the acceleration of the following vehicle at fixed time steps. Given the speed of the following vehicle in the previous time step and the predicted acceleration of the vehicle, the new speed of the vehicle can be predicted by

$$v_{pred}(t,\boldsymbol{\theta}) = v_{obs}(t-\Delta t) + a(t-\Delta t,\boldsymbol{\theta})\Delta t \qquad (13)$$

which is used in (11) to compute the error for the model.

In Table 1 a list of all models is given. In the first column, below the model name, references are included from which equations were obtained. In the second column the equations to determine the acceleration are given. In the equations $\Delta v_{ji}(t) = v_j - v_i$ equals the speed difference between follower $i$ and leader $j$, $\Delta v_{ij}(t) = v_i - v_j$ equals the approaching rate, $\Delta x_n(t)$ the net headway distance between a leader and a follower at time $t$, $\Delta x_g(t)$ the gross headway distance and $\Delta x_{nn}(t)$ the net headway minus the sum of vehicle lengths of all vehicles in between. The third column provides a description of the parameters of the model.

**Table 1 Overview of the used models. References are given to the papers from which the equations were obtained.**

| Model | Equation | Parameters |
|---|---|---|
| CHM [1] | $a(t,\boldsymbol{\theta}) = \gamma \Delta v_{ji}(t-\tau)$ | $\gamma$ response parameter (1/s)<br>$\tau$ reaction time (s) |
| Helly [4] | $a(t,\boldsymbol{\theta}) = \alpha \Delta v_{ji}(t-\tau) + \beta\left\{\Delta x(t-\tau) - \Delta x^{des}\left(v(t-\tau)\right)\right\}$<br>$\Delta x^{des}(v) = x_0 + Tv$ | $\alpha$ response parameter (1/s)<br>$\beta$ response parameter (1/s$^2$)<br>$x_0$ stopping distance (m)<br>$T$ minimum time headway (s)<br>$\tau$ reaction time (s) |
| OVM [56] | $a(t,\boldsymbol{\theta}) = \dfrac{v_{opt}\left(\Delta x_n(t)\right) - v(t)}{\tau_v}$<br>$v_{opt}(\Delta x_n) = \dfrac{v_0}{2}\left\{\tanh\left(\dfrac{\Delta x_n}{l_{\text{int}}} - \beta_s\right) - \tanh(-\beta_s)\right\}$ | $v_0$ desired velocity (m/s)<br>$\tau_v$ speed relaxation time (s)<br>$l_{\text{int}}$ interaction length (m)<br>$\beta_s$ shape factor (-) |
| IDM [2] **a)** | $a(t,\boldsymbol{\theta}) = a^{\max}\left(1 - \left(\dfrac{v(t)}{v_0}\right)^4 - \left(\dfrac{\Delta x^*(t)}{\Delta x_n(t)}\right)^2\right)$<br>$\Delta x^*(t) = s_0 + v(t)T + \dfrac{v(t)\Delta v_{ij}(t)}{2\sqrt{a^{\max}b}}$ | $a^{\max}$ maximum acceleration (m/s$^2$)<br>$b$ comfortable deceleration (m/s$^2$)<br>$s_0$ net stopping distance (m)<br>$T$ minimum time headway (s)<br>$v_0$ desired velocity (m/s) |
| Generalized Helly-3-1 **b)** [57] | $a(t,\boldsymbol{\theta}) = \sum_{j=1}^{m_1}\alpha_j \Delta v_{ji}(t-\tau) + \sum_{j=1}^{m_2}\beta_j\left\{\Delta x_g(t-\tau) - \Delta x_j^{des}\left(v(t-\tau)\right)\right\}$<br>$\Delta x_j^{des}(v) = x_0 + j\cdot Tv$ | $\alpha_1$ response parameter (1/s)<br>$\alpha_2$ response parameter (1/s)<br>$\alpha_3$ response parameter (1/s)<br>$\beta_1$ response parameter (1/s$^2$)<br>$x_0$ gross stopping distance (m)<br>$T$ minimum time headway (s)<br>$\tau$ reaction time (s) |
| Lenz-2 **c)** [57] | $a(t,\boldsymbol{\theta}) = \sum_{j=1}^{m}\kappa_j\left\{V\left(\dfrac{\Delta x_{g,j}(t-\tau)}{j}\right) - v(t-\tau)\right\}$<br>$V(\Delta x) = v_0\left(\left\{1 + \exp\left(\dfrac{1000}{\gamma_s \Delta x} - \dfrac{10}{2.1}\right)\right\}^{-1} - 5.34\cdot 10^{-9}\right)$ | $\kappa_1$ response parameter (1/s)<br>$\kappa_2$ response parameter (1/s)<br>$\gamma_s$ response parameter (1/m)<br>$v_0$ desired velocity (m/s)<br>$\tau$ reaction time (s) |
| HDM (adapted) **a/d)** | $a(t,\boldsymbol{\theta}) = a^{\max}\left\{1 - \left(\dfrac{v(t-\tau)}{v_0}\right)^4\right\} - a^{\max}\sum_{j=1}^{m}\left\{\left(\dfrac{\Delta x_j^*(t-\tau)}{\Delta x_{nn,j}(t-\tau)}\right)^2\right\}$<br>$\Delta x_j^*(t) = s_0 + v(t)T + \dfrac{v(t)\Delta v_{ij}(t)}{2\sqrt{a^{\max}b}}$ | $a^{\max}$ maximum acceleration (m/s$^2$)<br>$b$ comfortable deceleration (m/s$^2$)<br>$s_0$ net stopping distance (m)<br>$T$ minimum time headway (s)<br>$v_0$ desired velocity (m/s)<br>$\tau$ reaction time (s) |

