

Performance Boundary Identification for the Evaluation of Automated Vehicles using Gaussian Process Classification

Batsch, F, Daneshkhah, A, Cheah, M, Kanarachos, S & Baxendale, A

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Batsch, F, Daneshkhah, A, Cheah, M, Kanarachos, S & Baxendale, A 2019, Performance Boundary Identification for the Evaluation of Automated Vehicles using Gaussian Process Classification. in 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, pp. 419-424, 22nd IEEE International Conference on Intelligent Transportation Systems, Auckland, New Zealand, 27/10/19.

<https://dx.doi.org/10.1109/ITSC.2019.8917119>

DOI 10.1109/ITSC.2019.8917119

Publisher: IEEE

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Performance Boundary Identification for the Evaluation of Automated Vehicles using Gaussian Process Classification

Felix Batsch^{1,2}, Alireza Daneshkhah¹, Madeline Cheah², Stratis Kanarachos¹, Anthony Baxendale²

Abstract—Safety is an essential aspect in the facilitation of automated vehicle deployment. Current testing practices are not enough, and going beyond them leads to infeasible testing requirements, such as needing to drive billions of kilometres on public roads. Automated vehicles are exposed to an indefinite number of scenarios. Handling of the most challenging scenarios should be tested, which leads to the question of how such corner cases can be determined. We propose an approach to identify the performance boundary, where these corner cases are located, using Gaussian Process Classification. We also demonstrate the classification on an exemplary traffic jam approach scenario, showing that it is feasible and would lead to more efficient testing practices.

I. INTRODUCTION

We are seeing an increased introduction of automation technology into passenger cars and heavy goods vehicles, with the goal to ultimately reach fully automated vehicles (AVs). AVs offer the potential for significant reductions in accident rates, emissions and traffic congestion [1]. To that end, much research and development has gone into systems that automatically maintain adequate speed and keep within the lane by applying steering movements to aid the driver [2]. These systems rely on information from various sensors to perceive their environment. With this information about the environment, novel methods are able to shift the driving tasks of navigation and guidance, which are traditionally performed by the human driver, to be executed by the vehicle independently.

The barrier hindering higher levels of automation (SAE levels 3-5 [3]) to be introduced into public use is the assurance of their safety. In order to be accepted by the public, it needs to be validated that AVs are safe for them to use and can prevail in every situation that they encounter. Proving this with conventional testing methods would necessitate driving the AVs for hundreds of millions to billions of kilometres before rolling them out [4].

Current validation methods range from trials on public roads over Hardware-in-the-Loop (HiL) to simulation. Especially the validation of critical driving scenarios in simulation offers cost and feasibility benefits. A common approach to validate AVs in simulation is to conduct Monte Carlo experiments, where scenarios are randomly simulated. Both, public road trials and Monte Carlo simulation suffer from the same disadvantage however, which is that most

of the time spend driving on the road is uneventful [5]. Thus, a number of methods that limit the randomness and guide the search towards safety-critical scenarios have been proposed, mainly based on reducing the variation of the estimator, for example importance sampling. This accelerates the search significantly compared to Monte Carlo simulation [6], but requires knowledge of the distribution of safety-critical scenarios which can only be taken from real world tests.

In this paper, we present a novel approach to finding safety-critical scenarios by identifying the performance boundary of an AV. The performance boundary separates the scenario space into regions according to the outcome of the scenario [7]. The outcome of a scenario can be quantitatively judged by different criticality metrics, such as the frequently used Time-to-Collision (TTC) [8].

Along the performance boundary, small changes in the parameters that make up the scenario could result in a transition from a safe scenario to an unsafe one. These scenarios are often called *corner cases*, where the individual scenario parameters are within the capability of the system, but the combination challenges the system [9]. Identifying and testing for these corner case scenarios is crucial to enabling acceptable safety testing practices for AVs, and more widely, the widespread introduction of AVs [4].

To identify the performance boundary, Gaussian Process classification (GPC) is utilised (see Section III). The GPC can probabilistically predict the outcome of scenarios that were not tested, based on known scenarios and estimate where the performance boundary is located. Due to the probabilistic nature of the Gaussian Process (GP) emulator, the proposed method also efficiently enables us to provide a point-wise confidence measure around the predicted outcomes.

