Generation of Concrete Parameters from Logical Urban Driving Scenarios Based on Hybrid Graphs

Christoph Glasmacher[a], Hendrik Weber[b], Michael Schuldes[c], Nicolas Wagener[d] and Lutz Eckstein

Institute for Automotive Engineering, RWTH Aachen University, Aachen, Germany

{*christoph.glasmacher, hendrik.weber, michael.schuldes, nicolas.wagener, office*}*@ika.rwth-aachen.de*

Keywords: Automated Driving, Intelligent Vehicles, Safety Assurance, Scenario Generation, Parameter Sampling, Causal Networks, Constraint Graphs.

Abstract: Safety assurance of highly automated driving functions is a major challenge in today's research and requires the development of new validation methods. Scenario-based testing is a promising approach to handle the variety of possible situations efficiently. Due to the limited availability of real-world derived scenarios, they are increasingly generated synthetically. Whereas actual approaches to generate concrete parameters are mostly either knowledge- or data-driven, we propose a methodology to combine these approaches. We model the correlation of parameters in real-world data as multivariate probability functions by using copulas. In addition, we establish modular causal and constraint relations combining Bayesian networks and constraint graphs to add semantic knowledge about parameters and their interactions. Thereby, road user behavior and physical equations are represented. The application of our generation method on urban intersections shows the capability to sample high-dimensional parameter spaces with limited input data. Hereby, it offers the opportunity to create realistic scenarios to extend the database for scenario-based assessment.

1 INTRODUCTION

The development and safety assurance of automated driving functions is one of the big challenges in automotive engineering. Whereas traditional validation methods would require billions of kilometers driving on public roads to prove the safety (Wachenfeld and Winner, 2016), simulation-based methods offer an alternative and are a focus in current research (Riedmaier et al., 2020). One promising method is the approach of scenario-based testing. Within this method, a driving function is confronted with predefined scenarios. These scenarios are either generated based on expert knowledge or real-world data (Bussler et al., 2022).

Within data-driven approaches, concrete scenarios respectively their defining parameter values are extracted from real-world data and correlations are derived. In contrast to knowledge-based scenario parametrization, the direct link to real traffic allows conclusions about the probability of occurrence of the distribution in reality (de Gelder, 2022). If the amount of recorded data is sufficiently large, mapped correlations can represent the distribution of parameter values and improve the description of parameter interactions (Lotto et al., 2022). But since especially detailed urban scenarios need a comprehensive description and therefore a relatively large parameter set, a representative fitting needs more input data than for small parameter sets according to the curse of dimensionality (Fan and Li, 2006). Since the amount of real-world data is limited, actual correlation approaches are not sufficient for detailed scenarios needed in current safety assessment (Li et al., 2022). Other challenges of those scenarios are the understandability of the scenarios and the traceability of the generation process respectively their parameter values e.g. for safety argumentation (Beringhoff et al., 2022).

We address these problems with a new scenario parameter representation and sampling method combining knowledge-based and data-driven approaches. On the one hand, we acknowledge the realistic data distribution by mapping real-world extracted parameters to a multivariate Gaussian copula. On the other

[a] https://orcid.org/0000-0003-4826-9706
[b] https://orcid.org/0000-0003-3897-791X
[c] https://orcid.org/0000-0003-2339-8157
[d] https://orcid.org/0000-0002-9086-5061

Glasmacher, C., Weber, H., Schuldes, M., Wagener, N. and Eckstein, L.
Generation of Concrete Parameters from Logical Urban Driving Scenarios Based on Hybrid Graphs.
DOI: 10.5220/0011828400003479
In *Proceedings of the 9th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2023)*, pages 215-222
ISBN: 978-989-758-652-1; ISSN: 2184-495X

hand, we add semantic information about relations of parameters within individual modular concepts that can be combined. Therefore, a hybrid graph structure is proposed to combine specific mathematical constraints with not further specified causal relations. This expert knowledge-driven hybrid representation is used to generate new parameter values mathematically explainable following three steps: continuous parameter values are generated from data-driven correlation, those values are corrected according to knowledge-based constraints and in the end, the likelihood of occurrence is calculated to filter unrealistic distributions using causal graphs.

