A Realistic Cyclist Model for SUMO Based on the SimRa Dataset

Ahmet-Serdar Karakaya[1], Konstantin Köhler[1], Julian Heinovski[2], Falko Dressler[2], and David Bermbach[1]

[1]*TU Berlin & ECDF, Mobile Cloud Computing Research Group*, {ask,koko,db}@mcc.tu-berlin.de
[2]*TU Berlin, Telecommunication Networks*, {heinovski,dressler}@ccs-labs.org

Abstract—**Increasing the modal share of bicycle traffic to reduce carbon emissions, reduce urban car traffic, and to improve the health of citizens, requires a shift away from car-centric city planning. For this, traffic planners often rely on simulation tools such as SUMO which allow them to study the effects of construction changes before implementing them. Similarly, studies of vulnerable road users, here cyclists, also use such models to assess the performance of communication-based road traffic safety systems. The cyclist model in SUMO, however, is very imprecise as SUMO cyclists behave either like slow cars or fast pedestrians, thus, casting doubt on simulation results for bicycle traffic. In this paper, we analyze acceleration, velocity, and intersection left-turn behavior of cyclists in a large dataset of real world cycle tracks. We use the results to derive an improved cyclist model and implement it in SUMO.**

Index Terms—**Urban planning, Motion sensor data, Sensor data analysis, Traffic simulation**

I. INTRODUCTION

Active transportation modes such as cycling provide health benefits, alleviate traffic congestion, and reduce air pollution [1]. In practice, however, cyclists often face a car-centric traffic infrastructure which has a significant impact on their (perceived) safety and also affects the attractiveness of cycling routes [2]–[4]. Changing this infrastructure to better accommodate cyclists and pedestrians requires significant planning efforts of city planners and traffic engineers. Similarly, road traffic safety systems for vulnerable road users are often assessed using simulation. Particularly the interaction with cars is relevant when it comes to V2X-based safety systems for cyclists [5]. Many of these studies rely on the open source simulation platform SUMO[1] (Simulation of Urban Mobility), which allows them to study the effects of infrastructure changes before implementing them on the streets.

In SUMO, vehicles and their dynamics are simulated individually [6]. Unfortunately, the cyclist model is not particularly realistic – cyclists can either be modeled to behave as slow cars or as fast pedestrians. Several studies have already improved the bicycle model of SUMO. For instance, Kaths and Grigoropoulos [7] investigated the intersection behavior of cyclists using camera traces and transferred findings into SUMO. Also, Grigoropoulos et al. [8] improved modeling of bicycle infrastructure at intersections while Heinovski et al. [9] created a virtual cycling environment to import real bicycle behavior directly into SUMO. Nevertheless, the current cyclist behavior in SUMO is still rather unrealistic; so far, researchers have devoted much more effort to car models, e.g., [10]–[17]. One reason for this is that, until recently, not enough data on real-world cyclist behavior have been available. Today, crowdsourced data collection approaches such as SimRa[2] [2] have made thousands of cycle tracks available as open data.

In this paper, we analyze the SimRa dataset regarding acceleration and velocity of cyclists as well as their left-turn behavior in four-way intersections. We then use our findings to improve the cyclist model in SUMO. In this regard, we make the following contributions:

- We show that SUMO's default bicycle simulation is not realistic (Section IV),
- we improve bicycle simulation in SUMO by deriving new parameters for that vehicle type in SUMO (Section V),
- we develop an intersection model which captures cyclists' left-turn behavior at intersections in a more realistic way (Section V), and
- we compare our improvements to SUMO's default bicycle simulation, using the SimRa dataset as a ground truth (Section VI).

II. BACKGROUND

In this section, we give an overview of SUMO (see Section II-A) and SimRa (see Section II-B), which provided the dataset we used in our work.

A. SUMO

SUMO is an open source traffic simulation tool that offers macroscopic as well as microscopic simulation of vehicle mobility [6]. SUMO includes models for different types of "vehicles", including, among others, cars, bicycles, and even pedestrians. Due to its large feature set, it has become the de-facto standard for traffic simulation and is used even beyond the transport community, e.g., [18].

Traffic scenarios are, among other things, defined by road networks and vehicle traffic. The road network includes roads and their (sub-)lanes as well as exclusive lanes for cyclists and pedestrians, or road-side infrastructure such as traffic lights. Furthermore, connections between these lanes and traffic lights can be configured.

---

[1]https://www.eclipse.org/sumo/

[2]https://github.com/simra-project/

When modeling vehicle traffic, users specify demand for a specific road segment per vehicle type and can adjust vehicle-specific parameters of SUMO's simulation model to control their respective behavior. In general, vehicle parameters are usually specified in the vehicle type declaration (*vType*), applying the changes to all instances of the respective *vType*, e.g., to all cars. An alternative, however, is to obtain multiple *vType* realizations which typically differ in at least one parameter by using so-called *vTypeDistributions*. This way, when spawning a new vehicle, SUMO randomly picks a specific *vType* from the *vTypeDistribution* and instantiates the vehicle's parameters accordingly, e.g., cars can thus have individual maximum velocities.

In SUMO, vehicle behavior is, among other things, defined by *Car Following (CF) models* for the longitudinal kinematic behavior, *Lane Change (LC) models* for the lateral kinematic behaviour, and *junction models* for the behavior at junctions and intersections.

Despite including several of these models for cars and trucks, SUMO does not provide a dedicated movement model for cyclists. Instead, cyclists are simulated by modeling them either as slow cars or fast pedestrians. Both of these approaches use movement models of the corresponding vehicle type and adapt their respective shape and kinematic characteristics (e.g., velocity and acceleration profiles) to match cyclists. While this is obviously a rough approximation, it is unlikely to reflect the behavior of real-world cyclists [8].

