GENERATING REPRESENTATIVE TEST SCENARIOS: THE FUSE FOR REPRESENTATIVITY (FUSE4REP) PROCESS MODEL FOR COLLECTING AND ANALYSING TRAFFIC OBSERVATION DATA

Maximilian Bäumler
Matthias Lehmann
Günther Prokop,
Chair of Automobile Engineering,
Technische Universität Dresden
Germany

Paper Number 23-0122

ABSTRACT

Scenario-based testing is a pillar of assessing the effectiveness of automated driving systems (ADSs).
For data-driven scenario-based testing, representative traffic scenarios need to describe real road traffic situations in compressed form and, as such, cover normal driving along with critical and accident situations originating from different data sources. Nevertheless, in the choice of data sources, a conflict often arises between sample quality and depth of information. Police accident data (PD) covering accident situations, for example, represent a full survey and thus have high sample quality but low depth of information. However, for local video-based traffic observation (VO) data using drones and covering normal driving and critical situations, the opposite is true. Only the fusion of both sources of data using statistical matching can yield a representative, meaningful database able to generate representative test scenarios. For successful fusion, which requires as many relevant, shared features in both data sources as possible, the following question arises: How can VO data be collected by drones and analysed to create the maximum number of relevant, shared features with PD?

To answer that question, we used the Find–Unify–Synthesise–Evaluation (FUSE) for Representativity (FUSE4Rep) process model. We applied the first (“Find”) and second (“Unify”) step of this model to VO data and conducted drone-based VOs at two intersections in Dresden, Germany, to verify our results. We observed a three-way and a four-way intersection, both without traffic signals, for more than 27 h, following a fixed sample plan. To generate as many relevant information as possible, the drone pilots collected 122 variables for each observation (which we published in the ListDB Codebook) and the behavioural errors of road users, among other information. Next, we analysed the videos for traffic conflicts, which we classified according to the German accident type catalogue and matched with complementary information collected by the drone pilots. Last, we assessed the crash risk for the detected traffic conflicts using generalised extreme value (GEV) modelling. For example, accident type 211 was predicted as happening 1.3 times per year at the observed four-way intersection.

The process ultimately facilitated the preparation of VO data for fusion with PD. The orientation towards traffic conflicts, the matched behavioural errors and the estimated GEV allowed creating accident-relevant scenarios. Thus, the model applied to VO data marks an important step towards realising a representative test scenario database and, in turn, safe ADSs.

INTRODUCTION

Automated driving systems (ADSs), as an increasingly common part of road traffic today (Hohm, 2022), are designed to reduce the number of accidents and fatalities on the road and, as such, to play a significant role in making road traffic safer. To that end, ADSs first have to prove that they can drive more safely than attentive human drivers (Bergmann, 2022). One way to test ADSs for the safety of their intended functionality and thus safe driving is scenario-based testing, which entails using scenarios derived from real-world data (Nalic *et al.*, 2020), including naturalistic driving studies, police accident data (PD) and video-based traffic observations (VOs) using drones (Bock *et al.*, 2019; Nalic *et al.*, 2020). At best, real-world data sources cover all road traffic in the ADSs' operational design domain (ODD) and thus represent the ODD of road traffic (Lehmann *et al.*, 2019). Ideally, those data sources should also have the same depth of information needed to derive test scenarios.

However, the continuous collection of real-world data in all ODDs in which ADS are slated to operate is cost-intensive and technically complex. Beyond that, real-world data sources vary in the content of their information. Although PD represent entire regions or countries, they encompass information accessible only to police officers. Given that restriction, dynamic information about parties involved in accidents is not collected. By contrast, VOs afford a microscopic perspective on the dynamic behaviour of road users but are often spatially and temporally limited available.

In response to those setbacks, Bäumler and Prokop (2022) have proposed creating representative, information-rich databases for deriving test scenarios by fusing various real-world data sources. For instance, PD from Germany can be fused with data from local VOs, assuming that they belong to one unobserved, superordinate population (Bäumler *et al.*, 2020). In that case, dynamic information about road users (e.g. trajectory, speed and acceleration) can be assigned to the corresponding PD.

Given a common, unobserved, superordinate population, the quality of fusing two data sources depends primarily on the overlapping information between the sources—for instance, in the form of common variables (D'Orazio, Di Zio and Scanu, 2006). The more variables that coincide, the higher the probability of achieving good data fusion results (Rässler, 2002; D'Orazio, Di Zio and Scanu, 2006). For that reason, data collection should consider unifying information between the sources to be fused. However, regarding the fusion of, for example, German PD and VO data, changes in the nationwide standardised data collected by police are achievable only in the long term. Thus, VOs should be geared towards collecting information and/or variables comparable to PD. To that end, with reference to a real-world case, this paper answers a specific research question: How can VO data be collected by drones and analysed to create the maximum number of relevant, shared features with PD?

To answer this question, we introduce and apply the Find–Unify–Synthesise–Evaluation (FUSE) for Representativity (FUSE4Rep) process model (Bäumler and Prokop, 2022) to collect and analyse VO data for subsequent data fusion.

In what follows, we first introduce the general idea of the FUSE4Rep process model and the resulting requirements for collecting VO data using drones. Next, we demonstrate how VOs are collected using drones and a mobile app that we developed, after which we analyse the data collected and derive additional common variables concerning PD. We close the paper with a discussion of our results and directions for future research.

BACKGROUND

The FUSE4Rep model (see Figure 1), developed by Bäumler and Prokop (2022), proposes a holistic approach for fusing PD with VO data. In contrast to alternative approaches (Erbsmehl *et al.*, 2017; Erbsmehl, Lich and Mallada, 2019; Krause, 2019), the FUSE4Rep model explicitly seeks to maximise overlapping information between both sources of data to be fused and, in the process, to ensure valid fusion using statistical procedures, specifically statistical matching. As shown in Figure 1, the FUSE4Rep model starts by determining the shared, unobserved, superordinate population between two data sources as well as identifying potential common information that can be collected and/or analysed in both sources. Second, the common information identified needs to be mined and unified to be comparable. Third, both prepared data sets are synthesised using statistical matching, a process detailed by Bäumler et al. (2020). Last, data fusion is evaluated using statistical indicators and, if available, real-world data.

In light of our research question, this paper focuses on the first and second step of the FUSE4Rep model: "Find" and "Unify". Thus, possible information shared by PD and VO data is identified and subsequently unified. In the following, we use the terms *crash* and *accident* synonymously.