In the following, we give an overview of current methods of scenario generation and graph-based methods in scenario-based safety assurance. From that on we explain our methodology of the hybrid graph-based modeling and realistic scenario parameter value generation. This method is applied to urban intersections of the inD dataset (Bock et al., 2020). Thereby, new parameter sets are generated, compared to the real-world data and further discussed.

2 RELATED WORK

2.1 Scenario Generation

Scenario generation methods to create concrete scenarios can be subdivided into two fundamental types: real-world observation-based and knowledge-based approaches (Bussler et al., 2022). Knowledge-based approaches can include ontologies to create scenes (Bagschik et al., 2018), symbolic automates (Bannour et al., 2021), or based on pure expert knowledge-based scenario creation. Those approaches have one main shortcoming: They cover the realism of parameter values with regards to existing traffic insufficiently and can not estimate the probability of occurrence (Ding et al., 2022). By design, data-driven approaches can handle the problem of representativeness easier but need real-world data. (Li et al., 2022) uses real-world data to extract features and to map those to multiple agents. (Pegasus Project Consortium, 2019) extract individual parameter distributions from recorded data to generate new concrete scenarios. A similar approach to generate realistic scenarios using probability density functions (pdf) is proposed by (de Gelder et al., 2022). More recently, (Lotto et al., 2022) uses copulas to model parameters not independently but considering the correlations to create new scenario parameter values as points in a multidimensional and intercorrelated parameter space. Those approaches face one or multiple of the following limitations:

- Generated parameter values are not proven to be realistic and to potentially occur in reality (Li et al., 2022)
- Relations between individual scenario parameters are not considered explicitly. This results in wrong estimation of the possible parameter space potentially leading to the generation of unrealistic parameter values (Pegasus Project Consortium, 2019), (de Gelder et al., 2022).
- Whereas (Lotto et al., 2022) considers the correlations between scenario parameters, the generation of detailed concrete scenarios needs a significantly amount of data.
- Due to machine learning steps, the parameter value generation process cannot be traced mathematically completely (Li et al., 2022). This leads to uncertainties regarding completeness.

2.2 Graph-Based Representation

Graphs are recently used tools to describe scenarios. Within those graphs, states and their relations can be modeled (Bannour et al., 2021), (Bagschik, 2022). Especially, probabilistic graph models (PGMs) such as Bayesian networks (BN) are increasingly used to describe relations (Ding et al., 2022). A BN is a directed acyclic graph (DAG) consisting of nodes and edges. Nodes can be understood as a set of states with certain probabilities. Directed edges connect these nodes and describe their relations by a conditional probability. Graphs are used to describe scenes (Bagschik et al., 2018) or abstract interrelationships (Beck et al., 2022). (Adee et al., 2021) model perception phenomena within causal graphs. Thereby, only causal relations are represented. Combining causal graphs with BN leads to a causal BN. It differs from BN in that all relations between nodes within a BN are established because of direct causes (Pearl and Russell, 2000). This allows a better description and analysis of causal effects and relations.

Another type of graphs are constraint graphs (Friedman and Phan, 2017). Whereas PGMs consist of nodes and directed edges, bipartite graphs include nodes, undirected edges, and additional constraints representing mathematical equations. To the best of our knowledge, it is not yet used in automated driving, but a common tool in other domains. (Zhu et al., 2021) uses constraint graphs for language sorting algorithms. (Para et al., 2021) applies these to generate room layouts and (Friedman and Phan, 2017) states that the theory is capable to model complex problems.