B. SimRa

SimRa is an open source project started in 2019 which aims to identify hotspots of near miss incidents in bicycle traffic [2], [19]. For this, the project follows a crowdsourcing approach in which cyclists record their daily rides using a smartphone application available for both Android and iOS. Today, the project has managed to record more than 65,000 rides, most of them in Germany, approximately half of them in Berlin.

During the ride, SimRa records the GPS trace at 1/3Hz and the motion sensors, i.e., (linear) accelerometer, gyroscope, and rotation vector at 50Hz; the motion sensor readings are aggregated by calculating a moving average with a window size of 30 and then keeping only every fifth value. This was done for saving memory, battery, and mobile data usage while still being able to reconstruct the ride and detect near miss incidents. After the ride, SimRa shows the recorded ride as a route on the map which is then annotated, cropped (for privacy reasons), and uploaded by the user. In this paper, we only use measured data from the ride files and disregard user-annotated data on near miss incidents.

III. RELATED WORK

In this section, we give an overview of related work on improving intersection behavior (Section III-A) and longitudinal (Section III-B) behavior of cyclists in SUMO's simulation models.

A. Intersection Behavior of Cyclists

Kaths and Grigoropoulos [7] aim to address the shortcomings of SUMO's intersection model for cyclists. For this, they record video footage of an example intersection in Munich and derive cyclist trajectories. From the set of trajectories, they select one representative trajectory for each combination of start and end points in the intersection and make it available to SUMO via an external API. While this is a significant improvement in realism over SUMO's intersection model, it is hard to generalize to other intersections and cannot cover the plurality of trajectories chosen by real-world cyclists.

Similar to Kaths and Grigoropoulos [7], Grigoropoulos et al. [20] analyze video footage of intersections with the goal of better understanding the intersection behavior of cyclists. Their focus, however, is not on deriving an improved intersection model but rather on identifying best practices for traffic planners working on intersections with high volumes or cycling traffic. Grigoropoulos et al. [8] propose to adjust the default traffic infrastructure inside SUMO to achieve more realistic cyclist behavior at intersections. Here, they focus on the number and shape of bicycle lanes which, however, are highly specific and differ from intersection to intersection.

B. Longitudinal Behavior of Cyclists

Twaddle and Grigoropoulos [21] examine four models for the longitudinal kinematic behavior of cyclists, i.e., acceleration and velocity. The first, called Constant Model, is the most simple one and is the SUMO default: Cyclists accelerate and decelerate at a constant rate until the desired velocity is reached. This model works well when breaking to a full stop but leads to frequent acceleration jumps between a fixed positive or negative value and zero, which is not realistic cyclist behavior. In the Linear Decreasing Model, maximum acceleration is reached, when starting the acceleration maneuver and then linearly declines until the desired velocity is reached. This model is outperformed by all other models. In the third and fourth models, Polynomial and Two Term Sinusoidal Model, acceleration or deceleration start at zero and then gradually grow over time. In their paper, Twaddle and Grigoropoulos [21] analyze the video recordings of 1030 rides in four intersections in Munich, Germany and conclude that the Polynomial Model has overall the most realistic cyclist behavior but is, however, not trivial to implement in SUMO.

A different approach of achieving realistic cycling behavior in SUMO is taken by Heinovski et al. [9]. The authors simulate multiple traffic scenarios in which accidents between cars and cyclists occur to investigate the effects of wireless communication between cyclists and other road users in the context of accident prevention. In order to obtain realistic cycling behavior for SUMO, they set up a novel Virtual Cycling Environment (VCE) featuring an actual bicycle that is connected to the simulation via multiple sensors. The VCE supports interactive empirical studies in a physically safe environment and allows the authors to record the cyclists' behavior in the form of trajectories. They use a set of recorded trajectories from different cyclists for emulating realistic cycling

behavior inside SUMO to simulate accidents. Although their approach produces trajectories from cyclists created with an actual bicycle, it does only achieve limited realism, since no other road users were present when recording the trajectories. Furthermore, deriving a realistic set of trajectories requires a large number of test persons.

IV. Cycling Behavior in SimRa and SUMO

In this section, we analyze real-world cyclists' behavior extracted from the SimRa dataset and compare it to the behavior of SUMO's default bicycle model. We analyze acceleration and velocity behavior in Sections IV-A and IV-B before discussing left-turn behavior at four-way intersections in Section IV-C. We omit a detailed discussion of the right-turn behavior at intersections, since SUMO's default model does not deviate much from the behavior observed from SimRa's dataset. When referring to SUMO's bicycle model, we refer to the "slow car" model of SUMO as the "fast pedestrian" model occasionally leads to poor results and was therefore not considered further.

Aside from the public SimRa datasets [22], [23] and more recent rides available on GitHub,[3] we also used non-public rides which have, for privacy reasons, not been published yet. Table I summarizes the most important attributes of the dataset that we used.

Since SimRa's dataset stems from crowdsourced smartphone data generation, it suffers from poor sensor quality [24], [25] as well as heterogeneous hardware and users [26]. To achieve the best possible data quality, we tested various pre-processing techniques and filters. We also conducted an experiment in which sample trajectories were recorded in parallel on several SimRa client devices and compared to a ground truth trajectory recorded by a stand-alone GPS receiver. In the end, we used a Gaussian Kernel filter for improving location data and a Low Pass filter for the velocity data. After filtering semantically and syntactically defective files, we used data from 43 961 rides as input for our analysis scripts.[4]

Since most of the data were recorded in Berlin, Germany, all examples in the following focus exclusively on Berlin scenarios.