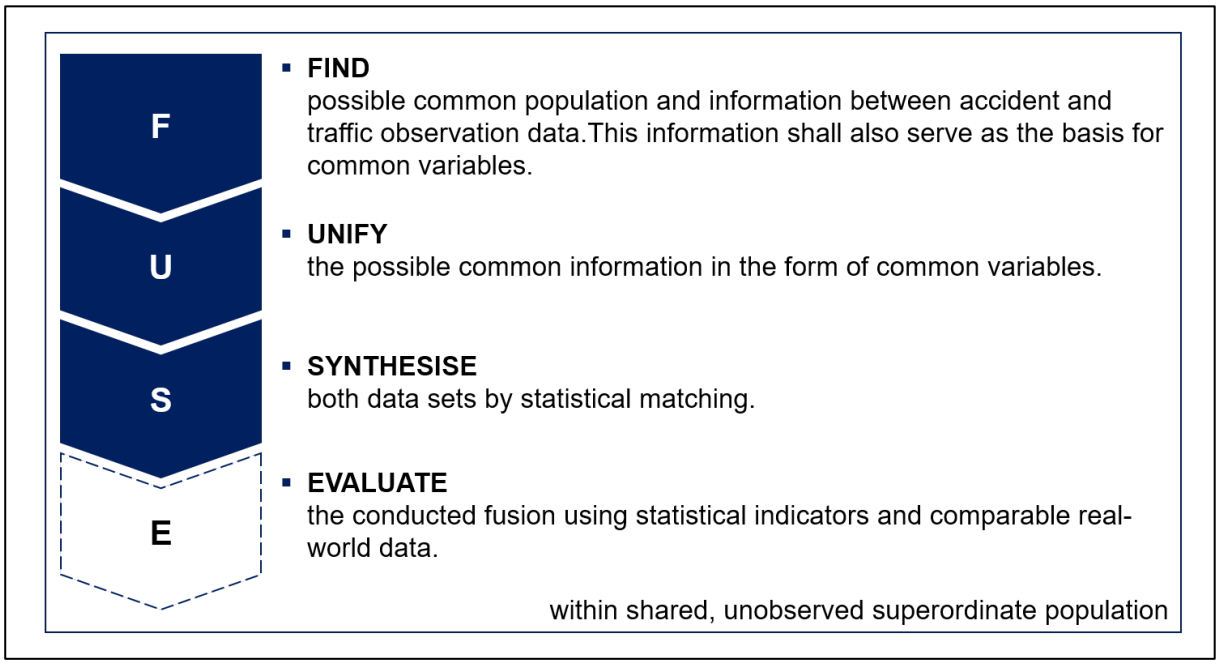


Figure 1. The four steps of the FUSE4Rep process model.

METHOD

This section describes our VOs, conducted to collect data to be fused with PD, and their subsequent analysis. Figure 2 shows the specific steps of the FUSE4Rep process model applied in this paper. In doing so, it anticipates that possible traffic conflicts, in the form of three-digit accident types (3AT), belong to the overlapping information within VOs and PD and should thus be collected and analysed.

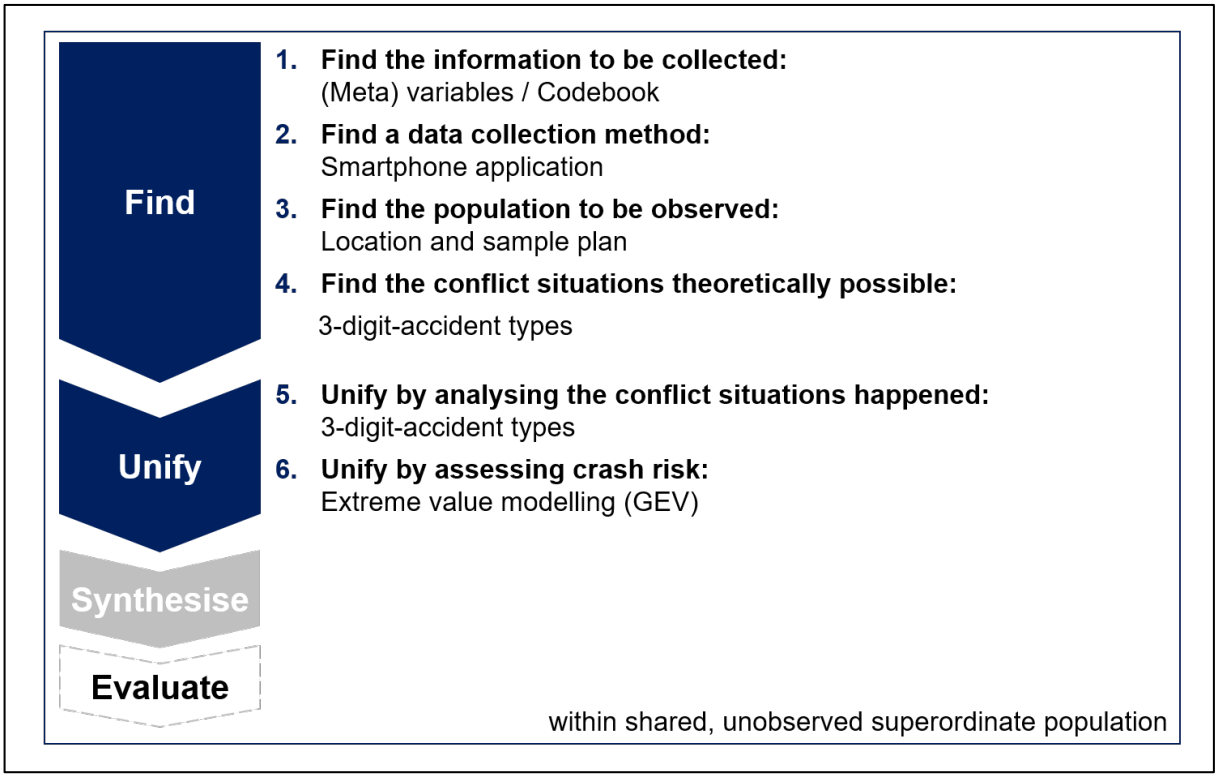


Figure 2. The FUSE4Rep process model applied to VO data.

Step 1: Find
In this subsection, we introduce all of the necessary steps for collecting VO data within the first step, "Find", of the FUSE4Rep process model (see Figure 2).

Information to be collected
To ensure that the VO data collected can be fused with PD, they need to fulfil the following four requirements:

1. Coverage of information collected by the police and thus published in national statistics (Destatis, 2021), because the PD to be fused are German;
2. Consideration of the traffic safety causality model (see Figure 3; Tarko, 2019; Orsini *et al.*, 2021), which represents the emergence of crashes;
3. Consideration of the six-layer model for scenario description (Scholtes *et al.*, 2021), because the fused data set should ultimately support the generation of test scenarios; and
4. Information collected by one drone pilot equipped with corresponding measurement instruments to keep personnel costs low.