3 METHODOLOGY

In order to generate parameter values for detailed urban driving scenarios we propose a combined methodology using data-driven and knowledge-based approaches. We use traffic data to extract realistic parameter distributions and additionally, use semantic information to reduce the realistic parameter space. Thereby, the overall process is split into two parts: traffic analysis and scenario parameter values generation. Within the traffic analysis, we extract parameter values from real-world data according to an underlying scenario concept. Those are stored as concrete scenarios within a database (see Fig. 1). In the generation part, multiple concrete scenarios are requested from the database and linked to the hybrid graph. This graph itself is linked to the scenario concept and describe the relations between parameters. After choosing a subset of parameters those concrete values are used to fit a multivariate probability distribution. This distribution is then used to generate initial parameter values. After generation, the parameter values are corrected to meet the included knowledge of the constraint graph. Afterward, they are checked by applying the causal graph and calculating probabilities.

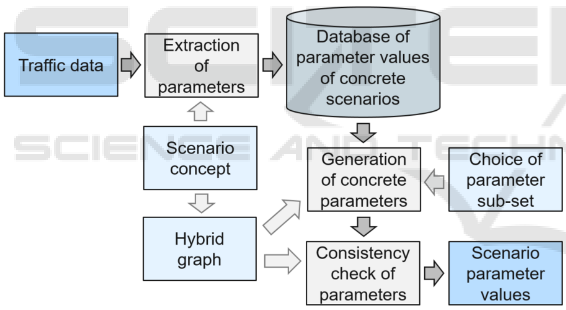


Figure 1: Scenario generation methodology including real traffic and expert knowledge.

3.1 Data-Driven Information Extraction

To ensure the generation of representative concrete scenarios, initial parameter value extraction is based on the analysis of traffic observations. The data is analyzed and transfered into parameter according to an underlying scenario concept. Because of the high complexity of urban traffic, a hierarchical approach is beneficial. Thereby, individual parameters or parameter combinations can be assigned modularly. By assigning multiple modules as maneuvers, or conflicts, the parameters can be combined within the scenario to describe scenarios in detail (see Fig. 2).

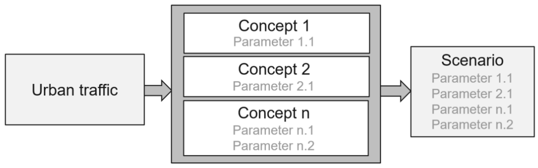


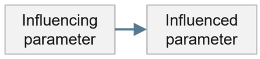
Figure 2: Composition of scenario parameters within a modular scenario concept.

3.2 Graph-Based Representation

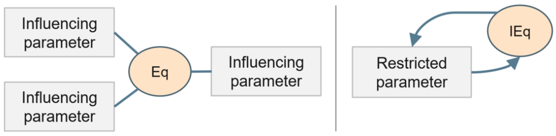
Whereas correlations on limited input data give only an incomplete picture of general relations, external expert knowledge is included via graphs. Although the usage of causal unrelated parameters may be possible for simple scenarios, it would be infeasible to avoid dependencies when parameterizing detailed and understandable scenarios. So, relations between the parameters are established to prevent inconsistencies and therefore modeled in graphs. Furthermore, adding knowledge about parameters leads to a possible reduction of the parameter space by constraints given by semantic relations a priori. For the setup of these graphs, a distinction is made between two types of edges:

- **Probabilistic Relations:** Probabilistic relations involve the causal influence of one parameter on another that cannot be adequately described deterministically with reasonable effort. Reasons for this can be the high complexity, fuzziness, or lack of better knowledge like e.g. the complexity of movement of an pedestrian (see Fig. 3a).
- **Deterministically Describable Relations:** These are relations between parameters that can be unambiguously represented by deterministic mathematical equations or inequalities. The mathematically describable relations can be derived from physical limitations or further model-based constraints. Descriptions can e.g. be used for metrics or reachability constraints for parameters (see Fig. 3b).