A. Acceleration

For analyzing cyclist acceleration, we extracted acceleration maneuvers from the dataset. For this, we slightly adapted the approach of [27] and found 140 736 acceleration maneuvers in the cleaned dataset. Distribution fitting processes showed

[3] https://github.com/simra-project/dataset
[4] https://github.com/simra-project/SimRaXSUMO

Table I
MOST IMPORTANT ATTRIBUTES OF ENTITIES THE SIMRA DATASET

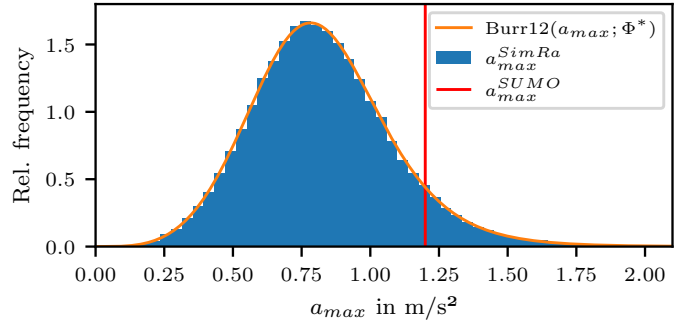|	Total	Used
Rides	57 662	43 961
Acceleration Maneuvers	1 922 087	140 736

Figure 1. Histogram of the empirical maximum acceleration capabilities of cyclists found in the SimRa dataset and the fitted EMG. The red scalar represents the default value in SUMO.

that the Burr (Type XII) distribution [28] $Burr(a_{max}; \Phi*)$ fits the data best (see also Fig. 1).

Comparing the acceleration capability of actual cyclists (the SimRa dataset) with the default SUMO bicycle model, differences become apparent. By default, SUMO specifies a_{max}^{SUMO} with $1.2 m/s^2$. This deviates significantly from the findings in the SimRa dataset where only 7.7% of the acceleration maneuvers are executed with a maximum acceleration of $1.2 m/s^2$ or higher. Furthermore, the empirical distribution is rather wide, indicating a broad variance across different cyclists and cycling situations, which is in stark contrast to SUMO's strategy of choosing a fixed maximum value.

B. Velocity

To gain insights into cyclists' behavior regarding their velocities, we calculate the maximum velocity for each ride file in the cleaned SimRa dataset. Using distribution fitting, we found that the S_U distribution Johnson [29] $JSU(v_{max}; \Phi*)$ fits the empirical data best (see also Fig. 2) and is therefore a valid fit for the specification of the empirical distribution of v_{max}^{SimRa}.

On the other hand, SUMO sets v_{max}^{SUMO} at $5.56 m/s$ by default. This deviates significantly from the findings in the SimRa dataset where 86.7% of the rides have a higher maximum velocity. Bringing this together with the acceleration findings, real-world cyclists often (but not always) cycle much faster than SUMO cyclists and vary much more in their acceleration behavior.

C. Left-turn Behavior at Intersections

According to the SimRa dataset, cyclists either behave like cars (using the normal road) or pedestrians (using the pedestrian crossing) to take left-turns at intersections. We call the former a *direct* left turn and the latter an *indirect* left turn.

SUMO's default model only provides cyclists with the *direct* left turn behavior, thus significantly limiting the simulation's degree of realism (see also Fig. 3).

Taking a closer look at real world intersections in the SimRa dataset revealed that there are two distributions of left-turn behavior in practice. In the first, the *indirect* path is chosen with a probability of $57 \pm 7\%$, while on intersections following the
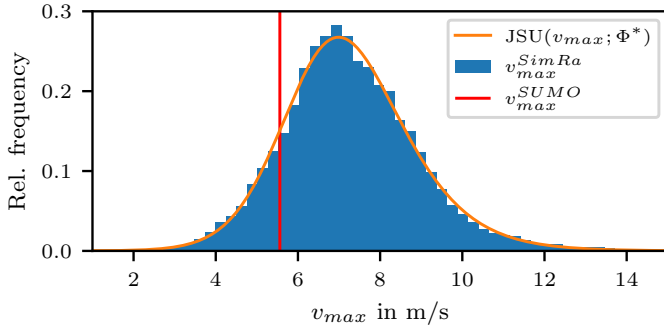
Figure 2. Histogram of the empirical maximum velocity data and the fitted JSU. The red scalar represents the default value in SUMO.
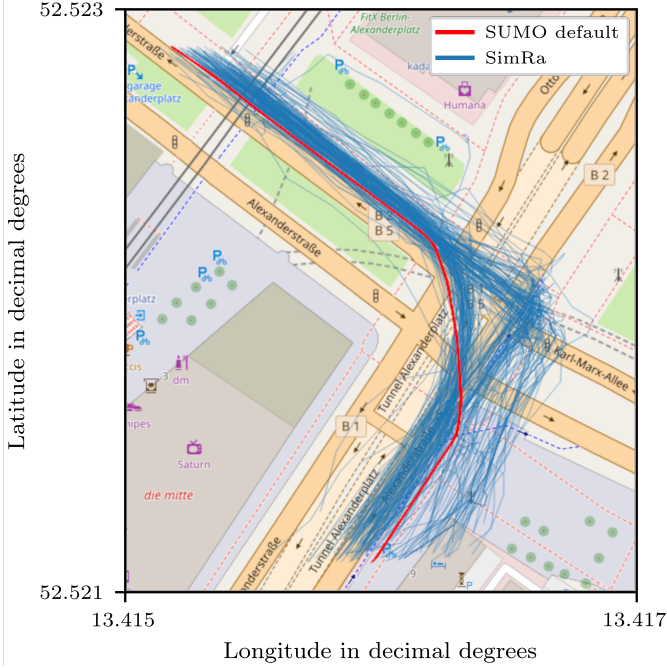

Figure 3. Qualitative comparison between the SUMO default intersection model and real world data given by SimRa for the intersection between Alexanderstraße and Karl-Marx-Allee in Berlin. SimRa shows two distinct left-turn paths (i.e., a direct and an indirect one) whereas SUMO default only models the direct path.

second, almost all cyclists choose the *indirect* path. Randomly investigating intersections of both types revealed that the first type has no specific characteristics while the second type actively encourages cyclists to indirect turns through the design of the intersection, e.g., by having a traffic island in the center. Since such information cannot be identified in Open Street Map (OSM) data reliably and in an abstract way, we will consider only the first distribution in the following.