Based on the four requirements, we created a codebook containing 122 different variables to be identified during and after a VO. The ListDB Codebook, published as part of the "Leverage Information on Street Traffic (ListDB)" project, has been made publicly available.[1]

[1] The ListDB Codebook can be accessed at https://w3id.org/listdb/.

According to the adapted causality model of traffic conflicts and crashes displayed in Figure 3, every crash is preceded by a traffic conflict. However, because not every traffic conflict necessarily leads to a crash, every accident recorded by the police is based on a traffic conflict, which in Germany can be described with the help of the 3AT classification (Ortlepp and Butterwegge, 2016; Destatis, 2021). Specifically, a traffic conflict represented as a 3AT describes the simultaneous approach of road users to a point on the road where they may collide (Ortlepp and Butterwegge, 2016). At the same time, traffic conflicts can be recorded and video-based analysed (Polders and Brijs, 2018). Therefore, a possible link between PD and VO data lies in the uniform description of traffic conflicts in both data sources according to the 3AT classification. The causality model (Figure 3, left) also shows that a set of different factors (e.g. road- and weather-related factors) can influence traffic conflicts and, in turn, crashes (Tarko, 2020). Thus, those factors should also serve as links between PD and VO. In that context, the six-layer model (6LM) for describing test scenarios for assessing ADSs already covers all factors except human factors. Because the 6LM was taken into account in the design of the ListDB Codebook, we here detail the collection and presentation of traffic conflicts in connection with human factors in the Codebook and refer to the detailed online ListDB Codebook for information from the other layers and factors.

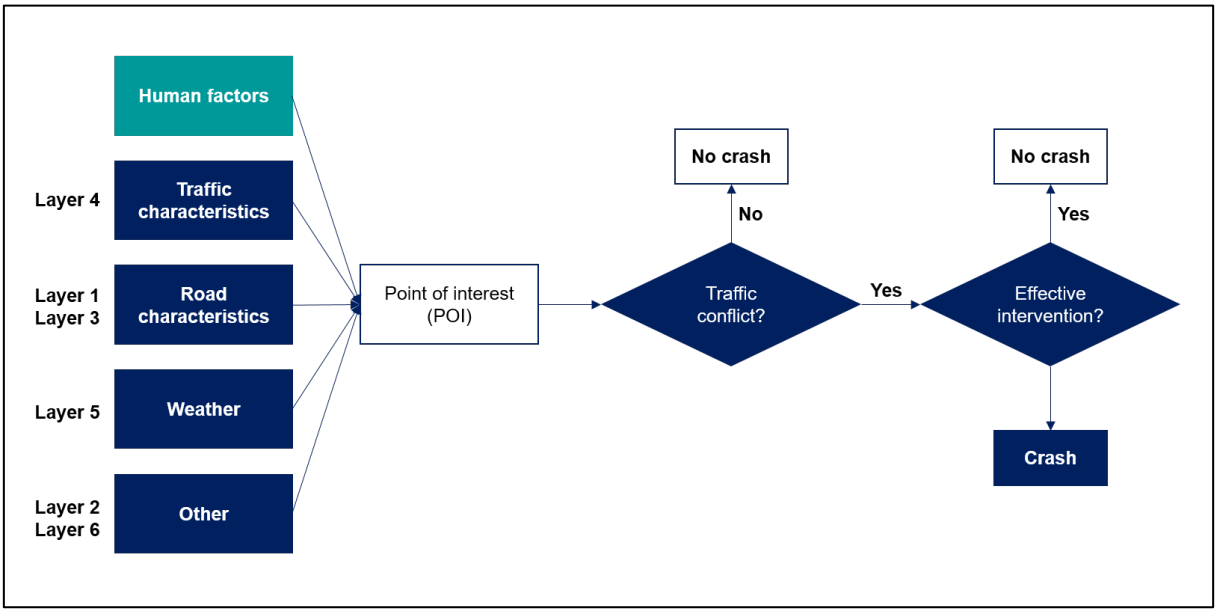


Figure 3. Causality model of traffic conflicts and crashes with assigned scenario layers, adapted from Tarko (2019) and Orsini et al. (2021).

In German PD, human factors primarily consist of drivers' behavioural errors, which can be assigned to different parties involved in accidents —namely, the primary contributor to the accident and other parties (Destatis, 2021). To each party, the police can assign up to three behavioural errors, including errors in observing the right of way, choice of speed or overtaking (Destatis, 2021). Therefore, it makes sense to adopt the police's categories of behavioural errors for VO data and assign them to observed traffic conflicts. However, the following challenges arise as a consequence:

1. **Limited detection of traffic conflicts:** Reliably detecting traffic conflicts in the recorded VO data needs to be possible. Test observations have shown that parking crashes, for example, are difficult to detect in VOs from the height at which drones fly (ca. 80 m) due to constraints in resolution and object detection. Traffic conflicts caused by unusual events, including ambulances with active blue lights, are also challenging to detect.

2. **Lack of ego-perspective:** Due to the missing ego-perspective of road users involved in traffic conflicts, the causes of behavioural errors, including visual obstacles, are challenging to detect and assess. Road users may also agree on the right of way at intersections, meaning that a seeming right-of-way error, especially when seen from the drone's bird's-eye view, may not in fact be one.

3. **Lack of personal information:** Drivers' behavioural errors, which affect drivers as people (e.g. excessive alcohol consumption and fatigue) are not detectable from the outside.

The first two challenges require a deep understanding of the traffic situation in question and its context, which cannot be achieved by analysing only the video data collected afterwards. However, the drone pilot

monitoring traffic situations during VOs can help to overcome those challenges. Therefore, we propose including the drone pilot in the detection of traffic conflicts and behavioural errors and linking it to real-time, drone-based video recording. That approach combines conflict techniques relying on human experts, so to speak, including parts of the "Swedish Traffic Conflict Technique" (Polders and Brijs, 2018), with automatic video analysis afterwards.

Nevertheless, because it is also difficult for drone pilots to recognise traffic conflicts, we introduced the concept of point of interest (POI), which generally precedes a potential traffic conflict (see Figure 3) and is easier for drone pilots to detect. As detailed in Table 1 in the Appendix, we differentiate four types of POIs:

1. **Single (1×):** One road user shows unusual behaviour or behavioural errors.
2. **Interaction (8×):** At least one road user reacts or should react to another road user. Interactions include the type of road user for interactions between a maximum of two users. Interactions with more than two users are multi-object interactions that do not specify the types of users.
3. **Predefined event (5×):** Predefined events are special and rare events (e.g. ambulances with active blue lights or slow-moving obstacles such as sweepers).
4. **Other (1×):** Everything that does not fit into the other three types is categorised as "Other".