**a)** *As was described in 2.2 all parameters are kept positive assuming a log-normal distribution for equation* (10). *This approach can however not be used to calculate the Hessian. The Hessian is numerically approximated assuming Gaussian distributions. Both the IDM and HDM take the square root of $a^{max}b$, which* *would give imaginary values if either is negative. To prevent this, both parameters are limited to a value of 0.01 only during the approximation of the Hessian, and only within the square root.*

**b)** *The speed difference with $m_1$ leaders is considered and the distance to $m_2$ leaders is considered. In this paper $m_1 = 3$ and $m_2 = 1$ as this model has the highest number of best fits over several other combinations in a paper by Hoogendoorn et al. [57]. Although an important conclusion of Hoogendoorn et al. is the fact that different values for $m_1$ and $m_2$ may apply to different drivers and to different situations (such as being behind a truck and being unable to look ahead), the integer nature of these parameters prevents us from including them directly in the calibration. Consequently, this paper discusses the GH-3-1 model and not the GH model in general.*

**c)** *Similar as in Hoogendoorn et al. [57], m = 2 is used, taking two leaders into account.*

**d)** *The HDM is a general framework that is applied here to the IDM. The HDM includes (1) finite reaction time, (2) multi-anticipation, (3) estimation errors and (4) temporary anticipation. Both (1) and (3) introduce instability while (2) and (4) introduce stability. A difficulty with (3) is that it uses a stochastic auto-correlation process. Estimating the state of the estimation error over time additional to estimating the parameters would surely result in an overfitted model with large uncertainty of the parameters, as the estimation error can absorb modeling errors. It is therefore chosen to exclude (3) from the HDM. Additionally (4) is excluded as well to balance the stabilizing factors.*

For car-following models with reaction time $\tau$, quantities $z$ (position, speed, acceleration) are linearly interpolated based on their nearest neighbors if the reaction time is not an integer multiple of the data time step:

$$z_{obs}\left(t-\Delta t-\tau\right)=\frac{t_2-t}{\Delta t}z_{obs}\left(t_1\right)+\frac{t-t_1}{\Delta t}z_{obs}\left(t_2\right) \tag{14}$$

where $t_1$ is the first time step before $t - \Delta t - \tau$ and $t_2$ the first after.

For each of the models prior distributions need to be defined. These prior distributions are important to contain the parameters within in a logical range in case there is little information in the data about the parameter. In case there is information about a parameter, the likelihood-term of the error function (11) will become more dominant, allowing the parameter to move away from its prior.

In order to obtain prior values a literature study was performed. Other calibration studies have been used and the outcomes of parameter values of these studies were used as a starting point for the calibration procedure of this paper. Variances reported in the papers were used to formulate a covariance matrix. In case no variances were available for parameters a large variance was used to reflect the uncertainty of the prior. The framework will then (need to) place more trust on the data. Note that no covariance between parameters is considered because numerical values could not be found for all models in literature. Inclusion of covariance between parameters could lead models with better predictions [58], but ignoring covariance does not prevent the illustration of how the framework can be applied. In Table 2 the chosen values are given, as well as the relevant references from which these values were obtained. Note that 'diag' stands for a diagonal matrix with the given values on the diagonal.

**Table 2 Prior values for the parameter values of the car-following models**

| Model | Parameters |
|---|---|
| CHM<br>[1,10,12] | $\bar{\boldsymbol{\theta}} = (\bar{\gamma}, \bar{\tau}) = (0.3, 1.6)$<br>$\Sigma = \text{diag}(0.2^2, 0.4^2)$ |
| Helly<br>[4,12] | $\bar{\boldsymbol{\theta}} = (\bar{\alpha}, \bar{\beta}, \bar{x}_0, \bar{T}, \bar{\tau}) = (0.3, 0.08, 20, 1, 1.2)$<br>$\Sigma = \text{diag}(0.3^2, 0.1^2, 6^2, 0.6^2, 0.9^2)$ |
| OVM<br>[56] | $\bar{\boldsymbol{\theta}} = (\bar{v}_0, \bar{\tau}_v, \bar{l}_{\text{int}}, \bar{\beta}_s) = (16, 1.4, 7, 2.5)$<br>$\Sigma = \text{diag}(6^2, 0.7^2, 9^2, 1.2^2)$ |
| IDM<br>[59] | $\bar{\boldsymbol{\theta}} = (\bar{a}^{\max}, \bar{b}, \bar{s}_0, \bar{T}, \bar{v}_0) = (1, 0.5, 7, 1, 28)$<br>$\Sigma = \text{diag}(0.2^2, 0.2^2, 3^2, 0.2^2, 2^2)$ |
| Generalized Helly-3-1<br>[57] | $\bar{\boldsymbol{\theta}} = (\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3, \bar{\beta}_1, \bar{x}_0, \bar{T}, \bar{\tau}) = (0.3, 0.07, 0.07, 0.06, 20, 1, 1.2)$<br>$\Sigma = \text{diag}(0.3^2, 0.1^2, 0.1^2, 0.08^2, 6^2, 0.6^2, 0.3^2)$ |
| Lenz-2<br>[57] | $\bar{\boldsymbol{\theta}} = (\bar{\kappa}_1, \bar{\kappa}_2, \bar{\gamma}_s, \bar{v}_0, \bar{\tau}) = (0.2, 0.15, 7, 32, 1)$<br>$\Sigma = \text{diag}(0.2^2, 0.2^2, 7^2, 7^2, 0.4^2)$ |
| HDM (adapted)<br>[59,60] | $\bar{\boldsymbol{\theta}} = (\bar{a}^{\max}, \bar{b}, \bar{s}_0, \bar{T}, \bar{v}_0, \bar{\tau}) = (1, 0.5, 7, 1, 28, 1)$<br>$\Sigma = \text{diag}(0.2^2, 0.2^2, 3^2, 0.2^2, 2^2, 0.7^2)$ |

Finally, a prior for the entire model needs to be chosen. In this paper, it is chosen to assign equal priors to all $M$ models, so that $P(H_q) = P(H) \forall q \in \left(q_1, q_2, ..., q_M\right)$.