In order to use these two types of relations simultaneously, a hybrid graph structure based on causal and constraint relations is used. Whereas in a usual BN conditional probabilities are used for modeling correlations between two nodes, the causal relation meaningfully restricts the parameter space leading to a causal Bayesian network. Thereby, not all causal effects have to be modeled but the more accurate effects are described the higher the validity of the restriction. If not all causal influences are considered, the space is less restricted whereas if non-causal relations are modeled, the restrictions may be caused by spurious correlations and diver from what is realistic. While



(a) Causal relation



(b) Constraint relation with equations (Eq) and inequations (IEq) within constraint graph

Figure 3: Classification of semantic parameter relations in graphs.

in a causal graph the relation between two parameters consists of a directed edge, the edges within constraint graphs are undirected but linked by an equation. This property of constraint graphs is used to model mathematically known and undirected relations. By combining the two types of edges, multiple concepts can be modeled with different or redundant parameterizations.

Within the modeling of a scenario, it is thus possible to combine different concepts (see Fig. 4). In doing so, identical nodes are merged and new connections between the graphs are established. This allows it to model even high-dimensional parameter spaces combining multiple elements such as road users, weather influences or traffic signs. So, complex scenarios as well as longer scenarios or sequences can be modeled. Thereby, establishing further nodes to describe relations between road users, temporal influences and spatial dependencies between scenario aspects can be added. This leads to a more constrained parameter space and thus the needed amount of input data decrease.

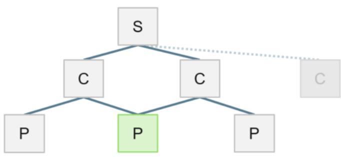


Figure 4: Combined graph with scenario (S), concepts (C), and their parameters (P) for generation of detailed scenarios.

3.3 Scenario Specification

Based on the previous steps, extracted data and applied knowledge have to be chosen to generate new parameter values. Thereby, the amount of concrete scenarios derived from the database can be filtered based on one or multiple concept annotations (cf. Sec. 3.1). The more restrictions are applied and the fewer parameters are used, the more focused further parameter value generation can be. The choice of both, the used concrete scenarios, and also parameters are dependent on the use case. When selecting parameters, however, it is important to note that not just any selection of defining parameters will result in a complete scenario. Therefore, two rules for parametrization are defined:

- A graph has to be specified sufficiently. It is considered as specified sufficiently if all direct descendants of the root are specified sufficiently.
- A node is considered specified sufficiently if either concrete parameter values for the node are set or descendants exist and are sufficiently specified.

The scenario parameterization rules lead to two consequences: Not every set of parameters selected results necessarily in an executable scenario. Not specifying all descendants can lead to semi-concrete scenarios. This can especially help clustering scenarios within a broader context like a safety argumentation. The fewer leaf-near nodes are specified, the less concrete the scenario becomes. The more nodes are described on a concrete level, the more likely the possibility to execute the parameter values as a concrete scenario.

The usage of a sub-set of parameters can be seen as a graph simplification deleting unused nodes and reconnecting relations. This may reduce the graph to a pure causal BN or a constraint graph. The simplification to a causal BN allows e.g. a more comprehensive analysis of dependencies using causal theory (Pearl and Russell, 2000) whereas a analytical solution for the parameter space can be found within pure constraint graphs (Friedman and Phan, 2017).

3.4 Parameter Generation

The generation of the parameter values can be subdivided into three steps taking the benefits of knowledge-based and data-driven generation methods into account. Firstly, in a data-driven step initial parameter values are created based on multivariate distributions of the extracted concrete scenarios. In the second step the generated parameter values are corrected applying the included knowledge before a hybrid approach is used in the end to validate the parameter values based on probability calculations:

1. **Initial Generation of Parameter Values:** To create initial parameter values, the extracted and chosen traffic information (cf. Sec. 3.3) is transformed into a multivariate probability distribution. Since the distribution on correlations of the individual parameters does not follow an a priori known distribution function, the parameter space is represented by a Gaussian copula. This repre-

sentation is also acceptable for more complex distributions since the data is transformed when fitting the copula. On the basis of this probability distribution, individual parameter values can be determined efficiently on the basis of traffic data.