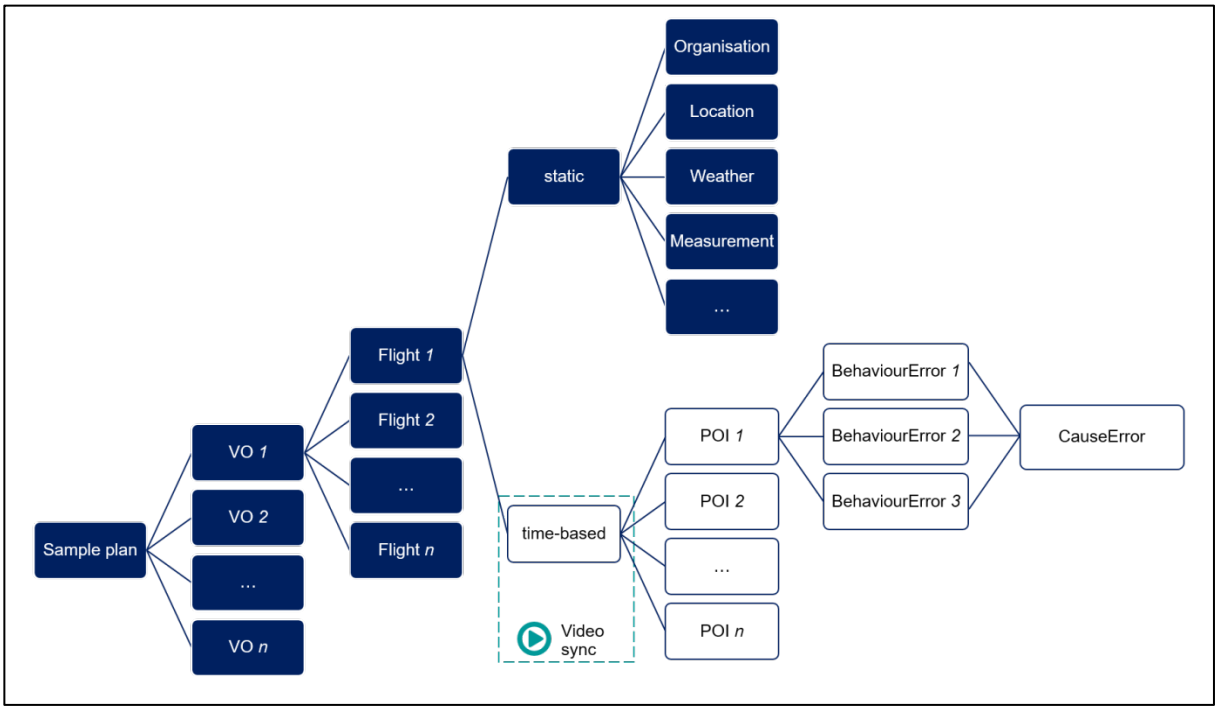


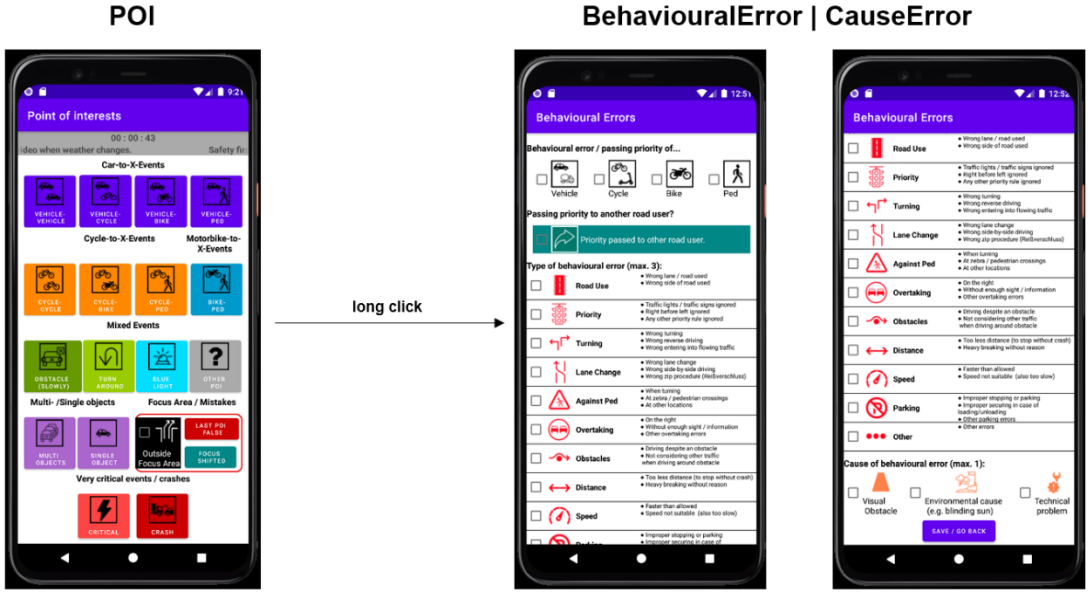
Figure 4. Relationships between sample plan, VOs, flights and identified variables.

For each POI that the drone pilot captures, three different behavioural errors can also be captured, along with one cause of the behavioural error (see Figure 4). Next, to overcome the third challenge, because the drone pilot can capture only behavioural errors that are visible to them, we have reduced possible behavioural errors in the PD (Destatis, 2021) to 11 major categories, including "PriorityError", "RoadUseError" and "DistanceError" (see Appendix, Table 5). The causes of such errors can be obstacles distracting the driver's sight (i.e. "VisualCause"), technical issues (i.e. "TechnicalCause") or weather-related issues (i.e. "EnvironmentalCause").

In general, every VO can consist of several drone flights, with the battery's capacity generally limiting flight times. There are also static and time-based variables for each flight. On the one hand, static variables (e.g. location, weather and measurement equipment) are considered to have stationary status during flights. For example, if the weather changes during a flight, then a new flight has to be started and the variable adjusted accordingly. As for time-based variables, on the other hand, to ensure the subsequent matching of POIs detected by a drone pilot with the traffic conflicts detected in subsequent video analysis, the timestamp of each detected POI is (manually) synchronised with that of the VO at the beginning of the recording (see Figure 4). Thus, POIs are treated as time-based variables, as shown in Figure 4.

