

Bayesian Calibration of Car-Following Models

Chris P.I.J. van Hinsbergen^{**}, Hans W.C. van Lint^{*}
Serge P. Hoogendoorn^{*}, Henk J. van Zuylen^{**}

^{*} *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 (Tel: +31 15 2784066; fax: +31 15 2783179; email: c.p.i.j.vanhinsbergen@tudelft.nl)*
^{**} *Member of the above department as well as TRAIL Research School*

Abstract: Recent research has revealed that there exist large inter-driver differences in car-following behavior such that different car-following models may apply to different drivers. This study applies Bayesian techniques to the calibration of car-following models, where prior distributions on each model parameter are converted to posterior distributions. The priors and posteriors are then used to calculate the so-called ‘evidence’, which can be used to quantitatively assess how well different models explain one driver’s car-following behavior. When considered over multiple drivers, the evidence represents probabilities for different models as a whole. These model probabilities can be used in a micro simulation, where for each driver first a model is drawn according to these probabilities, after which parameters are drawn from the posterior distribution for each parameter of that model that were obtained when calibrating the model. In a test case on actual data the Bayesian evidence indeed reveals inter-driver differences and it is shown how these differences can quantitatively be assessed.

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

1. INTRODUCTION

1.1 Inter- and intra-driver differences

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 (Brockfeld et al., 2004, Ossen et al., 2006, Hoogendoorn et al., 2007b). 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 (Hoogendoorn and Ossen, 2005, Hamdar et al., 2008). When microscopically modeling traffic, these inter- and intra-driver differences therefore need to be considered.

Existing methods to deal with the inter-driver differences, by comparing and selecting one car-following model for each driver from a set of possible models, involve investigating the calibration error (Ossen et al., 2006) or the Likelihood-Ratio Test (LRT) (Hoogendoorn et al., 2007a, Hoogendoorn et al., 2007b). The first approach does not take the model complexity into account, and therefore promotes over fitted models, always favoring more complex models over simpler models. The LRT approach does take model complexity into account and therefore prevents selecting over fitted models, but is only valid when used to compare hierarchically nested models, e.g. the simple model must be a special case of the more complex model by setting one or multiple parameters to

zero. As there are many different types of car-following models (Brackstone and McDonald, 1999, Tampère, 2004), this will not be the case when a modeler is interested in trying many different car-following models.

To remain focused, in this study the emphasis is on the inter-driver differences and on choosing the appropriate model for the appropriate driver; the intra-driver differences will not be considered. However, the approach that is presented in this paper does enable a modeler to deal with these intra-driver differences.

1.2 The Bayesian approach to calibrating car-following models

Recently, a new approach was proposed to calibrate and compare several car-following models (Hoogendoorn et al., 2007b). The main focus of the paper was to include prior information when calibrating the parameters of car-following models, to rule out unrealistic estimation results due to the fact that too little information is present within data. This prior information was included in a general Likelihood-estimator. It was shown that this in fact can lead to improved prediction accuracy of the car-following models, as well as more realistic estimates of the parameter values to ensure for example model stability and a correctly shaped fundamental diagram resulting from the model.

The approach of Hoogendoorn et al. used prior distributions and data to derive single (most likely) parameter values for each parameter. In this paper the approach of Hoogendoorn et

al. is extended, where the prior probabilities are transformed into *posterior probabilities* for each parameter in the car-following model, for which Bayes' rule is used. This approach allows for several things: (1) the most important feature is that it leads to a probabilistic approach to compare different models on the basis of the prior and posterior distributions of their parameters. 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; **any model** that is differentiable to its parameters can be used; (2) prior information can easily be included in the analysis, yielding realistic parameter estimates even if the data does not provide information on a certain parameter; (3) it can be used to combine the predictions of several models in a so-called *committee* (or *ensemble*) of models in which different models predict the behavior of one single driver, which may lead to a decrease in the error due to the intra-driver differences; (4) error bars can be constructed on the predictions of the car-following models.