### 3.2. Data collection

For the calibration and comparison a vehicle trajectory data set is available for the A2 motorway in The Netherlands, near the city of Utrecht, which was collected using a helicopter [61]. The traffic state at the data collection period was congested so that the cars were mainly in car-following mode. The data covers approximately 500m of motorway stretch; the data interval is 0.1s.

Both reaction time and multi-anticipation put demands on the evaluated data. A reaction time requires some history before the first considered time step and the multi-anticipation requires a set of vehicles larger than two. The reaction time is limited to a maximum of $\tau_{\max} = 2$s and the number of leaders is limited to a maximum of three, requiring a set of four vehicles (quadruplet). The data has been filtered to find episodes that obey the following constraints:

1. The quadruplet composition should be fixed during the episode and where each of the four vehicles did not change lanes.
2. The effective episode duration (excluding $\tau_{\max}$) should at least be 15s.
3. The follower, i.e. the last vehicle of the quadruplet, should experience a speed change of at least 5 m/s within the episode to include some level of car-following dynamics.

These constraints are similar as in [62]. The result is a set of 120 valid episodes (following drivers) for calibration.

The posterior distributions of the parameters of the models were found after which the posterior of the model was calculated for each model for each driver. Note that the *log evidence* is used, as the denominator of (8) is taken to the power of $K$, which means that the likelihood becomes very large if $\sigma_l<1$ and very small if $\sigma_l>1$ in case $K»1$. Given that the number of measurements and predictions is in the order of 150 to 350 for each driver, the log of the evidence is used to prevent numerical errors in the computations.

## 4. RESULTS

The results of the calibration are shown in Table 3. From this table the following observations can be made:

1. All multi-anticipative models perform better than their single-leader counterpart. In fact, Generalized Helly-3-1, which is the most complex model (with additional parameters for multiple leaders), performs best. Despite the lower Occam factor for these model due to their higher complexity the evidence indicates that multi-anticipation models are to be preferred. Apparently, multi-anticipation is a significant aspect of driver behavior.
2. Models without reaction time (OVM & IDM) perform very poor. At least on a microscopic scale, the reaction time is apparently a significant aspect of car-following. Remarkably, this is not necessarily true on macroscopic scale [60].
3. Models that include operations in the free regime (OVM, IDM, Lenz & HDM) have lower evidence than models that only describe constrained driving (CHM & Helly models). The data that was used for calibration and comparison mainly contains the car-following regime. Although the priors on the parameters prevent the calibration procedure to result in unreasonable values for the parameters describing the free

regime, these parameters still contribute to the Occam factor. For the description of car-following behavior in congested conditions only, models that include a free regime contain unnecessary complexity, leading to lower evidence.

**Table 3 Overview of all models and results of the evidence calculation for all 120 drivers**

| **Model** | **Leaders** ($v$) | ($x$) | **Parameters** | **Reaction time** | **Free regime** | **P(H\|D)** |
|---|---|---|---|---|---|---|
| CHM | 1 | - | 2 | yes | no | 9.8 % |
| Helly | 1 | 1 | 5 | yes | no | 18.9 % |
| OVM | - | 1 | 4 | no | yes | 0.3 % |
| IDM | 1 | 1 | 5 | no | yes | 0.0 % |
| Gen. Helly-3-1 | 3 | 1 | 7 | yes | no | 61.1 % |
| Lenz-2 | - | 2 | 5 | yes | yes | 5.5 % |
| HDM (adapted) | 3 | 3 | 6 | yes | yes | 4.4 % |

**Table 4 Optimal parameter values that result from the calibration. Parameters within one column with equal shading have an equal physical meaning**

| **Model** | **Optimal parameter value (mean±standard deviation)** | | | | | | |
|---|---|---|---|---|---|---|---|
| CHM | $\gamma$ 0.43±0.18 | | | | | | $\tau$ 1.36±0.45 |
| Helly | $\alpha$ 0.34±0.20 | | | $\beta$ 0.088±0.08 | $x_0$ 16.91±6.70 | $T$ 1.37±0.69 | $\tau$ 1.08±0.46 |
| OVM | $\tau_v$ 3.42±1.25 | $l_{\text{int}}$ 15.80±9.38 | $\beta_s$ 2.04±1.10 | $v_0$ 23.27±6.15 | | | |
| IDM | $a^{\max}$ 0.84±0.37 | $b$ 0.99±0.34 | | $v_0$ 28.44±2.09 | $s_0$ 10.84±4.94 | $T$ 0.81±0.38 | |
| Gen. Helly-3-1 | $\alpha_1$ 0.21±0.16 | $\alpha_2$ 0.084±0.079 | $\alpha_3$ 0.064±0.074 | $\beta_1$ 0.077±0.063 | $x_0$ 19.58±6.38 | $T$ 1.18±0.66 | $\tau$ 1.18±0.34 |
| Lenz-2 | $\kappa_1$ 0.17±0.13 | $\kappa_2$ 0.14±0.10 | $\gamma_s$ 10.03±4.06 | $v_0$ 28.17±8.88 | | | $\tau$ 0.87±0.51 |
| HDM (adapted) | $a^{\max}$ 0.93±0.40 | $b$ 1.19±0.36 | | $v_0$ 28.58±1.76 | $s_0$ 8.55±4.17 | $T$ 0.65±0.36 | $\tau$ 1.20±0.44 |