Data collection tools

Aside from a video drone and thermometers,[2] the essential survey instrument is an Android smartphone with the ListDB app, which the drone pilot can use to record all static and time-based variables defined in the ListDB Codebook. Figure 5 (left) illustrates the POI screen visible to the drone pilot during video-recording. As shown, the drone pilot can collect all POIs displayed on the screen by clicking on the corresponding buttons. Whereas a short click records a POI and the corresponding timestamp (see Figure 5, bottom), a long click opens another screen showing all possible behavioural errors and causes. After selecting the road user showing the behavioural error, the drone pilot can select up to three errors and one cause. Upon completion, the screen closes, and the data are saved together with the POI and the timestamp.



Date	Time	POI	Error of...	BehaviouralError 1	BehaviouralError 2	BehaviouralErrror 3	CauseError
20221123	130735	Videostart					
20221123	131020	VehicleVehicle	Vehicle	PriorityError	SpeedError	OvertakingError	VisualObstacle
20221123	131240	VehicleCycle	Cycle	PassedPriority	/	/	/
20221123	132010	Videoend					

Figure 5. ListDB App v0.3.2.4 – Screenshot of the POI and behavioural error screens. The table illustrates the format in which the information is stored.

[2] The video drone used was the DJI Mini 2, which offers video recording with a resolution of 3.840 × 2.160 pixels and a sampling rate of 29.97 frames per second.

Population to be observed

A shared, unobserved, superordinate population is a prerequisite for successfully fusing PD and VO data. For that reason, VOs have to target a population intersecting with the PD. Generally, a population and samples drawn from it should be described factually, spatially and temporally (Gabler and Häder, 2015; Lehmann *et al.*, 2019). From a factual perspective, the common population can be formed by all traffic conflicts at the examined locations, in which case traffic conflicts can be regarded as the binding element between PD and VOs (see Figure 3). Based on the factual component of the population, selecting locations where traffic conflicts are known to occur makes sense as a means to determine the spatial component of the population. Indeed, the conflicts need to have occurred in locations where accidents have already happened. Thus, we selected two intersections in Dresden, Germany, for VO, where five relevant crashes occurred between 1 January 2005 and 31 December 2021: a three-way intersection called "Tharandter Straße/ Frankenbergstraße" and a four-way intersection called "Dorfhainer Straße/Kohlenstraße" (see Figure 7).[3] Both intersections are located within the city and therefore have a speed limit of 50 km/h on the priority road. From the temporal perspective, it would make sense to have a permanent VO in place to record all traffic conflicts at the selected intersections. However, because the availability and number of survey personnel, as well as the research's budget, did not allow such monitoring, we defined a 3-month period in which the VO had to occur. As shown in Figure 6, June, July and August constituted the period chosen for VOs, as they are relatively accident-prone months in Dresden from 2017 to 2021 that usually have good weather conditions. Statistics of traffic accidents in Dresden (Figure 6) also illustrate that the occurrence of accidents does not depend on the day of the week; thus, all weekdays can be treated equally in the VO, though the weekend has to be excluded due to the limited availability of staff. Regarding the exact recording times, the following boundary conditions were used to confine continuous VO during all weekdays in the three selected months:

1. **Daylight recording:** Video recordings at night are impossible with the drones used.
2. **Good weather conditions:** The drones cannot fly in strong winds (>21 km/h), rain or snow.
3. **Limited flight time:** The average drone flight time is 20–25 min per battery charge. With the equipment available, about 90 min of recording at a time can be achieved.
4. **Limited access:** The responsible air traffic control authority has to inspect and approve each flight.
5. **Limited personnel budget:** The weekly working time of the two employed drone pilots cannot exceed nine hours incl. arrival and departure as well as data transfer.

Thus, it is necessary to develop a sampling plan to cover the targeted population, one able to ensure that every traffic conflict has the same odds of being considered in the sample and thus guarantee the random selection of traffic conflicts (Bischoff, 1995; Pfeiffer, 2006; Lehmann *et al.*, 2019). The requirements for the sampling plan were therefore:

1. **Fixed time slots:** Fixed time slots have to be considered for recording and covering accident-prone daylight hours. Each time slot can last 90 min maximum.
2. **Equal distribution of time slots:** Each time slot has to be observed once per month at each intersection. All time slots have to be observed the same number of times.
3. **Equal distribution of weekdays:** The time slots have to be evenly distributed across the weekdays. A time slot may only be observed again on the same day of the week when the other weekdays have already been fulfilled.
4. **Flexibility:** The drone pilot has to be free to choose the day of surveying with the given sampling requirements, because the weather conditions have to be suitable, and the flight permit has to be granted.

After analysing the statistics of accidents by hour of occurrence (see Figure 6), we defined four 90-min time slots distributed over the course of a day with daylight as follows:

1. **Time slot 1:** 7:30 a.m. to 9:00 a.m.
2. **Time slot 2:** 10:00 a.m. to 11:30 a.m.
3. **Time slot 3:** 1:00 p.m. to 2:30 p.m.

[3] In-house accident data begin on 1 January 2005. Accidents had to involve two cars, not involve a party under the influence of drugs or alcohol, not involve a trailer and be of accident type 2 (turning), 3 (turning in or crossing), or 6 (longitudinal traffic).

4. **Time slot 4:** 3:30 p.m. to 5:00 p.m.