SUMO and real world data differ decisively in all metrics considered, namely acceleration, velocity, and left-turn behavior at intersections. However, these three metrics are crucial for realistically simulating bicycle traffic. In the following, we try to adapt SUMO to simulate a more realistic cyclist behavior.

V. IMPROVING SUMO'S BICYCLE SIMULATION

To improve the simulation, we propose two changes to SUMO's bicycle model: First, the longitudinal kinematic parameters of SUMO's default bicycle model are (re-)parameterized based on the findings from the SimRa dataset. Second, a novel simulation model is derived from SimRa trajectories to exclusively simulate realistic left-turn bicycle behavior at intersections based on the findings in Section IV. The latter model is referred to as the intersection model in the following.

A. Longitudinal Kinematic Behavior

In Section IV, we derived maximum acceleration and maximum velocity characteristics from the SimRa dataset. We now use them to improve the longitudinal kinematic behavior of the default SUMO bicycle model. Contrary to the default parameterization, we use theoretical distribution functions instead of scalar values for the exposed kinematic parameters. This enables the model to produce more realistic bicycle simulation results since the heterogeneity of real world cycling styles is reflected.

We derive the theoretical distributions by aggregating the respective features from Sections IV-A and IV-B. For this, we rely on the *law of large numbers* which states that the average of the results obtained from a large number of trials of the same experiment eventually converges to its true expected value [30]. In the context of this work, this means that individual rides do not matter but that the aggregates of multiple rides will converge towards their actual expected value given a sufficiently large number of rides.

In our analysis, we found no correlation between a_{max} and v_{max} and, hence, decided to use both independently in our kinematic model. For the implementation, we used *vTypeDistributions* following the results of our previous analysis and sample both distribution independently.

It should be noted that through the parameterizations with theoretical probability density functions SUMO's *speedDev* parameter becomes obsolete as variance between the kinematic preferences among cyclists are already represented by the distribution function.

B. Left-turn Behavior at Intersections

To improve the degree of realism in cyclists' left-turn behavior at signaled intersections, we use an adapted version of the external intersection model (a Python script that steers cyclists via SUMO's *Traffic Control Interface*) as proposed by Kaths and Grigoropoulos [7] which is based on previously recorded real-world trajectories as their guidelines for cyclists across a single predefined intersection. Our approach algorithmically synthesizes the cyclists' trajectories (i.e., their respective guidelines across the intersection) for any regular four-way intersection and can therefore be seen as a step towards a more universal solution.

The left-turn maneuver distribution, as we call it, specifies the probability of the cyclists choosing either the *direct* or the *indirect* path to cross the intersection. For this, we use the distribution derived in Section IV-C as the default for our
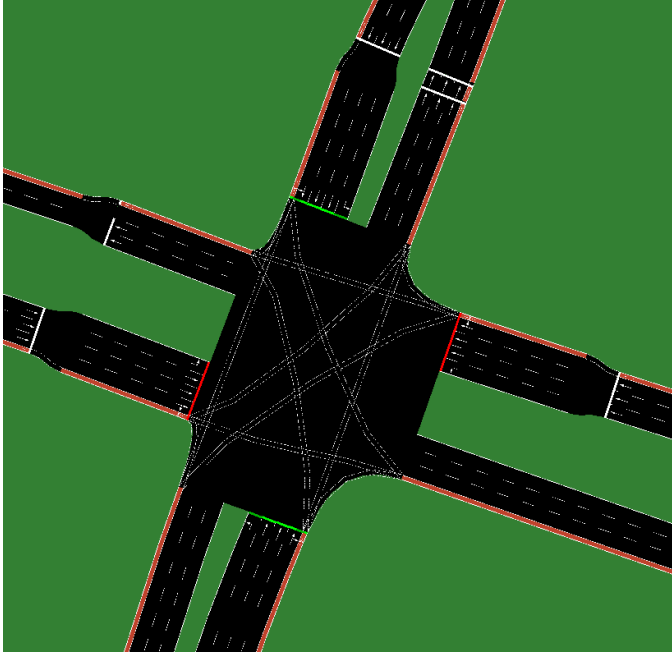
Figure 4. Excerpt from the *Mehringdamm* scenario in SUMO. The scenario was created using OSM data only.

intersection model. Users, however, can adjust the distribution if desired or needed for their specific purposes (see also the exception cases in Section IV-C).

VI. EVALUATION

In this section, we evaluate our new approach from Section V by comparing it to SUMO's default simulation model and the real-world data taken from the SimRa data set. We start by analyzing acceleration (Section VI-B), velocity (Section VI-C), and left-turn behavior at intersections (Section VI-D) before evaluating the combination of all model extensions (Section VI-E). Please note: While it may appear obvious that using the SimRa data set for both parameterization and evaluation should lead to perfect results, this is not the case as our extensions are subject to the design restrictions imposed by SUMO

A. Simulation Setup

As SUMO users can import real-world scenarios from OSM data, simulation results can be compared to real-world data and thus be evaluated. For our evaluation, we chose specific traffic scenarios that are representative and likely to showcase both strengths and weaknesses of our extensions. For evaluating the longitudinal behavior (acceleration and velocity), we chose urban traffic scenarios with long straight sections. As an example location, we use *Oranienstraße* in Berlin. For evaluating left-turn behavior, we chose compact scenarios around signaled intersections. For this, we study three intersections in Berlin, namely at *Mehringdamm* (see also Fig. 4), *Warschauer Straße*, and *Alexanderstraße*.