In this study, analytical expressions are derived, for which several assumptions will need to be made; in real-world applications, these assumptions could be relaxed and numerical approximations could be used. However, to illustrate the point of this paper, namely that the Bayesian approach can be used to select the most probable model for each driver in a dataset, and to show how this can be done, analytical solutions will be used.

In the following section, the Bayesian approach to calibrate and choose car-following models is developed, after which it is applied to two relatively simple car-following models: the CHM model and the linear Helly model. Next, the result of the Bayesian 'evidence' as a selection mechanism is shown, after which a discussion, a conclusion and recommendations are presented.

2. METHODOLOGY

2.1 Bayesian Inference: From Prior to Posterior

For the Bayesian analysis, the interest is in finding the posterior probability density function of a parameter vector $\theta = (\theta_1, \dots, \theta_N)^T$ which contains all N parameters of a car-following model under investigation after having used some data set D for calibration. This data set contains for example positions (lateral and longitudinal) and speeds of different vehicles, from which car-following models can be calibrated. This posterior probability is denoted by $p(\theta|D)$, e.g. the probability density function of the parameters θ given the data set D . Bayes' rule can be applied to find an expression for this posterior:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (1)$$

where $p(D|\theta)$ represents the distribution of noise on the data and corresponds to the likelihood function, $p(\theta)$ is a prior probability of the parameters, which represents our prior

knowledge of possible values for each parameter in our model, and where $p(D)$ is a normalization factor.

Now define the prior probability as a multivariate Gaussian with mean $\bar{\theta}$ and covariance matrix Σ :

$$p(\theta) = \frac{\exp\left(-\frac{1}{2}(\theta - \bar{\theta})^T \Sigma^{-1}(\theta - \bar{\theta})\right)}{(2\pi)^{N/2} |\Sigma|^{1/2}} \quad (2)$$

where N equals the number of parameters of the model. A Gaussian shape is chosen in this study because it simplifies the calculations and enables analytical expressions for the posterior distribution of the parameters. Note that this assumption can be relaxed and other shapes are possible. If it is assumed that the noise of the data is Gaussian distributed as well with mean zero and standard deviation σ_i , the likelihood function $p(D|\theta)$ can be defined as (Hoogendoorn et al., 2007b):

$$p(D|\theta) = \frac{\exp\left(-\frac{1}{2\sigma_i^2} \sum_{k=1}^K (v_{pred}(k, \theta) - v_{obs}(k))^2\right)}{(\sigma_i^2 2\pi)^{K/2}} \quad (3)$$

where $v_{pred}(k, \theta)$ is the predicted vehicle speed at time instant k with the parameter set θ , $v_{obs}(k)$ is the observed (measured) vehicle speed at time instant k , and where K equals the number of observations of vehicle speed and position. Note that in this study the models are calibrated on speeds alone, but that other likelihood functions which incorporate for example the predicted positions of the vehicles are also possible.

2.2 Description of the Posterior Distribution of the Parameters

Substituting (2) and (3) into (1) results in an expression for the posterior distribution of the parameters:

$$p(\theta|D) = \frac{1}{Z_p \sigma_i^K |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2\sigma_i^2} \sum_{k=1}^K (v_{pred}(k, \theta) - v_{obs}(k))^2\right) \cdot \exp\left(-\frac{1}{2}(\theta - \bar{\theta})^T \Sigma^{-1}(\theta - \bar{\theta})\right) \quad (4)$$

where Z_p is a constant that originates from $p(D)$ and the '2 π -constants' in (2) and (3). This posterior distribution of the parameters can be described by the *most probable parameter vector* θ^{MP} (the maximum of the posterior), and its *covariance matrix* Θ (the width of the posterior), with the knowledge that it has a Gaussian shape.