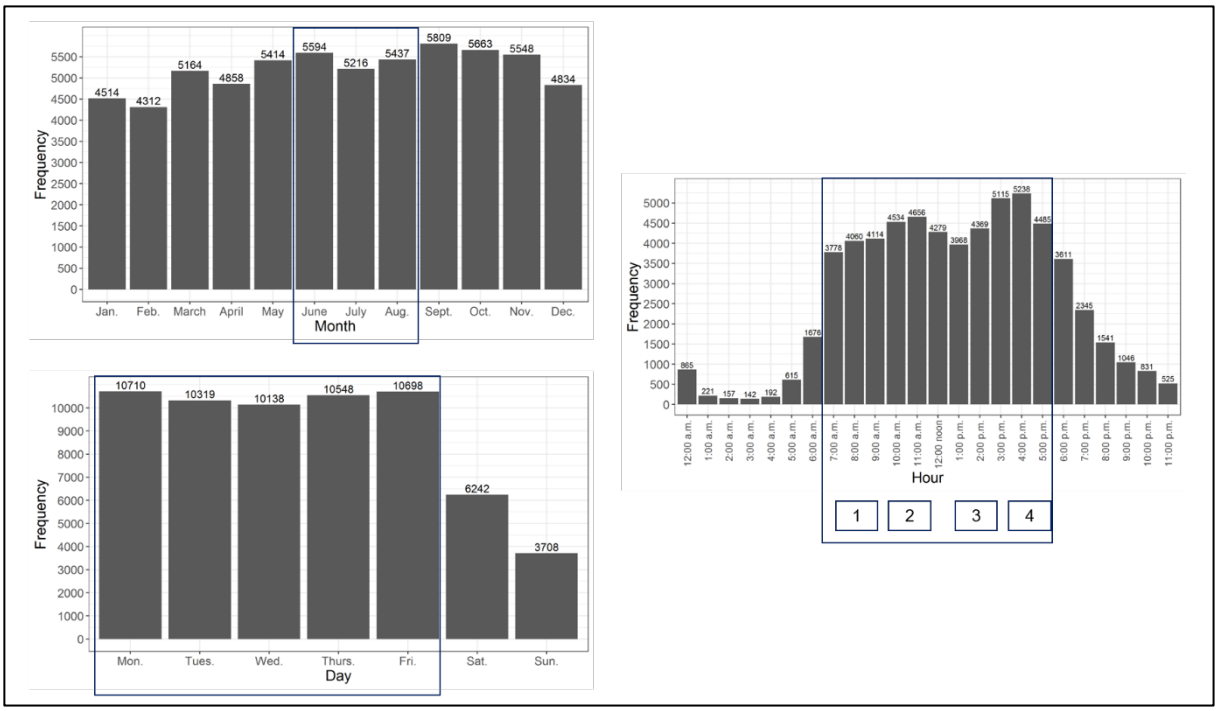


Figure 6. Road traffic accidents in Dresden, 2017–2021. All crashes involved no more than two parties.

Taken together, those boundary conditions and the derived sampling plan slightly changed the target population compared with the desired one (see Table 1). The most significant deviation was that the actual targeted population represented only traffic conflicts in good weather during daylight hours and the defined time slots.

Table 1.
Comparison of the desired and real target population for VOs.

	Factual	Spatial	Temporal
Desired target population	All traffic conflicts in any weather condition, regardless of daylight	Dresden, Tharandter Straße/Frankenbergstraße	Total period from 1 June to 31 August 2022
Actual target population	Traffic conflicts in good weather conditions during daylight hours	Dresden, Dorfhainer Straße/Kohlenstraße	Within the four time slots defined, from 1 June to 31 August 2022

Theoretically possible conflict situations

The FUSE4Rep process model will be applied to generate test scenarios for car specific ADSs in the first stage (Bäumler and Prokop, 2022). However, because subsequent video analysis cannot be used to mine pedestrian and bicycle trajectories reliably, we have focused on detecting traffic conflicts between two cars. At the same time, because the traffic conflicts should affect ADSs in their ODDs, we have not considered traffic conflicts involving only one road user—for example, veering off the road to the right due to the driver's carelessness. We have also not considered parking and animal-related traffic conflicts. Figure 7 illustrates the remaining 3ATs that can theoretically occur at the observed intersections, all 27 of which represent one of four accident types—(2) turning, (3) turning in or crossing, (6) longitudinal traffic, and (7) other (Ortlepp and Butterwegge, 2016)—defined as follows:

2. **Turning:** Conflicts between a road user turning and another road user coming from the same or opposite direction;
3. **Turning in or crossing:** Conflicts between a road user who is turning or crossing but obliged to wait for another road user with the right of way;
6. **Longitudinal traffic:** Conflicts between road users moving in the same or opposite direction; and
7. **Other:** Conflicts that cannot be classified into any other category.

Figure 7 also displays traffic conflicts that do not apply to the three-way intersection observed—namely, the blue-coloured traffic conflicts 215, 301, 321, 602, 612 and 651—and that, of the five crashes observed at either intersection, only the 3ATs 201, 211, 302 and 601 were represented.

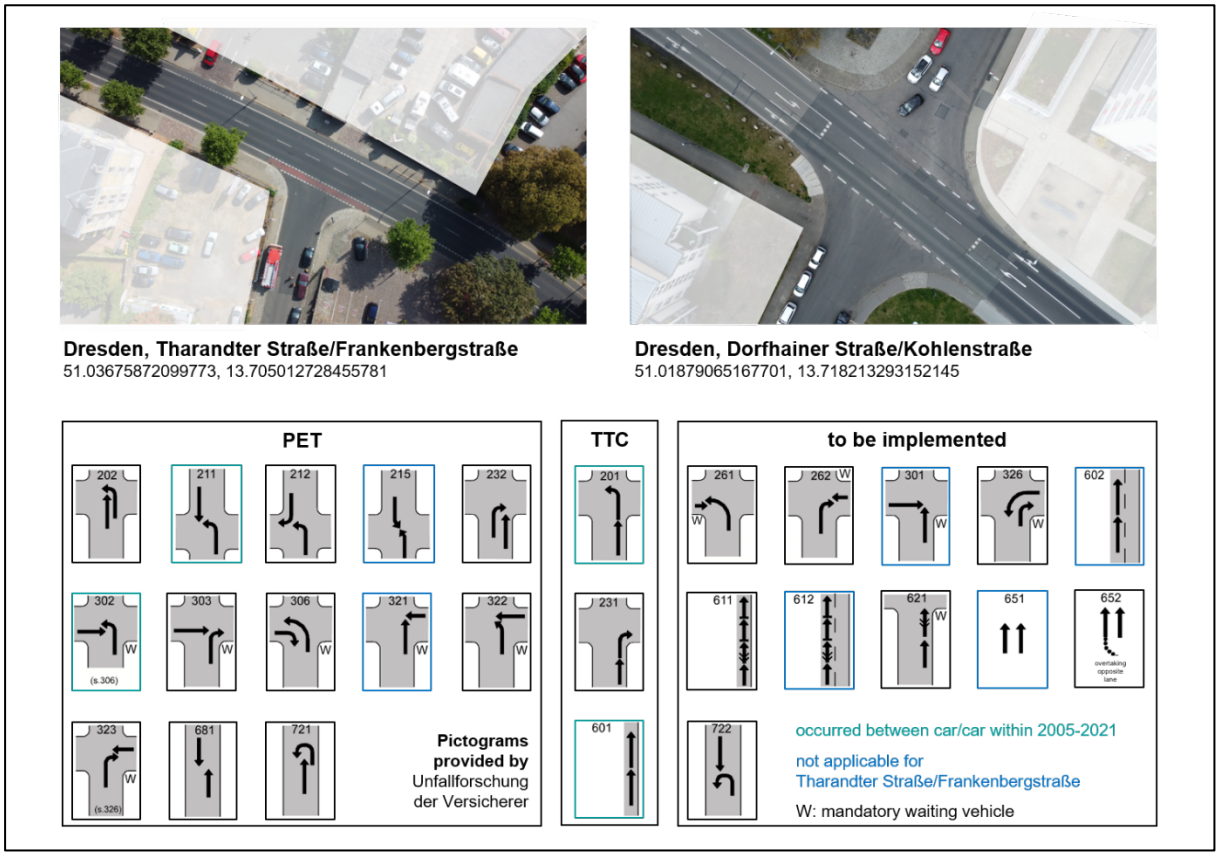


Figure 7. Intersections observed from a height of approximately 80 m with theoretically possible and implemented three-digit accident types.