The remainder of this paper is organised as follows. In Section II we review relevant literature in the field of scenario-based AV validation and supervised machine learning algorithms for classification. In Section III we formally introduce Gaussian Processes and how they are used to solve classification tasks. In Section IV, the exemplary scenario considered in this work is detailed and the acquisition of the training and test data using two different sampling methods is explained. Section V shows the results of the trained GPC prediction model and the estimated performance boundary. Furthermore, we compare and analyse the differences that result from the individual data sets. The paper concludes with an outlook on future research directions in Section VI.

¹Institute for Future Transport and Cities, Coventry University, Coventry CV1 5FB, United Kingdom

²HORIBA MIRA, Nuneaton CV10 0TU, United Kingdom
Corresponding Author: batschf@coventry.ac.uk

This research is funded by HORIBA MIRA

II. RELATED WORK

A. Validation of AVs based on Scenarios

Most research on the safety evaluation of AVs focuses on testing through carrying out a driving scenario involving the AV along with other actors and specified environment conditions [10]. This stems from the problem, that the safe operation of an AV depends on the correct decision making of the AV, based on the complex interaction with other actors. This scenario-based validation is usually carried out in simulation, which offers the possibility to conduct many scenarios at low cost [11].

One approach to validate an AV is to conduct Monte Carlo simulations [11]. Here, the parameters of a scenario are randomly selected from a distribution (mostly uniform distribution). After running the simulations, it is checked if the AV completed all scenarios successfully. This approach is computationally (and in case of physical testing, financially) expensive as a large number of scenarios must be run to be statistically significant. But since it achieves an even coverage of the parameter space, it is often used as baseline for comparison to other methods [12].

Monte Carlo methods can be improved by matching the sampling distribution to the actual distribution of the problem parameters, instead of a uniform distribution, and thus reduce the variance of the sampling. Especially importance sampling methods received a lot of attention to improve the search for scenarios that can be problematic for an AV, often called *rare events* [5].

It was found that the evaluation time of a car following scenario can be reduced by up to a factor of 100,000, and for a lane change scenario by up to 20,000 [6], [13]. In [2], importance sampling was used to evaluate an adaptive cruise control system.

Another sampling method, Subset Simulation, was used on a lane change scenario in [5]. Subset simulation offers advantages for high dimensional stochastic models and shows a similar performance improvement as importance sampling.

The general disadvantage of variance reduction techniques is their requirement of prior knowledge in order to shape the probability distribution used for sampling. Often this data is taken from accident databases or naturalistic driving trails [6]. This runs the risk however to exclude scenarios that emerge due to the new automation technology, which is often not considered in accident databases or naturalistic driving trails.

B. Supervised Machine Learning Algorithms for Classification

Supervised machine learning has become a topic of much discussion in recent years and they have been applied to a range of problems, from image classification to control [14], [15].

Especially Artificial Neural Networks have received a lot of attention, leading to big improvements in their classification capabilities. They unfold their potential particularly in the area of image recognition and object classification,

where large amounts of annotated data have become widely available, reducing the problem of overfitting [14]. This is the crux however that makes them less attractive in the area of AV validation, where data from tests is either costly, if tests involving hardware are conducted, or computationally expensive, if simulation is involved.

Another supervised learning method is the Support Vector Machine (SVM), originally proposed in [16]. The advantage of SVMs is that they stay effective in the prediction of large dimensional data sets, even if the number of samples is small. Their efficacy in scenario prediction of human driving scenarios was shown in [17] for example. Compared to GPs, SVMs have some disadvantages however. They inherently do not provide an estimate on the confidence of the prediction, which is intrinsically provided by the GP for example. Furthermore, GPs are more flexible on custom kernel functions and learning their hyperparameters from data [18].

The k-Nearest Neighbour (k-NN) algorithm can also be used to classify data [19]. The k-NN algorithms predict new data points from known data in the vicinity of the predicted point. This makes it unreliable however, if data points close to or on the boundary between classes should be predicted, which is critical to finding corner case scenarios for the evaluation of automated vehicles.