The maximum of the posterior is denoted by the vector θ^{MP} , which can be found by maximizing the logarithm of (4):

$$\begin{aligned}
(\boldsymbol{\theta}^{MP}) &= \arg \max_{\boldsymbol{\theta}} \ln(p(\boldsymbol{\theta} | D)) \\
&= \arg \max_{\boldsymbol{\theta}} -E(\boldsymbol{\theta}) \\
&= \arg \min_{\boldsymbol{\theta}} E(\boldsymbol{\theta})
\end{aligned} \tag{5}$$

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

$$E(\boldsymbol{\theta}) = K \ln(\sigma_l) + E_p(\boldsymbol{\theta}) + E_l(\boldsymbol{\theta}) \tag{6}$$

with E_l and E_p defined by:

$$E_p(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \Sigma^{-1}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \tag{7}$$

$$E_l(\boldsymbol{\theta}) = \frac{1}{2\sigma_l^2} \sum_{k=1}^K (v_{pred}(k, \boldsymbol{\theta}) - v_{obs}(k))^2. \tag{8}$$

Notice that in (6) the expressions resulting from Z_p and $|\Sigma|^{1/2}$ have been omitted, as these do not influence the solution of (5) and becomes zero for the derivatives that are defined next. For the maximization of (6) (so to find $\boldsymbol{\theta}^{MP}$), there is the condition (Hoogendoorn et al., 2007b):

$$\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) = \Sigma^{-1}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) + \nabla_{\boldsymbol{\theta}} E_l(\boldsymbol{\theta}) = 0 \tag{9}$$

which needs to be solved for the model under consideration.

The covariance matrix Θ of the posterior distribution (not to be confused with the covariance matrix Σ of the prior) can be found using the Cramér-Rao lower bound:

$$\Theta(\boldsymbol{\theta}^{MP}) = -(\mathbf{A}(\boldsymbol{\theta}^{MP}))^{-1} \tag{10}$$

where $\mathbf{A}(\boldsymbol{\theta})$ is the *Hessian*, given by:

$$\mathbf{A}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}^2 E(\boldsymbol{\theta}) = \Sigma^{-1} + \nabla_{\boldsymbol{\theta}}^2 E_l(\boldsymbol{\theta}). \tag{11}$$

Finally, for the description of the posterior a value for the standard deviation of the likelihood function σ_l needs to be found, for which the derivative $\partial E / \partial \sigma_l$ is set to zero, which leads to

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

2.3 The Bayesian Framework for Model Comparison

Consider a certain car-following model m with a set of assumptions H_m , and another model n with a different set of assumptions H_n . To compare these two models in how well they describe the car-following behavior of a certain driver, the posterior probability of a model $q \in (m, n)$ as a whole after it has been calibrated with data D for this driver, which we call $P(H_q | D)$, can be derived by again applying Bayes' rule:

$$P(H_q | D) = \frac{p(D | H_q) P(H_q)}{p(D)} \tag{13}$$

The term $P(H_q)$ represents the prior probability of each model. If a priori there is no preference of one type of model over the other (so there is belief that the assumptions H_m are as likely as H_n), then the prior $P(H_q)$ is equal for all q . As the denominator of (13) is independent of the models H_q , the posterior probabilities of the models m and n can be compared by only investigating the term $p(D | H_q)$, which is termed the *evidence* for the model q (MacKay, 1995):

$$P(H_q | D) \sim p(D | H_q) \tag{14}$$

This evidence can be recognized as the denominator of (1) if the conditional dependence on the model assumptions H_q is made explicit. The expressions used for deriving the *posterior distribution for the parameters* can therefore be used to derive expressions for the *evidence for the entire model*. First, from (1) the evidence can be written in the form

$$p(D | H_q) = \int p(D | \boldsymbol{\theta}, H_q) p(\boldsymbol{\theta} | H_q) d\boldsymbol{\theta}. \tag{15}$$