Step 2: Unify

In what follows, we introduce all steps necessary for analysing VO data in the second step, "Unify", of the FUSE4Rep process model (see Figure 2).

Trajectory mining

Our drone-based observations each delivered a video file, from which certain information about the dynamic behaviour and properties of the road users and their interactions has to be extracted. Such information encompasses:

- Properties of road users
  - Object type (i.e. car, van, truck, biker, cyclist and pedestrian)
  - (Transit) manoeuvres (i.e. turn right, turn left and go straight)
  - Kinematics (i.e. velocity, acceleration, location and manoeuvre-specific development over time)
- Interactions with other road users
  - Surrogate safety measures (SSM), including time to collision (TTC; Hayward, 1972), and post-encroachment time (PET; Allen, Shin and Cooper, 1978)

Extracting that information from the video files requires the steps shown in Figure 8, all adapted from Khan *et al.* (2017). In preprocessing, all unnecessary parts of the recording (e.g. flight start) are removed, followed by image rectification required by the non-ideal parameters of the camera lens.

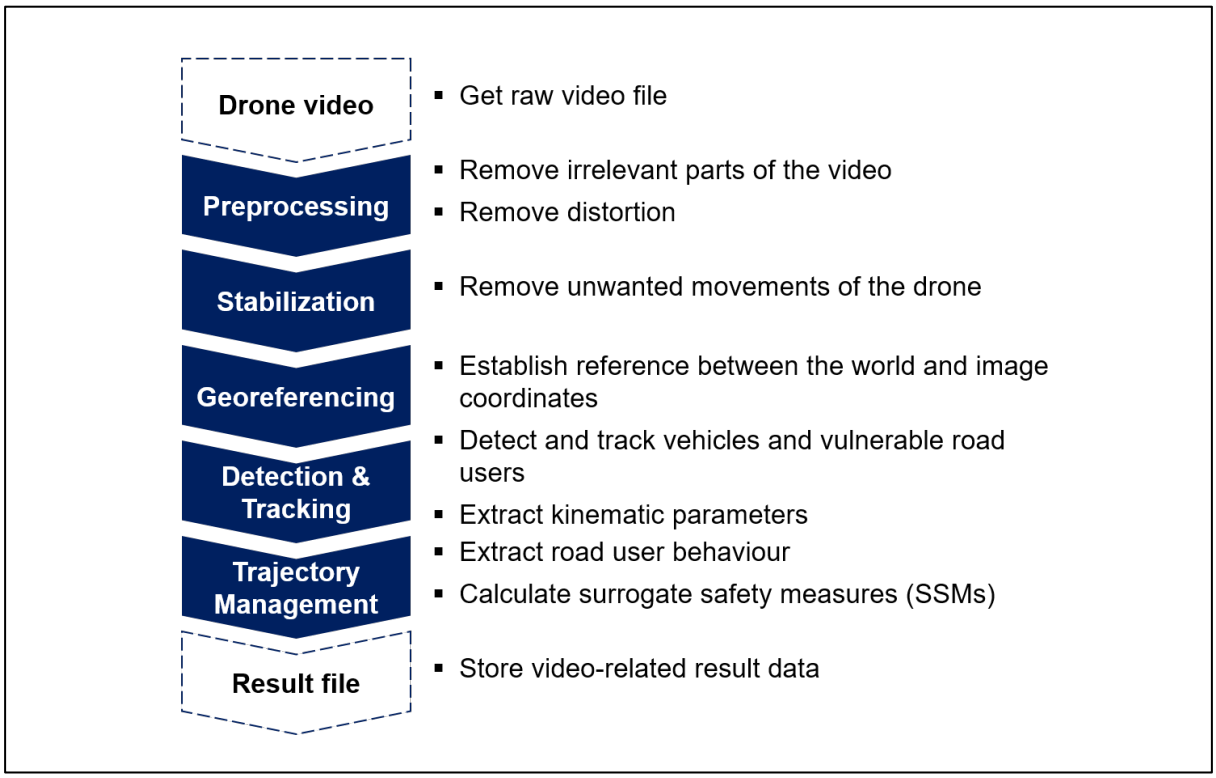


Figure 8. Steps of information extraction, adapted from Khan et al. (2017).

Before being analysed, the recordings have to be stabilised. When in operation, the drones are exposed to external influences from the wind that, despite each drone's internal control strategy, cause unwanted movements and thus changes in the recording angle relative to the ground (see Figure 9). For example, the detected vehicle in the red box in Figure 9 would be repositioned on the sensor due to such movements, which would give it a velocity (v) and acceleration (a) during detection that misrepresent reality. To overcome those movements, all video frames are mapped to a so-called base frame at the beginning of each recording. Once the recordings have been stabilised, the relationship between each

drone's coordinate system (x_{image}, y_{image}, see Figure 9) and a geodetic reference system—for example, WGS84 (United States - Defense Mapping Agency, 1987)—has to be established in order to be able to convert the kinematic parameters from pixel to metric units (Hackeloeer *et al.*, 2014). Next, the video data are prepared to allow the extraction of the required information. During detection, all objects of interest in the frames have to be detected so that the information between the frames can subsequently be merged into the tracking part. Correct detection is the only way to ensure that objects of interest are always clearly identifiable over time and to extract the kinematic parameters correctly. Last, in trajectory management, all required data about drivers' behaviour (e.g. manoeuvres) are derived and simple conflict situations measures (e.g. TTC) calculated.