For our evaluation, we use SUMO version 1.7.0 and a step size of $1\,\mathrm{s}$ in simulations. The SUMO default results
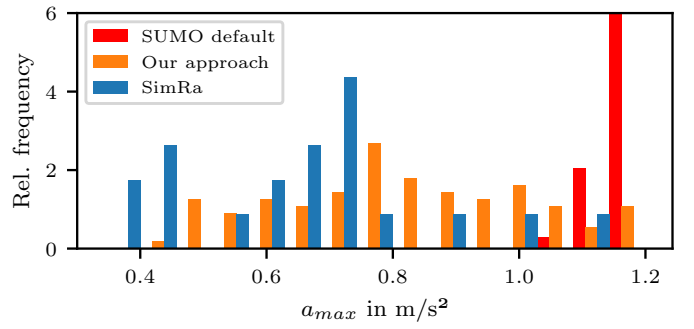

Figure 5. Histogram of SUMO's, SimRa's, and our approach's observed maximum accelerations inside the example *Oranienstraße* scenario. While the maximum accelerations are heterogeneously distributed in the real-world data and our approach, the default values are clustered. Here, the observed accelerations deviate from the configured default value ($1.2\mathrm{m/s}^2$) due to traffic effects inside the simulation.
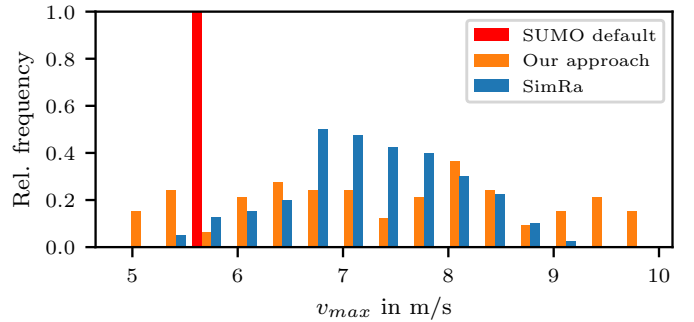

Figure 6. Histogram of SUMO's, SimRa's, and our approach's observed maximum velocities inside the example *Oranienstraße* scenario. While the maximum velocities are heterogeneously distributed in the real-world data and our approach, the default values are clustered.

are obtained with SUMO's *vType* `Bicycle` for cyclists. Thus, the maximum acceleration and maximum veocity are scalars and set to $1.2\,\mathrm{m\,s^{-2}}$ and $5.6\,\mathrm{m\,s^{-1}}$ respectively.

B. Acceleration

Fig. 5 shows the empirical distributions of the maximum acceleration among cyclists inside a specific simulation scenario and the corresponding real world data. It is evident that real-world acceleration maneuvers show heterogeneous maximum rates of acceleration. Apparently, the default parameterization is not suitable to describe this acceleration behavior among cyclists, as it provides homogeneous maximum acceleration rates within the simulation. Our new parameterization is significantly closer to the real-world behavior in the SimRa data set with its highly heterogeneous behavior across cyclists. That our new parameterization is not a perfect fit indicates that there are probably additional influence factors, e.g., the traffic density or the the weather situation, not covered in our kinematic model which aggregates data from all SimRa rides.

C. Velocity

Fig. 6 shows the empirical distributions of the cyclists' maximum velocities in the same simulation and the real-world
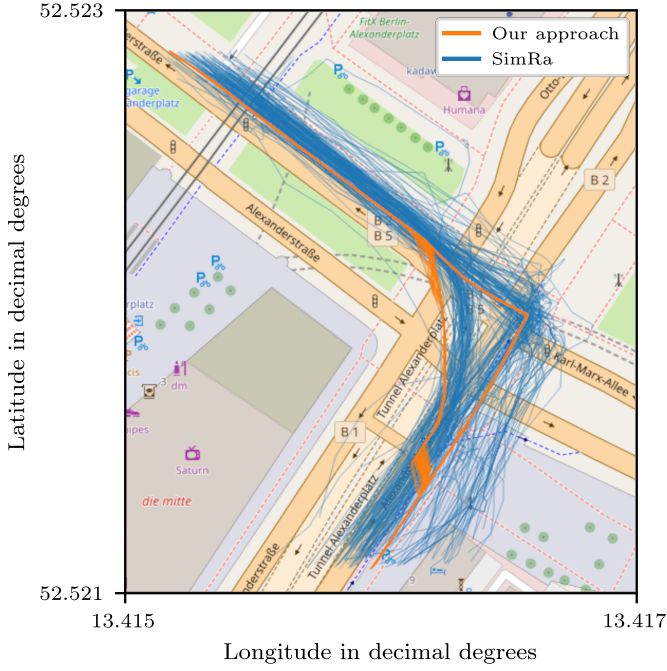
Figure 7. Qualitative comparison between the results of our approach for the intersection model and real-world data given by SimRa for the intersection between Alexanderstraße and Karl-Marx-Allee in Berlin. SimRa shows two distinct left-turn paths (i.e., a direct and an indirect one), which are also modeled by our approach.

Figure 8. ECDFs of the measured durations for crossing the scenario *Warschauer Straße*. It is apparent that our approach outperforms SUMO's default as the measured durations converge towards the real-world data.

Figure 9. ECDFs of the measured durations for crossing the scenario *Mehringdamm*. The results when using our new approach are only slightly more realistic than when using the standard SUMO model. However, when the *direct* path is blocked for cyclists, the simulation results outperform the default approach.

scenario. As with maximum acceleration rates, maximum velocities vary widely among real-world cyclists. Once more, the default parameterization is not able to reflect this characteristic. Our new parameterization is thus significantly closer to the real-world behavior of cyclists. As for acceleration behavior, the fact that our new parameterization is not a perfect fit to the real-world data indicates that there are likely to be additional influence factors not captured in our model.