2. **Constraint Verification and Correction:** Due to the mostly limited amount of data, the approximated distribution of the copula can not reflect causal relations or physical limits sufficiently. These inaccuracies can lead to contradictions within generated parameter values that must be corrected. These can include exact mathematical relations that are not considered in the correlation itself but are fulfilled by the input data. To correct those logical inconsistencies, graph constraints are reviewed and adjustments are made to satisfy all of those.

3. **Causal Probability Calculation:** Finally, the probability of occurrence of the parameter distribution is evaluated with respect to the real values. Contrary to the copula, only the causal relations of the hybrid graph are modeled within a causal Bayesian network. Thus, the possible parameter space is restricted by known causal relations. For the conditional occurrence probability of a parameter value, the traffic data is discretized and the conditional probabilities of parameter values (i) based on descendants (i_{pre}) are calculated ($p(i|i_{pre})$). The probability of occurrence of the complete parameter values is thus given by the product of the conditional probabilities (1). As long as the probability of a parameter values set (p_{set}) is greater than zero, a similar parameter distribution among the network nodes can be found in the data. If the probability is smaller than the desired threshold, the parameter values can be varied and must then be verified again according to step two.

$$p_{set} = \prod_i p(i|i_{pre}) \tag{1}$$

Using the hybrid graphs, other concretization steps are also conceivable but limit the design. For example, instead of the copula, an iterative set of valid scenarios constraining the parameter boundaries can be found, or in the case of a linear constraint graph, the problem can be solved analytically. However, this would limit the description and parameter value generation significantly.

4 RESULTS

To verify the proposed method, it is implemented in Python and pgmpy is used for graph representation (Ankan and Panda, 2015). For evaluation, parameter values for urban scenarios are generated based on observed real-world data from the inD dataset (Bock et al., 2020). Thereby, a scenario with two road users approaching the intersection from different directions, merging on the intersection, and ending on the same road is chosen exemplary. Maneuvers of the road users as well as relative direction and approaching arms are not further specified a priori, but implicitly included within the extracted parameters. The observed intersection (*Frankenburg*) contains four arms. 322 scenarios were extracted from this intersection. Besides the data-driven extraction, the conflict is modeled within the hybrid graph structure (see Fig. 5) according to (cf. Sec. 3.2). 23 of those parameters are directly extracted from the data. Due to the uniform road geometry intersection, related attributes are considered constant and therefore underlying parameters are combined to two nodes. Additionally, 2 parameters are unobserved because similar to the road geometry parameters they are to complex to process without further breakdown and only serve for a better scenario understanding (cf. Sec. 3.3). In addition to the causal relations, independent equations and inequations are used to describe mathematical relations. Furthermore, overarching conditions (gray) are inserted for parameter extraction and validation to check the correct expression of the concrete scenario. According to Sec. 3.3 it would be also valid to use less parameters, but all are used to show the methodology at a relatively simple use case.

From these input data, 200 parameter sets are created to describe the scenario according to Sec. 3.4 using 1000 randomly selected copula samples. Constraints and causal relations are used to adjust and filter the data. Thereby, the parameter values converge towards real-world extracted values as exemplary shown for the constellation of two vehicles. This is partially characterized by the predicted order of passing (predicted priority level adapted from (Hu and Li, 2017)) and the predicted timegap at the conflict zone (see Fig. 6). On average 50 percent of them meet all 4 additional overarching conditions. The incorrect 50 percent result from the inaccurate fitting of the copula since these semantic relations were not introduced. Nevertheless, it can be shown that the distribution of the generated data after the following steps including the hybrid graph have similar statistical properties in spite of the high dimensionality and complex distribution. Therefore, the relation of two constraint adjusted metrics and two unconstrained variables are shown exemplary (see Fig. 7).