In Table 4 the parameter distributions over all 120 drivers (not to be confused with the individual posterior distributions) are given by the mean and standard deviation. It is interesting to note that parameters with equal physical meaning have similar values between the models, as can be seen by comparing the values with an equal gray shading within one column of Table 4. Additional plausible relations between models can be found. For example, when the Helly and Generalized Helly are compared, $\alpha \approx \alpha_1 + \alpha_2 + \alpha_3$ (equal total sensitivity to speed difference). A relationship between the parameters of the OVM and the Lenz-2 model can be found in $1/\tau_v \approx \kappa_1 + \kappa_2$ (equal total sensitivity to optimized velocity deviation). Remarkably $\alpha \approx \alpha_1 + \alpha_2 + \alpha_3 \approx 1/\tau_v \approx \kappa_1 + \kappa_2$, indicating that the optimal velocity and the velocity of a leading vehicle approximate one another. However, the optimal velocity is determined using 3 parameters whereas the velocity of the leading vehicle is given. This explains for some part why the OVM and Lenz models perform worse than the (Generalized) Helly model.

Figure 2 shows the distributions of the posterior parameter values for all models over all individual drivers. The two dotted lines indicate the prior mean and posterior mean values. As can be seen from the figures, most of the parameter values for individual drivers fall within a reasonable range from the prior distributions, indicating that the parameter values found in this study are generally close to those found in the literature. For a few parameters this does not hold. For example, parameter $b$ from the IDM has only 5 drivers for which the most probable value falls below the prior mean. For $\tau_v$ of the OVM this number is only 2 and for $b$ of the HDM this number is even zero. Reasons for this could be that the prior for $\tau_v$ was determined in different circumstances (German 1-lane road in 1995) [56] and that the prior $b$ was determined partially with driver simulator data additional to the same dataset as in this study [59], which reduces the value.

It can also been seen from Figure 2 that models with a reaction time show a distribution of $\tau$ with *two* peaks. One peak is generally found near 1.25s while another smaller peak is found near 0.4s. Different explanations can be given: a) drivers have two modes of driving, where the larger reaction time corresponds to a more relaxed driving style in congestion and the lower reaction time a more aggressive driving style, b) the ‘true’ reaction time is the lower one, and the larger reaction time is not adjusted due to a lack of appropriate data in specific episodes, leading to reaction times in the neighborhood of the priors, which fall between 1.0 and 1.6s, c) the lower reaction time is in fact the same reaction time, but is the result from multi-anticipation on a leader that is not included in the calibration. The latter explanation can best be explained by the example as illustrated in Figure 1. The example considers a following vehicle and two leaders. In case leader B suddenly brakes, both the follower as well as leader A will respond with their reaction time $\tau_1$ and $\tau_2$. In case a single-leader model is calibrated to such a situation, the response to the action of leader B will only be noticed through the response of leader A. In case $\tau_2<\tau_1$, this could lead to a small reaction time; both the follower as well as leader A decelerate shortly after each other, which the single-leader model explains by taking a very small reaction time, but which in fact is the result of multi-anticipation. The adopted calibration procedure prevents negative reaction times, so that in case $\tau_2>\tau_1$ this second peak cannot be seen. This phenomenon can also occur for models that do incorporate multi-anticipation due to responses to leaders further away than the maximum number of leaders the model considers.