D. Left-turn Behavior at Intersections

As shown in Fig. 7, which shows the intersection between Alexanderstraße and Karl-Marx-Allee in Berlin, the 2D trajectories produced by the new intersection model converge towards the trajectories of the SimRa data set. While the trajectories produced by SUMO's default bicycle model only offer direct turns (we omit those in the figure), the new model is significantly closer to real-world intersection behavior of cyclists.

E. Combining Intersection Model and Kinematic Extensions

To achieve a holistic comparison between SUMO's default bicycle model and our new approach, we measure the durations of left-turn maneuvers at multiple intersections and compare the empirical distributions of these measurements. To specifically monitor the impact of our changes, we do not include any ride time before or after the intersection in the measurements.

Based on this, we identified the following three findings:

First, our new approach outperforms the default at most intersections, as its measured durations converge with real data,
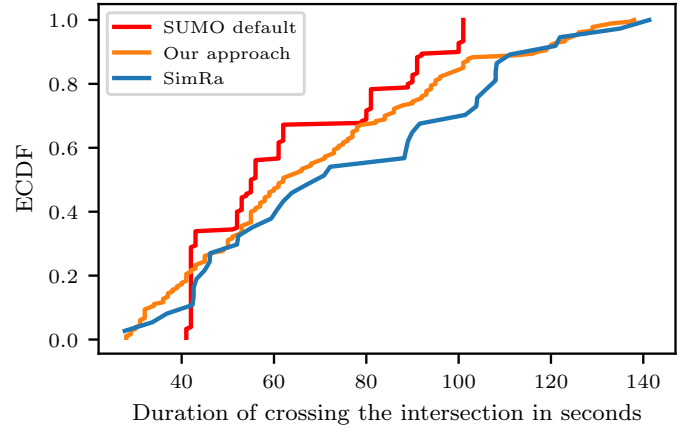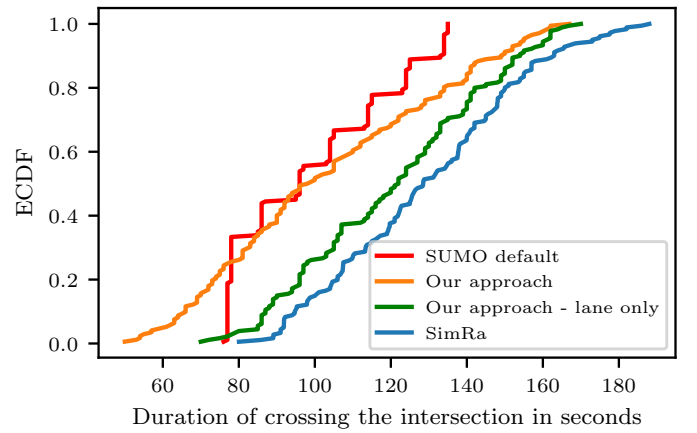
see for example Fig. 8. Especially when given the option to use the *indirect* path, cyclists take longer to cross an intersection as they need to stop at an additional traffic light. This is consistent with real-world data as we find it in the SimRa dataset at multiple intersections.

Second, in some cases, we were able to improve our results by adjusting the left-turn behavior distribution following the second distribution discussed in Section IV-C. The "lane only" results in Fig. 9 were achieved by prohibiting cyclists from using the *direct* path. Obviously, it takes much longer for cyclists to cross the intersection than SUMO's default simulation model suggests. When examining SimRa trajectories at this particular intersection, almost all cyclists chose the *indirect* path as the infrastructure guides cyclists to do so. Hence, manually adjusting the left-turn behavior distribution for such intersections is crucial.

Third, at a few intersections, the results of our approach do not yet sufficiently reflect real-world bicycle behavior (see
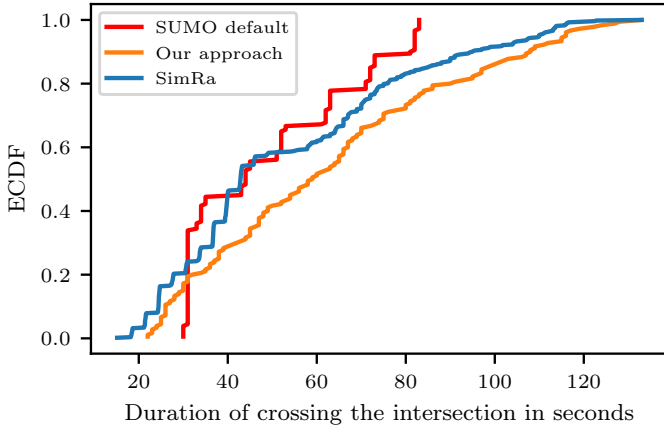
Figure 10. ECDFs of the measured durations for crossing the scenario *Alexanderstraße*. Here, our new approach is not more precise than SUMO default behavior.

Figure 11. Intersection Mehringdamm/Gneisenaustraße: a traffic island obstructs the direct turn path (dashed line) and, thus, makes the indirect path (solid line) more likely to be used.

Fig. 10). We discuss possible reasons for this in Section VII.

VII. DISCUSSION

Overall, the results presented in this paper show a significant improvement over the state-of-the-art. Nevertheless, they still have a number of shortcomings. In this section, we discuss the inherent limitations of our approach in general (Section VII-A) as well as problems resulting from the SimRa dataset as our ground truth data (Section VII-B).

A. Methodological Challenges

Our initial assumption was that the behavior of cyclists in a single intersection cannot be generalized to all intersections [7] but that the average behavior across a large number of intersections will be close enough to cyclists' behavior at arbitrary intersections. This seems to be true only for a (relatively large) subset of intersections – apparently, the intersection behavior of cyclists is more heterogeneous than expected. We believe that this is due to the fact that we averaged across all intersections in our dataset whereas there are apparently different classes of intersections that we did not account for.