For a quantitative comparison, the distance be-

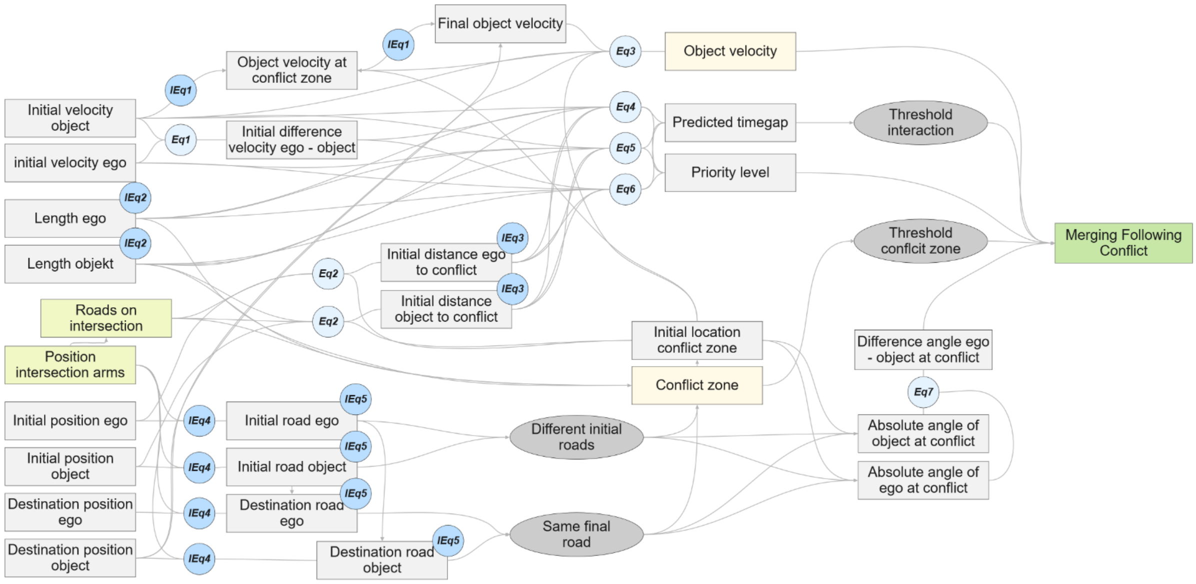
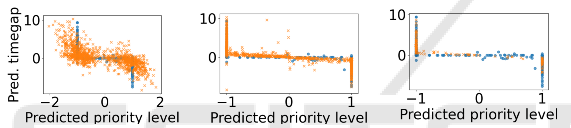
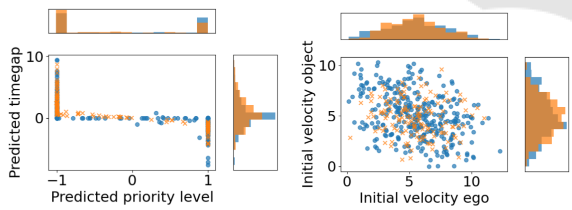


Figure 5: Hybrid graph of a merging following conflict including observed parameters (yellow), unobserved parameters (light yellow), infrastructure parameters (green), equations (Eq), inequalities (IEq) and thresholds (gray).



(a) Copula sampling (b) Constraints (c) Causal behavior

Figure 6: Parameter distribution of real parameter values (blue) and generated parameter values (red) along the generation steps each including previous steps (cf. Sec. 3.4).



(a) Constraint metrics (b) Unconstraint velocities

Figure 7: Distribution of extracted parameter values (blue) compared to generated parameter values (red).

tween two parameters is investigated. Comparing the parameters individually, the average and normalized Euclidean distance to the next observed real-world parameter value is used (see Fig. 8). The metric shows that the distance distribution of the generated values is similar to those of the real parameters. The generated parameter values are particularly close to real values for Gaussian distributions of real parameters and parameters constrained by rules. Deviations occur when the distributions consist of clusters with relatively sharp borders as in parameters as object length for pedestrians, bicycles, and vehicles. Since the borders are not modeled within constraints, a dispersion occurs because of the Gaussian modeling. Similar effects can be observed in the disappearance of road users.