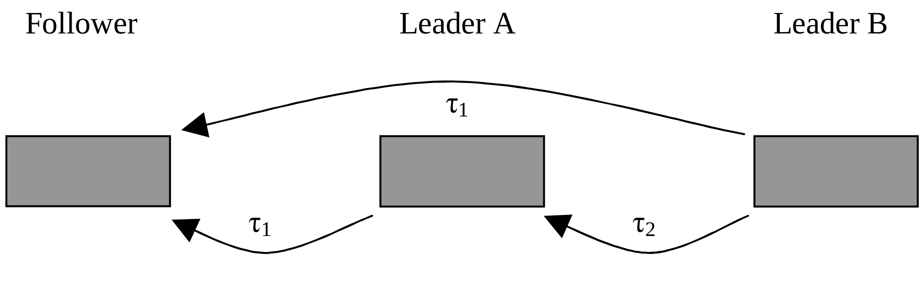
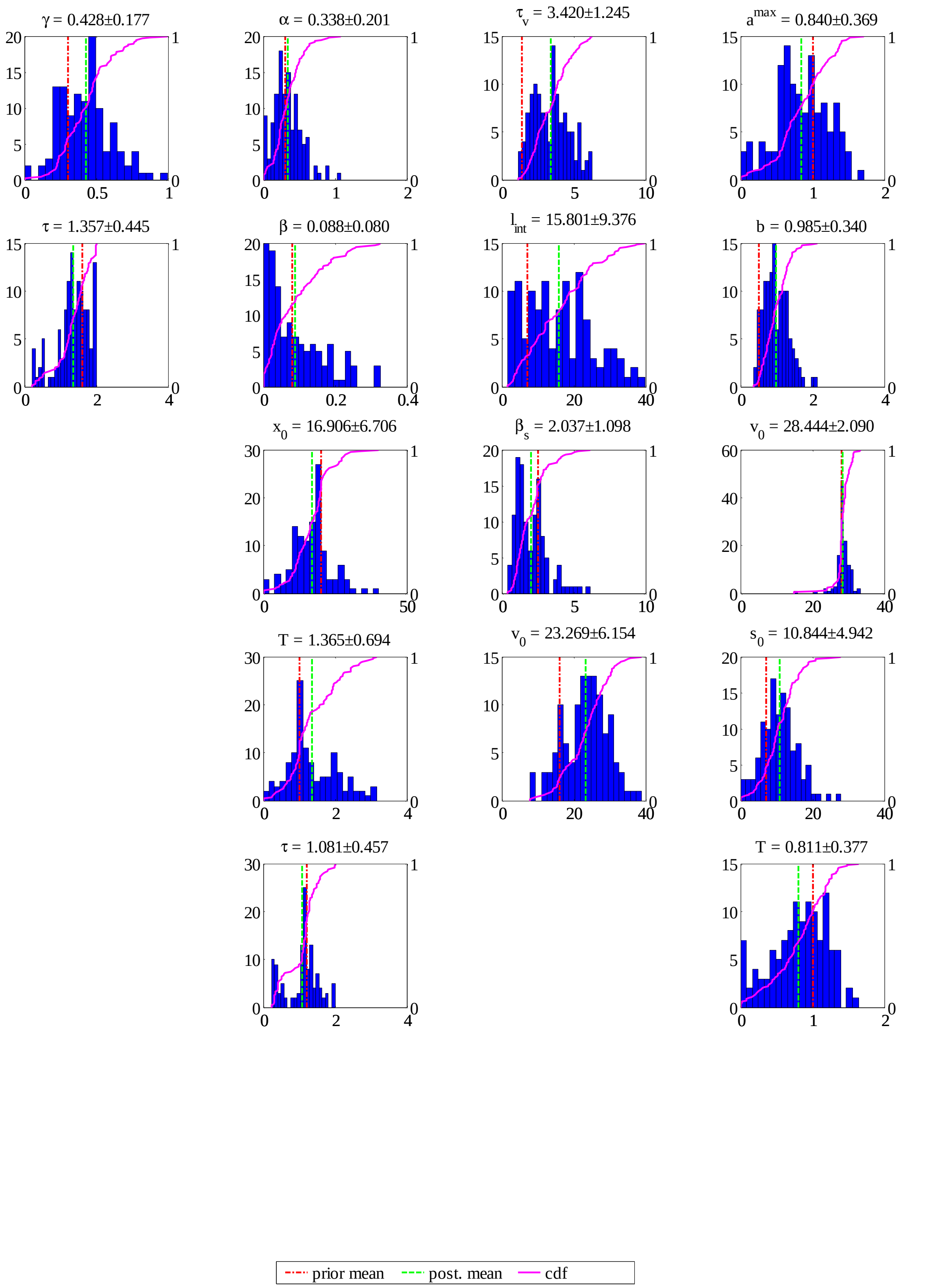
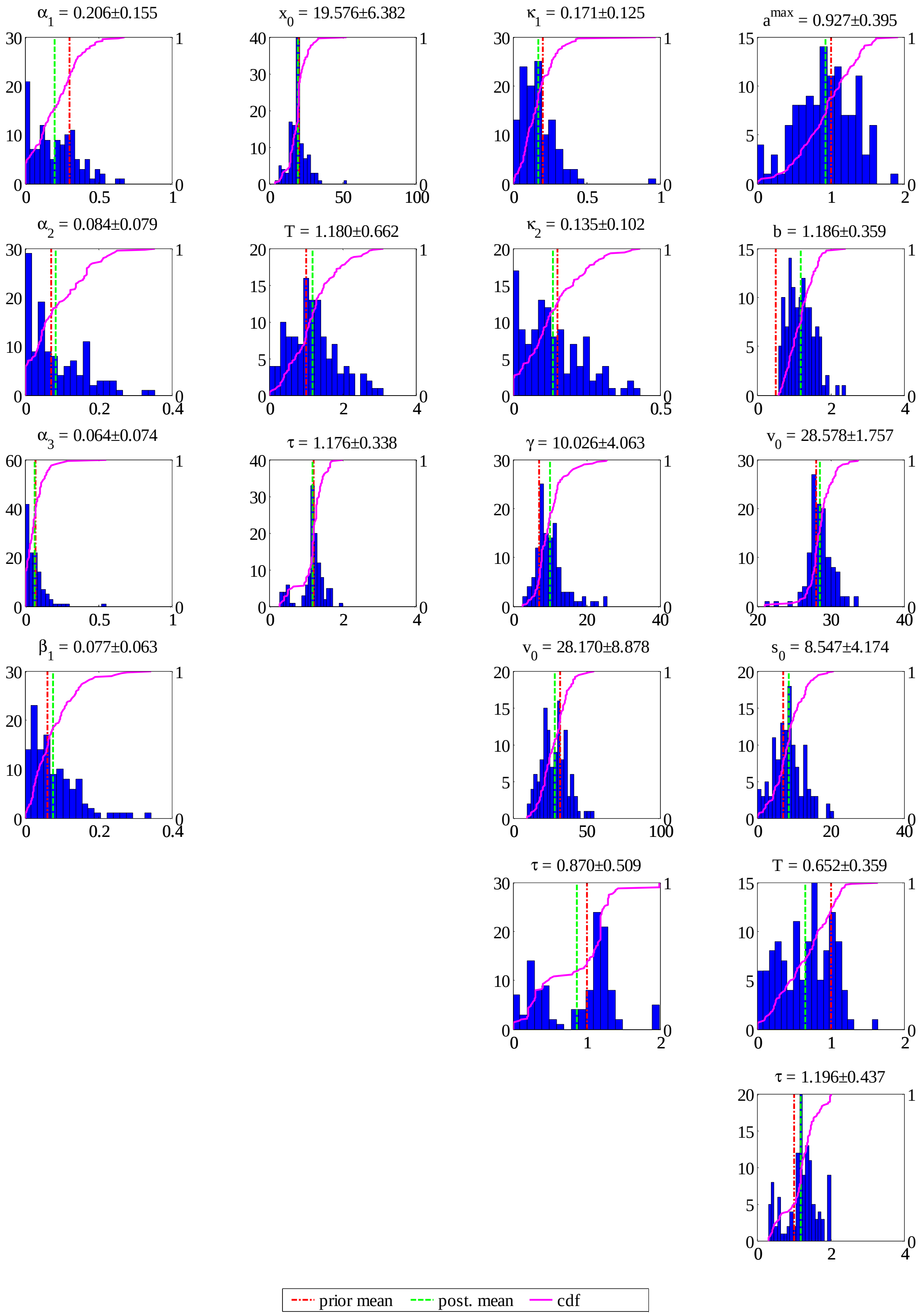