Since this term 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. Although a numerical approximation could be used, in this study an analytical approximation is chosen to be able to analytically describe the evidence. Assuming that the posterior distribution is sharply peaked around its maximum, the evidence is approximated as the value at this maximum times the width of the peak, which in the multivariate case leads to the expression (MacKay, 1995):

$$p(D | H_q) \approx p(D | \boldsymbol{\theta}^{MP}, H_q) \times p(\boldsymbol{\theta}^{MP} | H_q) \det^{-1/2}(\mathbf{A}(\boldsymbol{\theta}^{MP}) / 2\pi) \tag{16}$$

Together with (2), (3) and (11) a solution (approximation) is now found for the evidence. Note that values for the prior covariance matrix Σ and the prior mean $\bar{\boldsymbol{\theta}}$ are needed for this; the way the prior is defined will be treated later.

The evidence of (16) can be interpreted as consisting of two elements:

$$\text{Evidence} = \text{Best fit likelihood} \times \text{Occam factor} \tag{17}$$

A higher best fit likelihood favors models that can explain the data well, i.e. that have a low prediction error $\Sigma(v_{pred} - v_{obs})^2$. However, the model's performance is penalized by the Occam factor, which is always smaller than 1 and is named after Occam's Razor (Blumer et al., 1987). A model which has more parameters, so which is more complicated, has a lower Occam factor and therefore receives lower evidence. The evidence thus naturally reflects the trade-off between a good fit and over fitting. Extensive literature is available on the importance of this trade-off and other features of the evidence (Thodberg, 1993, Bishop, 1995, MacKay, 1995, Sivia, 1996, MacKay, 2003, van Hinsbergen and van Lint, 2008, van Hinsbergen et al., 2008).

In the remainder of this contribution, the evidence is used to rank different car-following models for individual drivers.

This is achieved by determining the evidence after the posterior distribution of its parameters has been found, after which a conclusion can be drawn to which model probably describes which driver's behavior best. The Bayesian analysis will be applied here to two simple car-following models, for which the evidence can be derived analytically.

2.4 The Evidence for the CHM Model

To illustrate the derivation of the evidence for a car-following model, consider the CHM model (Chandler et al., 1958). This stimulus-response model describes the delayed acceleration of a vehicle as a function of the relative speed with respect to its leading vehicle:

$$a(t + \tau, \theta) = \gamma \Delta v(t) \quad (18)$$

where $a(t + \tau, \theta)$ is the acceleration of the following vehicle at time $t + \tau$ given the parameter set θ and $\Delta v(t)$ the speed difference between the leader and the follower at time t . In this study, one-step-ahead predictions are made, where the observed speeds of the follower and its leader in the previous time step are used in the calculations. An explicit time stepping scheme is used to solve the model, resulting in the following numerical scheme for the speed at time t :

$$v_{pred}(t, \theta) = v_{obs}(t - \Delta t) + a(t - \Delta t, \theta) \Delta t \quad (19)$$

with $v_{obs}(t - \Delta t)$ the observed speed at time $t - \Delta t$, and Δt the size of the time step which should be sufficiently small. The acceleration is in this scheme determined by:

$$a(t - \Delta t, \theta) = \gamma \Delta v_{obs}(t - \Delta t - \tau) \quad (20)$$

The model has only one parameter that needs to be calibrated with data:

$$\gamma \quad \text{response parameter (1/s)}$$

For this model, the parameter vector is denoted as $\theta = \gamma$. Note that for the sake of this example, the reaction time τ is chosen to be a constant with a value of $\tau = 1$ s, and not as a parameter. This is done because in this case the partial derivative $\partial E(\theta) / \partial \tau$ cannot be derived analytically. In a real world application, the reaction time τ does need to be calibrated with data, and numerical derivatives would be needed instead of analytical ones.