Primarily, the intersection design is likely to have a strong impact: Consider the example in Fig. 11 where a traffic island partially blocks direct left turns and where markings on the ground suggest indirect turns. As another example, the intersection Bismarckstraße/Leibnizstraße had no direct left turns in the SimRa dataset. In this intersection, the reason would be that cyclists legally have to use a bike lane. When using that bike lane, a direct left turn would require cyclists to first pass through a row of parked cars, then to cross four car lanes of a major street before being able to turn left.

Aside from that, other possible influence factors include the amount and velocity of traffic (higher numbers of cars or faster cars can be expected to lead to more indirect turns), gender and age group distributions of cyclists in the respective intersections, as well as weather and light conditions or the grade of the street. In future work, we plan to explore these possible influence factors, focusing on the intersection design which we deem to have the strongest impact.

Another problem results from inaccuracy in the dataset used: GPS and motion sensors of smartphones provide only imprecise insights into actual "micro"-behavior of cyclists. Using a broad group of cyclists as input will always result in this limitation which we tried to overcome based on preprocessing and filtering of the SimRa dataset. Alternatives would be additional sensors (especially cameras) on bicycles or on intersections as in [7]. These, however, have the inherent limitation that they will either limit the number of bicycles producing data or the number of intersections covered.

B. Dataset Choice as Ground Truth

In this paper, we used the SimRa datasets as input for our analysis as it is, to the best of our knowledge, the first public dataset comprising a large number of rides that actually publishes individual rides in an anonymized but non-aggregated form. We need to keep in mind, however, that SimRa was designed for a different purpose: For example, the SimRa app records motion sensors at 50Hz but only persists every fifth value of a moving average over 30 values. While this suffices for detecting near miss incidents [2], it further limits the resolution of motion data (and thus any conclusions we can draw from that). Furthermore, the SimRa data which we used were recorded over a period of 1.5 years. During such as long period of time, physical changes to the bicycle infrastructure (both temporary and permanent) will occur, thus, adding additional noise to the data.

Finally, SimRa relies on crowdsourcing as a data collection method which often leads to participation inequality. As a result, individual users will be overrepresented in some intersections and street segments and not represented in others. Furthermore, based on the data collection method using smartphones, the user group of SimRa is likely to have a slight gender bias towards males and an age group bias towards cyclists between the ages 20 and 50. These biases will, of course, be reflected in our analysis results and cannot be compensated unless other cycling datasets become available in non-aggregated form.

C. Generalizability

Although the SimRa dataset contains rides from almost 100 regions, we only considered rides from Berlin for our data analysis and the development of cyclists' left turn behavior. The main reason for this is that almost half the rides are from the Berlin region [31]. To derive both turn model and acceleration behavior, however, we need data from "dense" street segments with many rides which are hence only available for Berlin at the moment. This is further aggravated by the fact that the preprocessing step (see Section IV) further reduces the number of eligible rides and with that the number of intersections with sufficient left-turn maneuvers. Nevertheless, this does not pose a threat to the generalizability of our findings – at least for Germany – as traffic infrastructure guidelines throughout Germany are standardized. Furthermore, adjacent countries such as Austria often also have comparable infrastructure. We hence believe that our findings also apply to at least these countries.

VIII. Conclusion

Increasing the modal share of cyclists to provide health benefits, alleviate traffic congestion, and reduce air pollution requires significant planning efforts of city planners and traffic engineers towards an improved cycling infrastructure. For this, city planners often rely on the open source simulation platform SUMO to study the effects of infrastructure changes before implementing them on the streets. Likewise, research on V2X-based safety systems for cyclists often relies on SUMO for evaluation. Unfortunately, SUMO cyclists are either modeled as slow cars or as fast pedestrians, neither of which is overly realistic.

In this paper, we used the recently published SimRa dataset, which to our knowledge is the first public dataset providing detailed insights into a large number of individual cyclists' rides, to improve SUMO's cyclist model. For this, we derived acceleration and velocity behavior and reparameterized the SUMO cyclist model. As a SUMO extension, we also developed a new intersection model describing left-turn behavior of cyclists in four-way intersections. While our work significantly improved the existing cyclist model, it is not as realistic as we wanted it to be. We, hence, discussed a number of research directions which we plan to explore in the near future.

References

[1] T. Götschi, J. Garrard, and B. Giles-Corti, "Cycling as a part of daily life: a review of health perspectives," *Taylor & Francis Transport Reviews*, vol. 36, no. 1, pp. 45–71, 2016.

[2] A.-S. Karakaya, J. Hasenburg, and D. Bermbach, "SimRa: Using crowdsourcing to identify near miss hotspots in bicycle traffic," *Elsevier Pervasive and Mobile Computing*, vol. 67, p. 101 197, Sep. 2020.

[3] F. E. Pedroso, F. Angriman, A. L. Bellows, and K. Taylor, "Bicycle Use and Cyclist Safety Following Boston's Bicycle Infrastructure Expansion, 2009–2012," *American Journal of Public Health (AJPH)*, vol. 106, no. 12, pp. 2171–2177, Dec. 2016.

[4] R. Aldred and A. Goodman, "Predictors of the frequency and subjective experience of cycling near misses: Findings from the first two years of the UK Near Miss Project," *Elsevier Accident Analysis & Prevention*, vol. 110, pp. 161–170, Jan. 2018.

[5] M.-C. H. Oczko, L. Stratmann, M. Franke, J. Heinovski, D. S. Buse, F. Klingler, and F. Dressler, "Integrating Haptic Signals with V2X-based Safety Systems for Vulnerable Road Users," in *IEEE International Conference on Computing, Networking and Communications (ICNC 2020)*, Kailua, HI: IEEE, Feb. 2020, pp. 692–697.

[6] P. A. Lopez, E. Wiessner, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flotterod, R. Hilbrich, L. Lucken, J. Rummel, and P. Wagner, "Microscopic Traffic Simulation using SUMO," in *21st IEEE International Conference on Intelligent Transportation Systems (ITSC 2018)*, Maui, HI: IEEE, Nov. 2018, pp. 2575–2582.