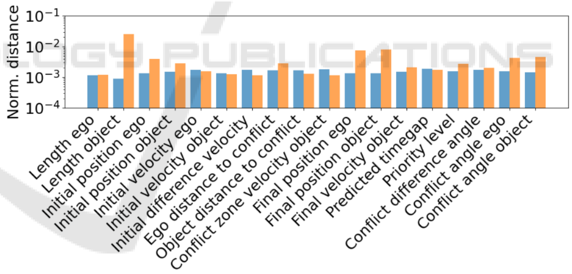


Figure 8: Average normalized distances between values for individual parameters.

To extend the evaluation to the multidimensional parameter space, it is necessary to bring the parameters into relations. For this purpose, the parameter values are normalized. The average Euclidean distance to the closest generation input parameter sets is used as a reference. The number of reference parameter values used for this purpose corresponds to the number of individual parameters used.

The average local minimum neighbor distance of a generated parameter values is 30 percent higher compared to the real-world and 17 percent higher compared to newly inserted real-world parameter values (see Tab. 1). Similarly, minimum and maximum distances are slightly higher, so it can be seen, that the parameter space is slightly extended. When using more parameter values as an input for the generation, both, the distances of new real-world and generated parameter value distances decrease.

Table 1: Deviation of locally average distances of new parameter values compared to generation input real-world parameter values.

Data type	Avg	Min	Max
real-world	1.13	0.79	1.63
synthetic	1.32	0.93	1.86

5 DISCUSSION AND OUTLOOK

The comparison of the generated parameter values to the observed parameters shows that the method is suitable for generating realistic concrete scenarios even with few input data and high dimensionality. This is shown by the fact that the generated parameter values reflect the distribution of the extracted parameters both qualitatively and quantitatively without the specification of a distribution structure. Thereby, the representation of discrete and continuous values is possible. According to the mean local distance to extracted parameter values, it can be shown that the generated data are similar to the original data and fill in gaps in the parameter space while satisfying the constraints of the model. Furthermore, it can also be seen that some parameter values are generated which have a relatively high distance to real points with regard to the local distances. This can also be observed adding new real-world values but is slightly higher for generated values. This is due to the moderate fit and partly close real points, whose restrictions were not modeled. Especially for expected distributions consisting of discrete and continuous values, further restrictions via constraints or more accurate relations between parameters should be used for closer results. Additionally, it is shown that single process steps like the exclusive use of a copula is not sufficient for the combination of many parameters and limited available data, because dependencies are not modeled and therefore have to be filtered out. Although the dependencies implicitly find their way into the copula through the real-world data, significantly more data or a reduction of the parameter space would be necessary for an accurate representation. This effect is expected to be magnified for high dimensional problems but can be counteracted by more detailed modeling via the hybrid graph structure.

Besides the quality of the generated data, additional advantages of the hybrid structure can be shown. The approach results in an improved understandability of the causal relations among each other due to the graphical representation. In addition, the structure yields simplified safety argumentability through the usage of semi-concrete scenarios and the comprehensive coverage of the parameter space. Furthermore, the methodology allows mapping of redundancies and alternative parameterization possibilities, which is especially interesting for the use in databases and simplifies the modular combination of different concepts for the creation of detailed scenario description.

In future work, the modular linking of hybrid graphs will be further investigated in order to address the complexity of urban traffic. In this context, more constraints should be established to link graphs sequentially and temporally in parallel, thus enabling the generation of realistic multi-vehicle parameter values. Besides physical or mathematical constraints, traffic regulations could also be investigated. Furthermore, the evaluation of the distance computation of scenarios results in another open field that can lead to a probability computation based on hybrid graphs (cf. Sec. 3.4) as well as to further similarity analyses. In addition, the introduced graphs can be a suitable to increasingly couple scenario parameter value generation with cause-effect relationships to analyze and understand them within causal theory.

6 CONCLUSIONS

In this work, we presented a comprehensive method for generating parameter values for detailed concrete scenarios. Starting from traffic data, we extracted concrete scenarios and semantically linked their parameters via hybrid graphs. Thus, causal relations as well as mathematical constraints could be modeled. Furthermore, we have shown that realistic and high-dimensional parameter values can be generated using distributions of few concrete input data in combination with the proposed graph structure. Those parameter values of concrete scenarios are located near real-world data and fill the parameter space between observed parameters. Further evaluation of the representativeness as well as the optimized generation of multi-vehicle scenarios remains for future work. Additionally, thresholds and probability analysis of scenarios within the distributions shall be further investigated.