To analytically derive the evidence for the CHM model, first the gradient of (9) needs to be computed:

$$\frac{\partial E(\theta)}{\partial \gamma} = \frac{1}{\sigma_{prior}^2} (\gamma - \bar{\gamma}_{prior}) + \frac{\Delta t}{\sigma_l^2} \sum_{k=1}^K v_q (v_p + \gamma v_q \Delta t - v_s) \quad (21)$$

where $\bar{\gamma}_{prior}$ is the mean of the prior distribution and σ_{prior}^2 is the prior variance (previously $\bar{\theta}$ and Σ , but now for the one-dimensional case), and where v_q , v_p and v_s are defined by:

$$\begin{aligned} v_s &= v_{obs}(k) \\ v_p &= v_{obs}(k - \Delta t) \\ v_q &= \Delta v_{obs}(k - \Delta t - \tau) \end{aligned} \quad (22)$$

The Hessian of (11) is given by:

$$\frac{\partial^2 E_l(\theta)}{\partial \gamma^2} = \frac{1}{\sigma_{prior}^2} + \frac{\Delta t^2}{\sigma_l^2} \sum_{k=1}^K v_q^2 \quad (23)$$

To calculate the evidence, the most probable parameter γ^{MP} is required, for which (9) needs to be solved. This is done numerically using standard Matlab optimization tools as the analytical solution becomes rather complex. Then σ_l^{MP} is calculated using (12), γ^{MP} and σ_l^{MP} are substituted in (2), (3) and (11), and the resulting equations into (16) together with $\bar{\gamma}_{prior}$ and σ_{prior}^2 , resulting in the evidence for the model.

2.5 The Evidence for the Helly Model

As a second example of the derivation of the evidence for a car-following model, consider the Helly model (Helly, 1959), another stimulus-response model, which is defined by:

$$\begin{aligned} a(t + \tau, \theta) &= \alpha \Delta v(t) + \beta \{ \Delta x(t) - \Delta x^{des}(v(t)) \} \\ \Delta x^{des}(v) &= x_0 + Tv \end{aligned} \quad (24)$$

where $a(t + \tau, \theta)$ is the acceleration of the following vehicle at time $t + \tau$ given the parameter set θ , $\Delta v(t)$ the speed difference between the leader and the follower at time t , $\Delta x(t)$ the distance headway between the leader and the follower at time t and $\Delta x^{des}(v(t))$ the desired distance headway of driver when driving at speed $v(t)$, the speed of the follower at time t . Again, one-step-ahead predictions are made, where the observed speeds and distances of the follower and its leader in the previous time step are used in the calculations. The same numerical scheme as in (19) is used, but with the acceleration now determined by:

$$\begin{aligned} a(t - \Delta t, \theta) &= \alpha \Delta v_{obs}(t - \Delta t - \tau) \\ &+ \beta \{ \Delta x_{obs}(t - \Delta t - \tau) - \Delta x^{des}(v_{obs}(t - \Delta t - \tau)) \} \end{aligned} \quad (25)$$

The model has the following four parameters that need to be estimated from data:

$$\begin{aligned} \alpha &\quad \text{response parameter (1/s)} \\ \beta &\quad \text{response parameter (1/s}^2\text{)} \\ x_0 &\quad \text{stopping distance (m)} \\ T &\quad \text{minimum time headway (s)} \end{aligned}$$

For this model, the parameter vector is denoted as $\theta = (\alpha, \beta, x_0, T)$. Again, as with the CHM model, the reaction time is chosen to be a constant with a value of $\tau = 1$ s, and not as a parameter.

The gradient and Hessian for the Helly model are derived analytically again, the result of which will be omitted here as it involves quite lengthy equations. The most probable parameter vector θ^{MP} is estimated numerically using standard numerical tools in the Matlab software package, as the

condition (9) is not easily solvable analytically. Then, the same procedure as with the CHM model is used to calculate the evidence.