III. GAUSSIAN PROCESSES

Gaussian Processes are a class of supervised machine learning algorithms, that describe a functional relation as a multivariate Gaussian distribution and can thus be used for non-linear regression and classification problems [18]. They have been used to model and predict trajectories of vehicles and pedestrians, as shown in [20] and [21]. Furthermore, GPs have been used to model the driving intention of human drivers, particularly for scenarios at intersections [22], [23].

Gaussian Processes have been relatively unexplored in the context of performance classification of AVs. In [7] Gaussian Process Regression (GPR) is used to adaptively search the state space of an autonomous, unmanned underwater vehicle. The authors also define a performance boundary, where the performance of the system transitions from one mode to another, due to changes in the environment. The concept of a performance boundary is further developed and adapted to ground AVs in this paper.

An application of GPR in an automotive context is presented in [24], where the GPR was used to estimate the probability distribution of a scenario, which was subsequently used for importance sampling. They study their procedure on a lane change scenario and show an improvement of the evaluation effort compared to crude Monte Carlo sampling.

The GPC in this application is trained on data obtained from a traffic jam approach scenario, which is further described in Section IV. The scenarios were executed using the vehicle simulation software CarMaker [25], and for the purpose of this paper, the simulation is regarded as ground truth. For this to be valid in the overall scope of validating

the AV, physical tests have to follow up and validate the simulation.

A. Formal Description of Gaussian Processes

The GP model used in this paper is based on the extensive work of [18]. The formal definition of a Gaussian Process is denoted by a prior distribution

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (1)$$

with a mean of $m(\mathbf{x})$ and the kernel function $k(\mathbf{x}, \mathbf{x}')$. The kernel function used in this paper is the radial basis function (RBF), also known as squared exponential kernel [18].

We consider a data set $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$, consisting of n samples, wherein \mathbf{x}_i denotes the vector of input data taken from the input space \mathcal{X} , and $y_i = f(\mathbf{x}_i)$ the corresponding output observations. With the definition of a Gaussian Process from Eq. (1), we can describe a joint prior distribution for the observed outputs \mathbf{f} and the predicted outputs \mathbf{f}_* :

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right), \quad (2)$$

with an assumed mean of zero and the covariance matrices $K = k(\mathbf{X}, \mathbf{X})$
 $K_*^T = k(\mathbf{X}, \mathbf{X}_*)$
 $K_* = k(\mathbf{X}_*, \mathbf{X})$
 $K_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$
for all observed and predicted data points. Here \mathbf{X} denotes a $d \times n$ matrix of the training inputs $\{\mathbf{x}_i\}_{i=1}^n$ (also known as the design matrix), d stands for the dimension of input space \mathcal{X} , and \mathbf{X}_* is the matrix of test inputs. The subscript * differentiates the test/predicted data from the observed data. To simplify the problem, the mean function is usually assumed to be zero, which does not limit the mean of the posterior to zero.

To obtain the posterior distribution over the predicted value we can condition the joint prior distribution to

$$\mathbf{f}_* | \mathbf{f}, \mathbf{X}, \mathbf{X}_* \sim \mathcal{N}(K_* K^{-1} \mathbf{f}, K_{**} - K_* K^{-1} K_*^T). \quad (3)$$

The predicted value \mathbf{f}_* and its uncertainty are thus given by the mean and covariance of the posterior distribution evaluated over \mathbf{X}_* .

B. Gaussian Process Classification

Considering a new data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, wherein y_i now describes a number of discrete class labels according to \mathbf{x}_i . GPs can be used to solve classification problems by giving predictions in form of class probabilities y_* . This is done by squashing the output of a regression model through a logistic function (e.g. sigmoid function, $\sigma(\cdot)$) to transform it from a domain of $(-\infty, \infty)$ to $[0, 1]$.

The classification is done in two steps, by first predicting a latent variable \mathbf{f}_* corresponding to an input value \mathbf{x}_*

$$p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int p(f_* | \mathbf{X}, \mathbf{x}_*, \mathbf{f}) p(\mathbf{f} | \mathbf{X}, \mathbf{y}) d\mathbf{f}. \quad (4)$$

The probabilistic prediction can then be calculated in the second step using

$$p(y_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int \sigma(f_*) p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) df_*. \quad (5)$$

The likelihood function in Eq. (4) is non-gaussian due to the discrete class labels in y . Therefore, a computationally feasible Laplace approximation must be used to approximate the integral.