ACKNOWLEDGEMENTS

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project "Verifikations- und Validierungsmethoden automatisierter Fahrzeuge im urbanen Umfeld". The authors would like to thank the consortium for the successful cooperation.

REFERENCES

Adee, A., Gansch, R., and Liggesmeyer, P. (2021). Systematic modeling approach for environmental perception limitations in automated driving. In *2021 17th European Dependable Computing Conference (EDCC)*, pages 103–110. IEEE.

Ankan, A. and Panda, A. (2015). pgmpy: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*. Citeseer.

Bagschik, G. (2022). *Systematischer Einsatz von Szenarien für die Absicherung automatisierter Fahrzeuge am Beispiel deutscher Autobahnen*. Universitätsbibliothek Braunschweig.

Bagschik, G., Menzel, T., and Maurer, M. (2018). Ontology based scene creation for the development of automated vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1813–1820. IEEE.

Bannour, B., Niol, J., and Crisafulli, P. (2021). Symbolic model-based design and generation of logical scenarios for autonomous vehicles validation. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 215–222. IEEE.

Beck, H. N., Salem, N. F., Haber, V., Rauschenbach, M., and Reich, J. (2022). Phenomenon-signal model: Formalisation, graph and application.

Beringhoff, F., Greenyer, J., Roesener, C., and Tichy, M. (2022). Thirty-one challenges in testing automated vehicles: Interviews with experts from industry and research. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 360–366. IEEE.

Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., and Eckstein, L. (2020). The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1929–1934. IEEE.

Bussler, A., Wagener, N., Pütz, A., Menzel, T., Braun, T., Fuchs, J., Körtke, F., Weber, H., and Glasmacher, C. (2022). A story of scenarios - from data and knowledge to scenarios for testing.

de Gelder, E. (2022). *Safety assessment of automated vehicles using real-world driving scenarios*. PhD thesis, Delft University of Technology.

de Gelder, E., Hof, J., Cator, E., Paardekooper, J.-P., den Camp, O. O., Ploeg, J., and de Schutter, B. (2022). Scenario parameter generation method and scenario representativeness metric for scenario-based assessment of automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):18794–18807.

Ding, W., Xu, C., Arief, M., Lin, H., Li, B., and Zhao, D. (2022). A survey on safety-critical driving scenario generation – a methodological perspective.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery.

Friedman, G. J. and Phan, P. (2017). *Constraint Theory*, volume 23. Springer International Publishing, Cham.

Hu, M. and Li, Y. (2017). Drivers' avoidance patterns in near-collision intersection conflicts. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE.

Li, A., Chen, S., Sun, L., Zheng, N., Tomizuka, M., and Zhan, W. (2022). Scegene: Bio-inspired traffic scenario generation for autonomous driving testing. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):14859–14874.

Lotto, K., Nagler, T., and Radic, M. (2022). Modeling stochastic data using copulas for application in validation of autonomous driving.

Para, W., Guerrero, P., Kelly, T., Guibas, L., and Wonka, P. (2021). Generative layout modeling using constraint graphs. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.

Pearl, J. and Russell, S. (2000). *Bayesian networks*. UCLA, Department of Statistics Papers.

Pegasus Project Consortium (2019). Pegasus method - an overview.

Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., and Diermeyer, F. (2020). Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, 8:87456–87477.

Wachenfeld, W. and Winner, H. (2016). The release of autonomous vehicles. In Maurer, M., Gerdes, J. C., Lenz, B., and Winner, H., editors, *Autonomous Driving*, pages 425–449. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zhu, Y., Zhou, K., Nie, J.-Y., Liu, S., and Dou, Z. (2021). Neural sentence ordering based on constraint graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14656–14664.