[7] H. Kaths and G. Grigoropoulos, "Integration of an External Bicycle Model in SUMO," in *SUMO User Conference 2016 (SUMO 2016)*, vol. 30, Berlin, Germany: DLR, May 2016.

[8] G. Grigoropoulos, L. Lücken, J. Erdmann, and H. Kaths, "Modelling Bicycle Infrastructure in SUMO," in *SUMO User Conference 2019 (SUMO 2019)*, M. Weber, L. Bieker-Walz, R. Hilbrich, and M. Behrisch, Eds., Berlin, Germany: EasyChair, Aug. 2019, pp. 187–198.

[9] J. Heinovski, L. Stratmann, D. S. Buse, F. Klingler, M. Franke, M.-C. H. Oczko, C. Sommer, I. Scharlau, and F. Dressler, "Modeling Cycling Behavior to Improve Bicyclists' Safety at Intersections – A Networking Perspective," in *20th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2019)*, Washington, D.C.: IEEE, Jun. 2019.

[10] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic Dynamics: Studies in Car Following," *INFORMS Operations Research*, vol. 6, no. 2, pp. 165–184, Apr. 1958.

[11] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear Follow-The-Leader Models of Traffic Flow," *INFORMS Operations Research*, vol. 9, no. 4, pp. 545–567, Aug. 1961.

[12] P. G. Gipps, "A Behavioural Car-Following Model for Computer Simulation," *Elsevier Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, Apr. 1981.

[13] W. Leutzbach and R. Wiedemann, "Development and applications of traffic simulation models at the Karlsruhe Institut für Verkehrswesen," *Traffic engineering & control (TEC)*, vol. 27, no. 5, pp. 270–278, 1986.

[14] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *APS Physical Review E (PRE)*, vol. 51, no. 2, pp. 1035–1042, Feb. 1995.

[15] S. Krauß, "Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics," PhD Thesis, Mathematical Institute, Köln, Germany, Apr. 1998.

[16] M. Treiber, A. Hennecke, and D. Helbing, "Congested Traffic States in Empirical Observations and Microscopic Simulations," *APS Physical Review E (PRE)*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.

[17] D. Salles, S. Kaufmann, and H.-C. Reuss, "Extending the Intelligent Driver Model in SUMO and Verifying the Drive Off Trajectories with Aerial Measurements," in *SUMO User Conference 2020 (SUMO 2020)*, Virtual Conference, Oct. 2020.

[18] J. Beilharz, P. Wiesner, A. Boockmeyer, F. Brokhausen, I. Behnke, R. Schmid, L. Pirl, and L. Thamsen, "Towards a Staging Environment for the Internet of Things," in *19th IEEE International Conference on Pervasive Computing and Communications (PerCom 2021), PerCom Work in Progress on Pervasive Computing and Communications (PerCom Workshops)*, Virtual Conference: IEEE, Mar. 2021, pp. 312–315.

[19] M. Temmen, J. Eckart, J. Merk, A.-S. Karakaya, and D. Bermbach, "Crowdsourcing in der Radverkehrsforschung," *Internationales Verkehrswesen*, vol. 74, no. 1, pp. 66–69, Jan. 2022.

[20] G. Grigoropoulos, A. Leonhardt, H. Kaths, M. Junghans, H. Baier, and F. Busch, "Traffic flow at signalized intersections with large volumes of bicycle traffic," *Elsevier Transportation Research Part A: Policy and Practice*, vol. 155, pp. 464–483, Jan. 2022.

[21] H. Twaddle and G. Grigoropoulos, "Modeling the Speed, Acceleration, and Deceleration of Bicyclists for Microscopic Traffic Simulation," *Transportation Research Record (TRR)*, vol. 2587, no. 1, pp. 8–16, Jan. 2016.

[22] D. Bermbach, "SimRa Rides Berlin 06/19-12/20," 2021.

[23] D. Bermbach and A.-S. Karakaya, "SimRa Rides Berlin 01/21-09/21," 2021.

[24] A. Chowdhury, T. Chakravarty, and P. Balamuralidhar, "Estimating true speed of moving vehicle using smartphone-based GPS measurement," in *IEEE International Conference on Systems, Man and Cybernetics (SMC 2014)*, San Diego, CA: IEEE, Oct. 2014.

[25] Y. Usami, K. Ishikawa, T. Takayama, M. Yanagisawa, and N. Togawa, "Bicycle Behavior Recognition using Sensors Equipped with Smart-

phone," in *8th IEEE International Conference on Consumer Electronics (ICCE 2018)*, Berlin, Germany: IEEE, Sep. 2018.

[26] A. Basiri, M. Haklay, and Z. Gardner, "The Impact of Biases in the Crowdsourced Trajectories on the Output of Data Mining Processes," in *21th AGILE International Conference on Geographic Information Science (AGILE 2018), VGI-ALIVE - AnaLysis, Integration, Vision, Engagement*, Lund, Sweden, Jun. 2018.

[27] X. Ma and D. Luo, "Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data," *Elsevier Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 40, pp. 130–144, Jul. 2016.

[28] I. W. Burr, "Cumulative Frequency Functions," *The Annals of Mathematical Statistics*, vol. 13, no. 2, pp. 215–232, Jun. 1942.

[29] N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," *MAA Biometrika*, vol. 36, no. 1/2, p. 149, Jun. 1949.

[30] N. Etemadi, "An elementary proof of the strong law of large numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 55, no. 1, pp. 119–122, Feb. 1981.

[31] A.-S. Karakaya, T. Ritter, F. Biessmann, and D. Bermbach, *CycleSense: Detecting Near Miss Incidents in Bicycle Traffic from Mobile Motion Sensors*, 2022. [Online]. Available: https://arxiv.org/abs/2204.10416.