In this paper, we predict AV scenario outcomes using GPC. In the remaining sections we describe this application of GPC on an exemplary scenario.

IV. DATA ACQUISITION THROUGH SIMULATION

The GPC was trained on data obtained from simulations conducted with the CarMaker simulation software [25]. We consider an exemplary scenario where an AV approaches a traffic jam in which vehicles are moving considerably slower than the approaching ego vehicle. Additionally, the end of the traffic jam is situated in a left turn of a curve with a radius of 50 m. To simplify the simulation, only the last vehicle in the traffic jam is modelled, and there are no additional obstructions between the ego vehicle and the traffic jam. The system under test is an automated vehicle equipped with a radar sensor to detect objects within its sensing arc. Sensor uncertainty was neglected here but will be considered in future work. If a vehicle is detected, the longitudinal control initiates a braking manoeuvre in order to avoid a collision. The scenario is illustrated in Fig. 1. The preset vehicle and radar sensor models of CarMaker were used in this work.

Three parameters were considered in this paper: the speed of the approaching ego vehicle (*speed ego*), the speed of the last vehicle in the traffic jam (*speed target*), and the aperture angle of the radar sensor of the ego vehicle (*aperture angle*). For feasibility purposes, the variation of the scenario has been restricted to these three parameters but adding more parameters, such as additional sensors and perception uncertainties, is in the scope of our future research. The parameters are varied within fixed ranges, which can be found in Table I. We note that excluding extreme parameter

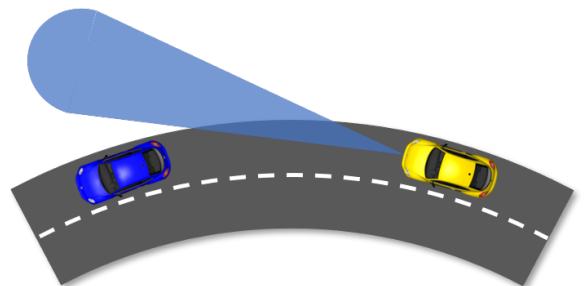


Fig. 1. Traffic jam approach scenario with the blue target vehicle and the yellow ego vehicle equipped with a radar sensor

TABLE I
SIMULATION PARAMETERS AND RANGES

Simulation Parameters	Lower limit	Upper limit
speed ego [km/h]	40	70
speed target [km/h]	5	20
aperture angle [deg]	10	25

combinations a priori is avoided, as this would make a possibly invalid assumption and exclude potential corner case scenarios.

The outcome of the simulated scenarios was evaluated on whether the AV was able to prevent a collision or not, resulting in a binary classification problem. It should be mentioned at this point however, that GPC can also be used for multi-class classification problems, in case multiple events should be considered and this is something we also consider for future work (see Section VI).

Two sampling methods were used to create data sets of different size in order to compare the efficacy of the sampling methods, and the amount of necessary training data. The data sets were obtained through Monte Carlo sampling and the Minimax Latin Hypercube design method. The rationale in including different sampling methods is explained in the following.

A. Monte Carlo Sampling

Baseline data sets to train and test the GPC prediction model were generated by sampling from a uniform distribution, confined to the limits described in Table I. The limits were chosen on representative values for the scenario, to skew the parameter space and exclude parameter combinations of no interest.

In the applied Monte Carlo sampling, every parameter had its own independent distribution. Monte Carlo sampling does not guarantee that the samples are homogeneously spread throughout the parameter space but is included here as it is frequently used as basis for comparison [12].

B. Latin Hypercube Sampling

Furthermore, data sets were created using Latin Hypercube (LHC) sampling [26]. LHC ensures that the parameter space is evenly covered by dividing each parameter into intervals which are then randomised. This reduces the risk that large areas of the parameters space remain uncovered or that samples are too close together.

The generated data sets were split into a training and a test set. The training set was used to tune the hyperparameters of the GPC. The different data sets and their sizes can be found in Table II.