2.6 Prior Distributions for the Parameters

For both models, prior distributions need to be defined for all parameters in the model. The prior information can have various origins: it can be based on expert knowledge, a literature study or by restrictions such that the resulting model has specific desired properties such as model stability or a correct shape of the fundamental diagram. In this study, a brief literature review was conducted to find prior distributions for the parameters.

2.6.1 Prior Distribution for the CHM Model Parameter

The original work of Chandler, Herman and Montroll showed high variations between subjects for the constant γ , between $0.17s^{-1}$ and $0.74s^{-1}$ with a mean of $0.37s^{-1}$ (Chandler et al., 1958, Brackstone and McDonald, 1999). A benchmarking study (Ossen et al., 2006), conducted on a Dutch motorway using helicopter data, as is the case in this study, showed the distribution of parameter values for the CHM model as shown in Fig. 1, more or less confirming the spread of the original study of Chandler, Herman and Montroll. From the results of these studies, a prior distribution $N(\bar{\gamma}_{prior}, \sigma_{prior}^2) = N(0.3, 0.04)$ is chosen.

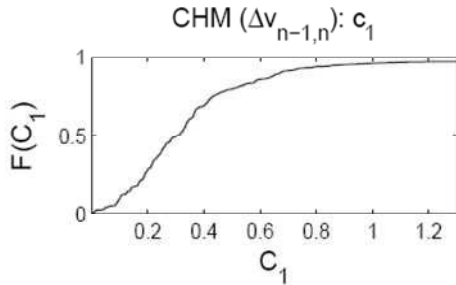


Fig. 1. The cumulative distribution of the parameter γ (denoted by c_1 in the figure) given by (Ossen et al., 2006)

2.6.1 Prior Distribution for the CHM Model Parameter

Helly in his original work (Helly, 1959) estimated the mean parameter values $\alpha=0.5s^{-1}$, $\beta=0.125s^{-2}$, $x_0=20m$ and $T=1s$. The earlier mentioned benchmarking study (Ossen et al., 2006) only presents CDFs for α and β , as shown in Fig. 2. Taking both these studies into account, the following prior mean and covariance matrix are chosen (not taking into account covariance between the different parameter).

$$\bar{\theta} = (\bar{\alpha}, \bar{\beta}, \bar{x}_0, \bar{T})^T = (0.25, 0.075, 20.0, 1.0)^T$$

$$\Sigma = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 40.0 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix}$$

Here, large variances are taken for x_0 and T , as there is no reference study available for estimates of the variance of these two parameters. These broad priors represent the fact that little knowledge is available on the most probable values of these parameters. However, the variances are chosen in such a way that it is ensured that most of the mass (at least 95%) of the CDF is for values >0 , which is sensible in the light of the physical meaning of these two parameters.

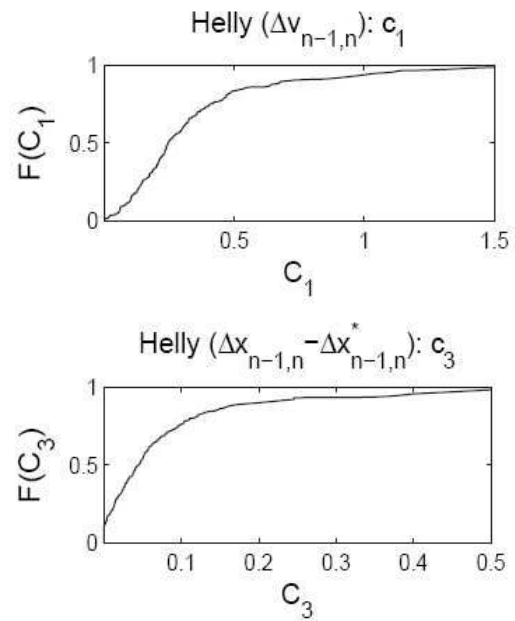


Fig. 2. The cumulative distribution functions of α (c_1) and β (c_3) given by (Ossen et al., 2006)

3. EXPERIMENT

To illustrate the workings of the Bayesian evidence, the two models described in the methodology section are applied to a vehicle trajectory data set of the A2 motorway in The Netherlands, near the city of Utrecht, which was collected using helicopter data (Hoogendoorn et al., 2003). The traffic state at the data collection period was congested, in which cars were mainly in car-following mode. The data covers approximately 500m of motorway stretch; the data interval is 0.1s.