**Figure 1 Example of multi-anticipation of a Follower and two Leaders**




**(a) CHM** **(b) Helly** **(c) OVM** **(d) IDM**




**(e) Generalized Helly-3-1** **(f) Lenz-2** **(g) HDM (adapted)**

**Figure 2 Distributions of the most probable parameter values over all 120 drivers for all models. Axis titles have been omitted for readability; on the vertical axes the number of occurrences is placed, on the horizontal axis the value in the units that are described in Table 1**

## 5. DISCUSSION AND CONCLUSION

In this paper a Bayesian framework has been used for calibrating and comparing car-following models. This framework has several benefits over existing calibration and comparison frameworks that can be found in literature. Many ways of comparing models do not take the complexity of the model into account. The Bayesian framework balances how well a model fits to a calibration data set with the complexity of the model. This prevents overly complex, overfitted models to be chosen. Opposed to alternative frameworks that do take the complexity into account, such as the Likelihood-Ratio Test, *any* car-following model can be compared (or any model in general, the framework is not restricted to car-following models). Furthermore, the comparison can be made on a single data set; no separate validation data set is required. This is a very beneficial feature, as obtaining two different data sets for one driver is expensive. In a different study the use of the evidence was compared to the use of a validation dataset [47], and it was concluded that when datasets are small, the evidence framework should be preferred over the use of a separate validation set because it leads to equally well choices for models but better predictions because more data is available for calibration. Finally, the framework uses prior distributions for parameters, which prevents parameters to take on unrealistic values in case no information is available in the data set on that specific parameter.

In the experiment with seven different car-following models it is illustrated how the framework can be applied. The Bayesian framework is used to express the probability that a model is the correct model for the description of the car-following behavior of an individual driver. Averaged over all drivers, the framework leads to a probability of a model for the description of a group of drivers. In the case study of this paper, the Generalized Helly (3-1) has a probability of 61% of being the correct model (compared to the six other models), while the Helly model (19%), the CHM (10%), Lenz with two leaders (6%) and HDM (4%) are also probable. The OVM and the IDM have a very low probability, which can be explained by their lack of a reaction time. Based on these results, it can be concluded that the reaction time is an important feature in the description of microscopic car-following behavior. Furthermore, all models that include multi-anticipation are more probable than their single-leader counterparts. From this result, it can be concluded that multi-anticipation is a second important feature in car-following behavior.

As the Bayesian framework leads to distributions of both entire models as well as the parameters of models, an interesting possibility emerges: the framework can be used to set up a heterogeneous micro-simulation. In such a simulation, for each driver that is entered into the network first a model is drawn using the probabilities of Table 3. Then, for this model parameters are drawn using the probabilities of Table 4. Note that for the second step, the drawing from the posterior distributions, it is advisable to use the posterior covariance between parameters as well, as research has shown that there exists covariance between parameters of car-following models [58]. Through the Hessian matrix that is computed during the Bayesian analysis, a posterior covariance matrix of the parameters can be obtained. The outcomes of such a heterogeneous simulation are of course uncertain. Future research will need to show any possible benefits and disadvantages.

Future research will also need to investigate the performance of the models and the Bayesian framework in case covariance between parameters is taken into account in the *prior* distributions. This may lead to different parameter distributions and possibly better performance [58]. Furthermore, future research will need to investigate whether the more complex models with a free driving regime perform better (receive a higher probability) on a data set which contains both congested as well as free flow conditions.

Other benefits of the Bayesian approach that have not been illustrated in this study are the possibility to use the evidence to create a committee, and to construct prediction intervals. A committee may improve the description of individual behavior (because it may deal with the intra-driver differences), while the prediction intervals may become useful when predicting the trajectory of a single driver, in for example vehicle-to-vehicle or vehicle-to-roadside architectures. Future studies will need to investigate these benefits of the Bayesian calibration framework.

## ACKNOWLEDGEMENTS

This research was sponsored by the ATMO (www.atmo.tudelft.nl) research program. The data was used at the courtesy of the Traffic Research Centre (AVV) of the Dutch Ministry of Transportation and the project Tracing Congestion Dynamics sponsored by the Dutch Foundation of Scientific Research NWO-MagW.