V. RESULTS

A. Finding the Performance Boundary

Separate Gaussian Process classifications were trained and tested on the four data sets described in Table II. For the scenario described in Section IV, a clear performance

TABLE II
DATA SETS

Name	Sampling Method	Size training set	Size test set
MC100	Monte Carlo	90	10
MC1000	Monte Carlo	900	100
LHC100	Latin Hypercube	90	10
LHC1000	Latin Hypercube	900	100

boundary was found, which separates the scenarios that ended successfully, i.e. the AV could prevent a collision, from the ones that were unsuccessful and ended in a collision. The performance boundary along with the training and test data of the MC100 data set is visualised in Fig. 2.

While the MC100 data set yielded good classification results, with no misclassifications on the test set, there is the possibility that a misclassification occurs, if a test data point is very close to the performance boundary and thus not included in its prediction. For the LHC100 test data set one misclassification occurred, where a test data point was on the performance boundary, indicated by the white circle in Fig. 3. If this data point would be included in the training set, the performance boundary would have shifted accordingly.

One exemplary scenario on the performance boundary is depicted as a black dot in Fig. 3 (white arrow). The explicit parameter values of this predicted scenario on the performance boundary and the two closest, simulated scenarios on

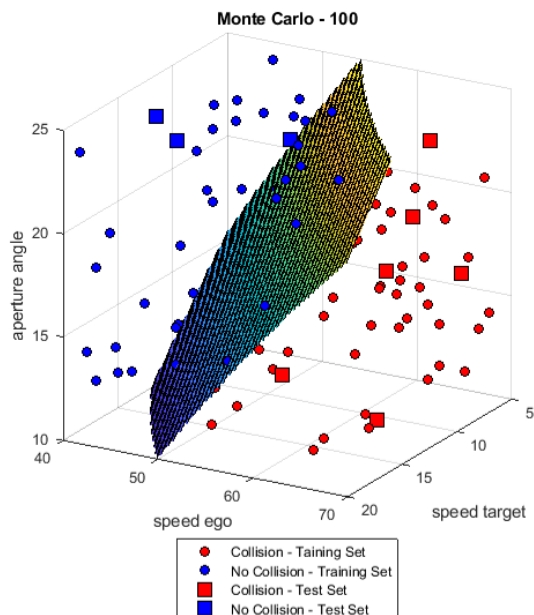


Fig. 2. Boundary estimation based on the MC100 data set

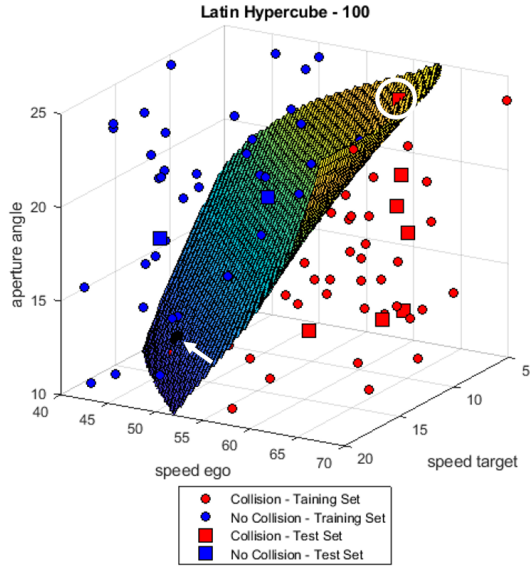


Fig. 3. Boundary estimation based on the LHC100 data set with a single misclassification marked by a white circle and an exemplary scenario on the performance boundary indicated by a black dot (white arrow)

either side of the boundary are listed in Table III.

Furthermore, it can be seen that training the GPC on the two different data sets yields slightly different performance boundary estimates. This is due to the sparsity and different distributions of data in the sets and leads to a maximum Euclidean distance of 4.5 between the two performance boundary estimates. This deviation of the estimated performance boundary can be minimised by increasing the number of data points. The GPC on the MC1000 and LHC1000 data sets with 1000 data points reduced the maximum Euclidean distance between the performance boundary estimates to 1.25.

Additionally, we analysed which sampling method yields a better estimation of the performance boundary on sparse data. The Latin Hypercube sampling method had a slightly better estimation, as the maximum Euclidean distance between the performance boundary estimations of the LHC100 and LHC1000 data set was 3.75. The maximum Euclidean distance between the performance boundary estimations of the MC100 and the MC1000 data set was found to be 4.75.