For each trajectory (driver) in the dataset that was following one leader without any lane-changes of either follower or leader (222 drivers in total), the posterior distributions of the parameters of the two models were found after which the evidence was calculated for each model for each driver. Note that the *log evidence* is used, as the denominator of (3) is taken to the power of K , which means that the likelihood becomes very large if $\sigma_i < 1$ and very small if $\sigma_i > 1$ in case $K \gg 1$.

Given that the number of measurements and predictions is in the order of 100 to 400 for each driver, the log of the evidence is used to prevent numerical errors in the computations.

4. RESULTS

Fig. 3 shows the evidence for the two models for some of the 222 drivers. It can be seen that in general, the differences between the two models are small, and that the Helly model is favored slightly in most cases. The small difference is to be expected, because the two models share equal basic assumptions.

It can also be concluded that the added complexity (the additional set of parameters) of the Helly model in some cases does help in explaining the car-following behavior in the collected data set, as its log evidence is in those cases larger than the CHM's log evidence, and that in other cases its additional parameters are useless for explaining a driver's behavior, in which cases it receives a lower evidence. The number of occasions where each of the models had the largest log evidence is shown in Table 1.

Table 1. Results

Model	# largest log evidence	$P(H D)$	Mean log evidence
CHM	77	31.5%	206.7
Helly	152	68.5%	216.4

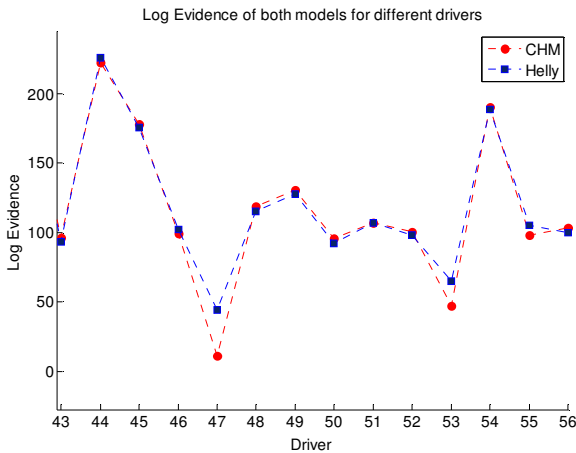


Fig. 3. The log evidence for the two models for 14 drivers

5. DISCUSSION AND CONCLUSION

The Bayesian evidence that has been developed for the car-following models in this paper is shown to be useful as a tool for quantitatively analyzing inter-driver differences. As can be seen from the experiment, these inter-driver differences indeed do exist: for some of the drivers the CHM model suffices and the additional parameters of the Helly model do not contribute to explaining their car-following behavior, in which cases the Helly model is punished for its higher complexity, while for others the additional parameters do lead to a better explanation of the car-following behavior in

which cases the Helly model is rewarded for this. The Bayesian evidence thus acts as a natural selection mechanism when choosing between different car-following models. Note that for the two models chosen in this study, the Likelihood Ratio Test (LRT) could also be applied, but that the evidence is favorable over the LRT in the general case, **because the evidence can be used for any model** that is analytically or numerically differentiable to its parameters, while the LRT can only be applied to hierarchically nested models.

The evidence, when normalized, represents a distribution of different models' probabilities to describe one drivers' behavior i , i.e. $P(H_i|D) \sim p(D|H_i)$. If the evidence is integrated over all drivers i of a certain dataset, an approximation of $P(H|D)$ for an entire population of drivers can be made, as is done in Table 1, indicated by $P(H|D)$. This enables a modeler to use different car-following models with different probabilities in one single micro-simulation. In such a simulation, for each driver first a model is randomly drawn according to $P(H|D)$, after which the parameters are drawn from $p(\theta|H,D)$; both $P(H|D)$ and $p(\theta|H,D)$ follow from the Bayesian calibration procedure.