## REFERENCES

[1] Chandler, R.E., R. Herman and E.W. Montroll: 'Traffic Dynamics: Studies in Car Following', Operations Research, 1958, 6, pp. 165-184

[2] Treiber, M., A. Hennecke and D. Helbing: 'Congested traffic states in empirical observations and microscopic simulations', Physical Review E, 2000, 62, (2), pp. 1805-1824

[3] Bando, M., K. Hasebe, A. Nakayama, A. Shibata and Y. Sugiyama: 'Dynamical model of traffic congestion and numerical simulation', Physical Review E, 1995, 51, (2), pp. 1035-1042

[4] Helly, W.: 'Simulation of bottlenecks in single lane traffic flow'. International symposium on the theory of traffic flow, New York, NY, USA, 1959

[5] Bexelius, S.: 'An extended model for car-following', Transportation Research, 1968, 2, (1), pp. 13-21

[6] Gipps, P.G.: 'A behavioural car-following model for computer simulation', Transportation Research Part B, 1981, 15, (2), pp. 105-111

[7] Addison, P.S. and D.J. Low: 'A novel nonlinear car-following model', Chaos, 1998, 8, (4), pp. 791-799

[8] Lenz, H., C.K. Wagner and R. Sollacher: 'Multi-anticipative car-following model', The European Physical Journal B, 1999, 7, (2), pp. 331-335

[9] Tampère, C.M.J.: 'Human-Kinetic multiclass traffic flow theory and modelling'. PhD thesis, Delft, The Netherlands, 2004

[10] Brackstone, M. and M. McDonald: 'Car-following: a historical review', Transportation Research Part F, 1999, 2, pp. 181-196

[11] Brockfeld, E., R.D. Kühne and P. Wagner: 'Calibration and validation of microscopic traffic flow models', Transportation Research Record, 2004, 1876, pp. 62-70

[12] Ossen, S.J.L., H. S.P. and B.G.H. Gorte: 'Interdriver differences in car-following: a vehicle trajectory-based study', Transportation Research Record, 2006, 1965, pp. 121-129

[13] Hoogendoorn, S.P., S.J.L. Ossen and J.W.C. van Lint: 'Advanced calibration of car-following models'. 11th World Conference on Transport Research, Berkeley, California, 2007

[14] Hoogendoorn, S.P. and S.J.L. Ossen: 'Parameter estimation and analysis of car-following models'. 16th international symposium on transportation and traffic theory, College Park, Maryland, USA, 2005

[15] Hamdar, S.H., M. Treiber, H.S. Mahmassani and A. Kesting: 'Modeling driver behavior as a sequential risk taking task', Transportation Research Record, 2008, 2088, pp. 208-217

[16] Hoogendoorn, S.P., S.J.L. Ossen and M. Schreuder: 'Properties of a microscopic heterogeneous multi-anticipative traffic flow model', R.E. Allsop, M.G.H. Bell, and B.G. Heydecker, Editors (Ed.): 'Transportation and Traffic Theory 2007' (Elsevier, 2007), pp. 583-606

[17] Panwai, S. and H. Dia: 'Comparative Evaluation of Microscopic Car-Following Behavior', IEEE Transactions on Intelligent Transportation Systems, 2005, 6, (3), pp. 314-325

[18] Chakroborty, P. and S. Kikuchi: 'Evaluation of the General Motors based car-following models and a proposed fuzzy inference model', Transportation Research Part C: Emerging Technologies, 1999, 7, pp. 209-235

[19] Kesting, A. and M. Treiber: 'Calibrating Car-Following Models by Using Trajectory Data', Transportation Research Record: Journal of the Transportation Research Board, 2008, 2088, pp. 148-156

[20] Punzo, V. and A. Tripodi: 'Steady-state solutions and multi-class calibration of Gipps' microscopic traffic flow model', Transportation Research Record: Journal of the Transportation Research Board, 2007, 1999, pp. 104-114

[21] Ranjitkar, P., T. Nakatsuji and A. Kawamura: 'Experimental Analysis of Car-Following Dynamics and Traffic Stability', Transportation Research Record: Journal of the Transportation Research Board, 2005, 1934, pp. 22-32

[22] Ossen, S.J.L. and S.P. Hoogendoorn: 'Car-Following Behavior Analysis from Microscopic Trajectory Data', Transportation Research Record: Journal of the Transportation Research Board, 2005, 1934, pp. 13-21

[23] Ranjitkar, P., T. Nakatsuji and M. Asano: 'Performance Evaluation of Microscopic Traffic Flow Models with Test Track Data', Transportation Research Record: Journal of the Transportation Research Board, 2004, 1876, pp. 90-100

[24] Aycin, M.F. and R.F. Benekohal: 'Comparison of Car-Following Models for Simulation', Transportation Research Record: Journal of the Transportation Research Board, 1999, 1678, pp. 116-127

[25] Rakha, H. and B. Crowther: 'Comparison and calibration of FRESIM and INTEGRATION steady-state car-following behavior', Transportation Research Part A, 2003, 37, pp. 1-27

[26] Abbas, M.M., B. Higgs, Z. Adam and A. Medina: 'Comparison of car-following models when calibrated to individual drivers using naturalistic data'. 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011

[27] Hawkins, D.M.: 'The problem of overfitting', Journal of Chemical Information and Modeling, 2004, 44, pp. 1-12

[28] Punzo, V. and F. Simonelli: 'Analysis and Comparison of Microscopic Traffic Flow Models with Real Traffic Microscopic Data', Transportation Research Record: Journal of the Transportation Research Board, 2005, 1934, pp. 53-63

[29] Wu, J., M. Brackstone and M. McDonald: 'The validation of a microscopic simulation model: a methodological case study', Transportation Research Part C: Emerging Technologies, 2003, 11, pp. 463-479

[30] Brockfeld, E., R.D. Kuhne, A. Skabardonis and P. Wagner: 'Towards Benchmarking of Microscopic Traffic Flow Models', Transportation Research Record: Journal of the Transportation Research Board, 2003, 1852, pp. 124-129

[31] Hoogendoorn, S.P., S.J.L. Ossen and M. Schreuder: 'Empirics of Multianticipative Car-Following Behavior', Transportation Research Record: Journal of the Transportation Research Board, 2006, 1965, pp. 112-120

[32] van Hinsbergen, C.P.I., J.W.C. van Lint and H.J. van Zuylen: 'Bayesian committee of neural networks to predict travel times with confidence intervals', Transportation Research Part C: Emerging Technologies, 2009, 17, pp. 498-509