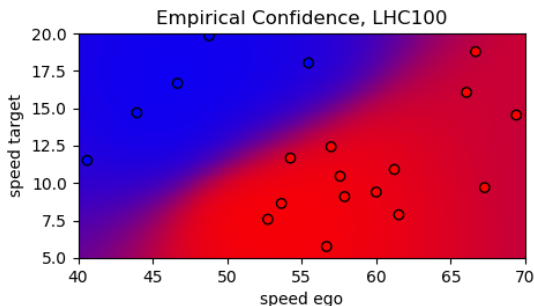


Fig. 4. Empirical confidence provided by the Gaussian Process classification at a constant aperture angle of 17.5° . The data points displayed are in the range $[16, 19]^\circ$.

TABLE III
EXEMPLARY SCENARIOS ON THE PERFORMANCE BOUNDARY
(LHC100 DATA SET)

speed ego [km/h]	speed target [km/h]	aperture angle [deg]	Scenario Outcome
47.27	15.76	11.36	Collision
46.97	15.30	13.33	No-Collision
47.25	15.5	12.25	Boundary

B. Uncertainty Measure of the Classification

An advantage of the GPC is that it provides a measure on the uncertainty of the prediction model. Since the prediction is probabilistic, we can calculate the empirical confidence of estimations, which gives additional information on the prediction.

In Fig. 4 the prediction output of the GPC model trained on the LHC100 data set is shown for a constant aperture angle of 17.5° . Also displayed are the data points that have the biggest influence on the estimation of the performance boundary. These are the data points located in the vicinity of the constant third parameter; data points with an aperture angle on the interval $[16, 19]^\circ$. As can be seen from Fig. 4, the performance boundary is fuzzy, especially in the areas close to the limits of the data set. This gives a direct indication on the confidence level of predictions from the GPC model.

With an increased number of data points, especially close to the true location of the performance boundary, the empirical confidence is increased and the performance boundary can be drawn much sharper. Fig. 5 shows the empirical confidence of the LHC1000 data set. It is clearly visible that the increased number of data points, which are here on an interval of $[17, 18]^\circ$, increase the confidence of the prediction along the performance boundary area. Similar to Fig. 4 the empirical confidence decreases towards the limits of the data set, due to missing data beyond the limit.

VI. CONCLUSIONS AND FUTURE WORK

In this work, we applied Gaussian Process classification to the problem of automated vehicle validation. It was found

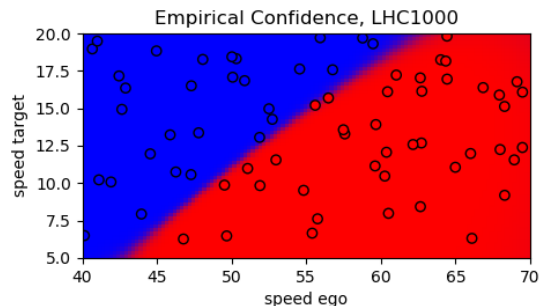


Fig. 5. Empirical confidence provided by the Gaussian Process classification at a constant aperture angle of 17.5° . The data points displayed are in the range $[17, 18]^\circ$.

that the performance boundary, which separates successful from unsuccessful scenarios, can be adequately estimated from simulated data.

Knowledge about the location of the performance boundary can be used to identify corner cases, which are the scenarios on the performance boundary. We detailed the predicted parameters of such a scenario in Table III. These scenarios can be used to test the automated vehicle and help to identify erroneous behaviour or limitations in the design of the systems.

It was shown that a Latin Hypercube design of the data set yields a better prediction than randomised Monte Carlo sampling when looking at sparse data sets, an advantage which decreases, however, with larger data sets. Furthermore, the amount of available data around the performance boundary naturally improves the confidence of the prediction model on the location of the performance boundary.

The disadvantage of Gaussian Process classification lies in its lack of scalability, as it scales with $\mathcal{O}(n^3)$. High dimensionality of the input might thus necessitate dimensionality reduction methods such as sparse Gaussian Processes to stay computationally viable [27].

In future work, we plan to integrate the prediction model in an adaptive framework, to concentrate the sampling around the performance boundary and thus reduce the necessary simulations to find the performance boundary. Additionally, we plan to extend the model to include a multi-dimensional parameter space and multiple criticality classes, which the scenarios are classified in. These could for example include a classification of near miss scenarios. Further investigation is also necessary to look at the influence of different kernel functions and their impact on the classification.