Other benefits of the Bayesian approach that have not been illustrated in this study are the possibility of using the evidence to create a committee, and to construct error bars. A committee may improve the description of individual behavior (because it may deal with the intra-driver differences), while the error bars 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

- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford, Oxford University Press.
- Blumer, A., Ehrenfeucht, A., Haussler, D. & Warmuth, M. K. (1987). Occam's Razor. *Information Processing Letters*, 24, 377-380.
- Brackstone, M. & McDonald, M. (1999). Car-following: a historical review. *Transportation Research Part F*, 2, 181-196.
- Brockfeld, E., Kühne, R. D. & Wagner, P. (2004). Calibration and validation of microscopic traffic flow models. *Transportation Research Record*, 1876, 62-70.
- Chandler, R. E., Herman, R. & Montroll, E. W. (1958). Traffic Dynamics: Studies in Car Following. *Operations Research*, 6, 165-184.
- Hamdar, S. H., Treiber, M., Mahmassani, H. S. & Kesting, A. (2008). Modeling driver behavior as a sequential

- risk taking task. *Transportation Research Record*, 2088, 208-217.
- Helly, W. (1959). Simulation of bottlenecks in single lane traffic flow. *International symposium on the theory of traffic flow*. New York, NY, USA.
- Hoogendoorn, S. P. & Ossen, S. J. L. (2005). Parameter estimation and analysis of car-following models. *16th international symposium on transportation and traffic theory*. College Park, Maryland, USA.
- Hoogendoorn, S. P., Ossen, S. J. L. & Schreuder, M. (2007a). Properties of a microscopic heterogeneous multi-anticipative traffic flow model. IN Allsop, R. E., Bell, M. G. H. & Heydecker, B. G. (Eds.) *Transportation and Traffic Theory 2007*. Amsterdam, The Netherlands, Elsevier.
- Hoogendoorn, S. P., Ossen, S. J. L. & Van Lint, J. W. C. (2007b). Advanced calibration of car-following models. *11th World Conference on Transport Research*. Berkeley, California.
- Hoogendoorn, S. P., Van Zuylen, H. J., Schreuder, M., Gorte, B. G. H. & Vosselman, G. (2003). Microscopic Traffic Data Collection by Remote Sensing. *Transportation Research Record*, 1855, 121-128.
- Mackay, D. J. C. (1995). Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6, 469-505.
- Mackay, D. J. C. (2003). *Information theory, inference and learning algorithm*. Cambridge, UK, Cambridge University Press.
- Ossen, S. J. L., S.P., H. & Gorte, B. G. H. (2006). Interdriver differences in car-following: a vehicle trajectory-based study. *Transportation Research Record*, 1965, 121-129.
- Sivia, D. S. (1996). *Data analysis: a Bayesian tutorial*. New York, USA, Oxford University Press.
- Tampère, C. M. J. (2004). Human-Kinetic multiclass traffic flow theory and modelling. Delft, The Netherlands, Delft University of Technology.
- Thodberg, H. H. (1993). *Ace of Bayes: Application of Neural Networks with Pruning*. Roskilde, The Danish Meat Research Institute.
- Van Hinsbergen, C. P. I. & Van Lint, J. W. C. (2008). Bayesian combination of travel time prediction models. *Transportation Research Record*, 2064, 73-80.
- Van Hinsbergen, C. P. I., Van Lint, J. W. C. & Van Zuylen, H. J. (2008). Neural Network Committee to Predict Travel Times: Comparison of Bayesian Evidence Approach to the Use of a Validation Set. *IEEE-ITSC*. Beijing, China.