[33] Bishop, C.M.: 'Neural Networks for Pattern Recognition' (Oxford University Press, 1995)

[34] Petridis, V., A. Kehagias, L. Petrou, S. Bakirtzis, S. Kiartzis, H. Panagiotou and N. Maslaris: 'A Bayesian multiple models combination method for time series prediction', Journal of intelligent and robotic systems, 2001, 31, pp. 69-89

[35] Kuchipudi, C.M. and S.I.J. Chien: 'Development of a Hybrid Model for Dynamic Travel-Time Prediction', Transportation Research Record, 2003, 1855, pp. 22-31

[36] Zheng, W., D. Lee and Q. Shi: 'Short-term freeway traffic flow prediction: Bayesian combined neural network approach', Journal of Transportation Engineering, 2006, 132, (2), pp. 114-121

[37] Schwarz, G.: 'Estimating the dimension of a model', The Annals of Statistics, 1978, 6, (2), pp. 461-464

[38] MacKay, D.J.C.: 'Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks', Network: Computation in Neural Systems, 1995, 6, pp. 469-505

[39] van Hinsbergen, C.P.I. and J.W.C. van Lint: 'Bayesian combination of travel time prediction models', Transportation Research Record, 2008, 2064, pp. 73-80

[40] van Hinsbergen, C.P.I., J.W.C. van Lint and H.J. van Zuylen: 'Bayesian training and committees of State-Space Neural Networks for online travel time prediction', Transportation Research Record, 2009, 2105, pp. 118-126

[41] Beal, M.J. and Z. Ghahramani: 'The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures', J.M. Bernardo, et al., Editors (Ed.): 'Bayesian statistics 7: Proceedings of the Seventh Valencia International Meeting' (Oxford University Press, 2003), pp. 453-463

[42] Lisboa, P.J.G., H. Wong, P. Harris and R. Swindell: 'A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer', Artificial Intelligence in Medicine, 2003, 28, pp. 1-25

[43] He, H., J. Wang, W. Graco and S. Hawkins: 'Application of neural networks to detection of medical fraud', Expert Systems With Applications, 1997, 13, pp. 329-336

[44] Baesens, B., S. Viaene, D. van den Poel, J. Vanthienen and G. Dedene: 'Bayesian neural network learning for repeat purchase modelling in direct marketing', European Journal of Operational Research, 2002, 138, pp. 191-211

[45] Blumer, A., A. Ehrenfeucht, D. Haussler and M.K. Warmuth: 'Occam's Razor', Information Processing Letters, 1987, 24, pp. 377-380

[46] Gull, S.F.: 'Developments in maximum entropy data analysis' (Kluwer, 1989)

[47] van Hinsbergen, C.P.I., J.W.C. van Lint and H.J. van Zuylen: 'Neural Network Committee to Predict Travel Times: Comparison of Bayesian Evidence Approach to the Use of a Validation Set'. 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, 2008

[48] MacKay, D.J.C.: 'Information theory, inference and learning algorithm' (Cambridge University Press, 2003)

[49] Sivia, D.S.: 'Data analysis: a Bayesian tutorial' (Oxford University Press, 1996)

[50] Thodberg, H.H.: 'Ace of Bayes: Application of Neural Networks with Pruning' (The Danish Meat Research Institute, 1993)

[51] MacKay, D.J.C.: 'Bayesian interpolation', Neural Computation, 1992, 4, pp. 415-447

[52] Fletcher, R.: 'Practical Methods of Optimization. Volume 1: Unconstrained Optimization' (John Wiley & Sons Ltd, 1980)

[53] Fletcher, R.: 'Practical Methods of Optimization. Volume 2: Constrained Optimization' (John Wiley & Sons Ltd, 1981)

[54] Broyden, C.G.: 'The Convergence of a Class of Double-Rank Minimization Algorithms', Journal of the Institute of Mathematics and Its Applications, 1970, 6, pp. 76-90

[55] D'Errico, J.R.: 'Description of DERIVEST, http://www.convexoptimization.com/TOOLS/DERIVEST.pdf' (2007)

[56] Kesting, A. and M. Treiber: 'Calibration of Car-Following Models Using Floating Car Data', C. Appert-Rolland, et al., Editors (Ed.): 'Traffic and Granular Flow ’07 part 1' (2009), pp. 117-127

[57] Hoogendoorn, S., S. Ossen and M. Schreuder: 'Multi-Anticipative Car-Following Behavior: An Empirical Analysis', A. Schadschneider, et al., Editors (Ed.): 'Traffic and Granular Flow ’05 part 4' (2007), pp. 687-697

[58] Kim, J.-W. and H.S. Mahmassani: 'Correlated Parameters In Driving Behavior Models: Car-Following Example and Implications For Traffic Microsimulation'. 90th Annaul Meeting of the Transportation Research Board, Washington, DC, 2011

[59] Hoogendoorn, S.P. and R. Hoogendoorn: 'Calibration of microscopic traffic-flow models using multiple data sources', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2010, 368, (1928), pp. 4497-4517

[60] Treiber, M., A. Kesting and D. Helbing: 'Delays, inaccuracies and anticipation in microscopic traffic models', Physica A: Statistical Mechanics and its Applications, 2006, 360, (1), pp. 71-88

[61] Hoogendoorn, S.P., H.J. van Zuylen, M. Schreuder, B.G.H. Gorte and G. Vosselman: 'Microscopic Traffic Data Collection by Remote Sensing', Transportation Research Record, 2003, 1855, pp. 121-128

[62] Ossen, S.: 'Longitudinal Driving Behavior: Theory and Empirics'. PhD thesis, TRAIL Research School, Delft, the Netherlands, 2008