REFERENCES

- [1] U. Montanaro, S. Dixit, S. Fallah, M. Dianati, A. Stevens, D. Oxtoby, and A. Mouzakitis, "Towards connected autonomous driving: review of use-cases," *Vehicle System Dynamics*, pp. 1–36, jul 2018
- [2] O. Gietelink, B. De Schutter, and M. Verhaegen, "PROBABILISTIC VALIDATION OF ADVANCED DRIVER ASSISTANCE SYSTEMS," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 97–102, 2005
- [3] SAE, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE, Tech. Rep., 2018
- [4] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, dec 2016
- [5] S. Zhang, H. Peng, D. Zhao, and H. E. Tseng, "Accelerated Evaluation of Autonomous Vehicles in the Lane Change Scenario Based on Subset Simulation Technique," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, HI: IEEE, nov 2018, pp. 3935–3940
- [6] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc, "Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 733–744, mar 2018
- [7] G. E. Mullins, P. G. Stankiewicz, R. C. Hawthorne, and S. K. Gupta, "Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles," *Journal of Systems and Software*, vol. 137, pp. 197–215, mar 2018
- [8] M. Schreier, V. Willert, and J. Adamy, "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2751–2766, oct 2016
- [9] DIN SAE SPEC 91381:2019-06, "Terms and Definitions Related to Testing of Automated Vehicle Technologies; Text in German and English," DIN, Standard, 2019
- [10] A. Zlocki, L. Eckstein, and F. Fahrenkrog, "Evaluation and Sign-Off Methodology for Automated Vehicle Systems Based on Relevant Driving Situations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2489, pp. 123–129, jan 2015
- [11] S. Khastgir, G. Dhadyalla, S. Birrell, S. Redmond, R. Addinall, and P. Jennings, "Test Scenario Generation for Driving Simulators Using Constrained Randomization Technique," *SAE Technical Paper*, vol. 2017-01, no. 1672, pp. 1 – 7, mar 2017
- [12] M. Althoff and A. Mergel, "Comparison of Markov Chain Abstraction and Monte Carlo Simulation for the Safety Assessment of Autonomous Cars," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1237–1247, dec 2011
- [13] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan, "Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 595–607, mar 2017
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, vol. 25, pp. 1097 – 1105, 2012
- [15] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, and H. Radha, "Deep learning algorithm for autonomous driving using GoogleNet," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. Los Angeles: IEEE, jun 2017, pp. 89–96
- [16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, sep 1995
- [17] F. Remmen, I. Cara, E. de Gelder, and D. Willemsen, "Cut-in Scenario Prediction for Automated Vehicles," in *2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. Madrid: IEEE, sep 2018, pp. 1–7
- [18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive computation and machine learning. Cambridge, Mass.: MIT Press, 2006
- [19] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, aug 1992
- [20] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," in *2011 International Conference on Computer Vision*. Barcelona: IEEE, nov 2011, pp. 1164–1171
- [21] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectory patterns with Gaussian processes," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. Kyoto: IEEE, sep 2009, pp. 1229–1234
- [22] Q. Tran and J. Firl, "Modelling of traffic situations at urban intersections with probabilistic non-parametric regression," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. Gold Coast: IEEE, jun 2013, pp. 334–339
- [23] A. Armand, D. Filliat, and J. Ibanez-Guzman, "Modelling stop intersection approaches using Gaussian processes," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. The Hague: IEEE, oct 2013, pp. 1650–1655
- [24] Z. Huang, H. Lam, and D. Zhao, "Towards affordable on-track testing for autonomous vehicle A Kriging-based statistical approach," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. Yokohama: IEEE, oct 2017, pp. 1–6
- [25] IPG Carmaker, "https://ipg-automotive.com/," 2019
- [26] M. D. McKay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, vol. 21, no. 2, pp. 239 – 245, may 1979
- [27] N. Lawrence, M. Seeger, and R. Herbrich, "Fast Sparse Gaussian Process Methods: The Informative Vector Machine," in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, Mass.: MIT Press, 2002, pp. 625 – 632