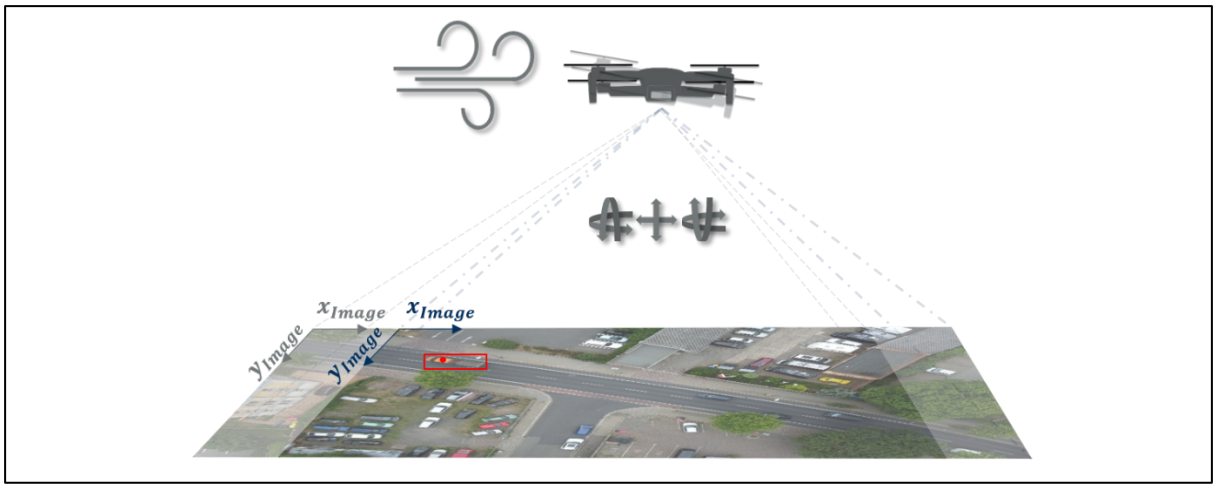


Figure 9. Drone movement due to external influences.

In video analysis, a wide variety of algorithms are needed to perform individual tasks. To that end, and to respond to shifting insights and requirements, a custom tool called "track in drone view" (tidv) has been developed, namely on a microkernel architecture (Richards and Ford, 2020). By outsourcing the individual tasks (e.g. keypoint detection as part of stabilisation) to plug-ins, different algorithms can be used, and the functionalities can be easily extended. The structure of tidv is outlined in Figure 10.

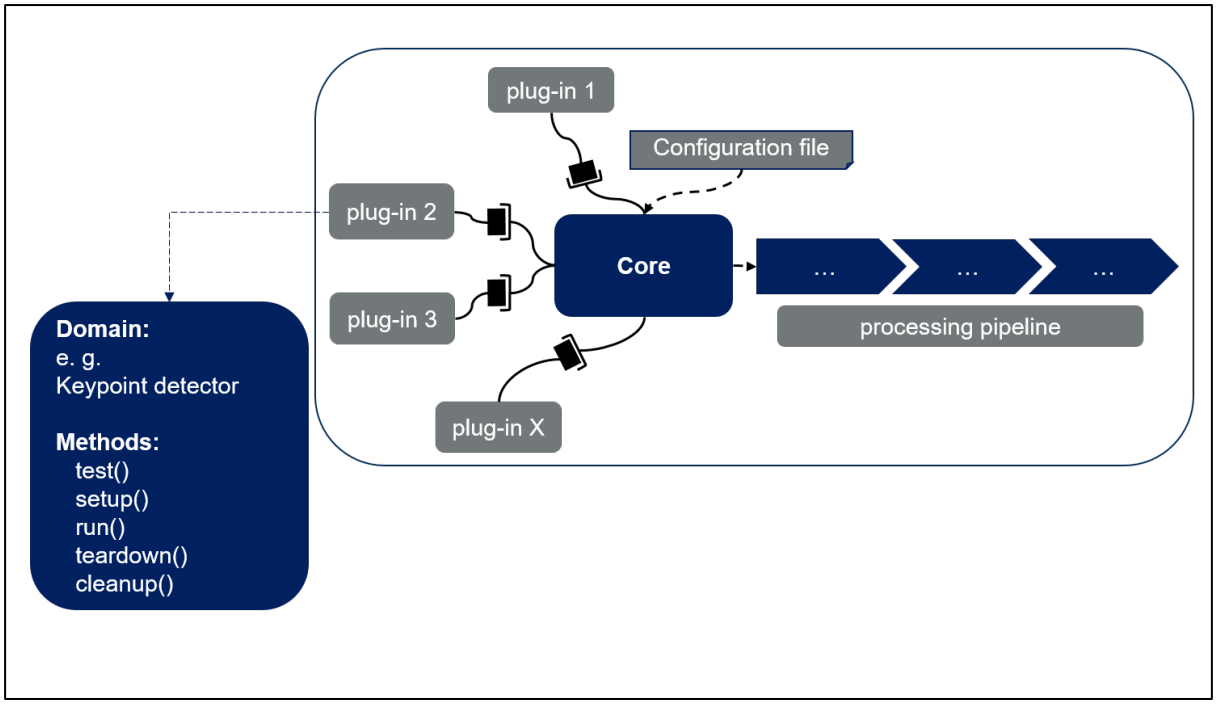


Figure 10. Track in drone view structure (tidv) and plug-in structure.

The core of tidv completely takes over the control and monitoring of the workflow processes. For its part, the workflow is defined via a configuration file written in YAML (Ben-Kiki, Evans and Ingerson,

2021) that can be created manually by the user or with the help of a graphical user interface. With reference to the processing steps, the processing pipeline is checked for consistency by the core before the plug-in's execution. Each plug-in follows a fixed structure (see Figure 10 left side), in which the associated domain is first defined, which determines the data interfaces because the data input and output formats are defined for each domain within tidv. A plug-in's five defined methods to be implemented are test, set-up, run, clean-up and teardown. First, *test* involves checking the behaviour and presence of the necessary data before the setup's execution. Second, *set-up* involves preparing the *run* method for its execution; for example, the necessary neural network weights have to be loaded such that the trained network is available. Third, *run*, occurring at the core of the plug-in, performs the task, after which *clean-up* involves eliminating data that are no longer needed. Last, *teardown* involves terminating the pipeline if an unexpected error occurs. In all, the plug-in's structure allows performing a modular, continuously expandable analysis of traffic observations using tidv.

Because the implementation of the plug-ins is beyond this paper's focus, we discuss only the two plug-ins that are essential in our work: one for recognising road users' manoeuvres, the other for calculating TTC. First, for manoeuvre recognition, the user defines gates on the base frame of the stabilised video (i.e. reference mapping frame for stabilisation), as shown in Figure 11, where white lines indicate the start and end of the junction arm and the black lines the start and end of the central intersection area. Thus, and as shown, the intersection is divided into respective arms and a central area. The start and end points of the trajectories in the defined areas can be used to determine the manoeuvre performed by the road user (i.e. turning right or left or driving straight), which is connected with the individual analysis of the dynamics and heading of the road users performed to extract manoeuvres inside the specified areas. By comparing the position within the areas, the current state of the intersection can also be determined (e.g. access to or departure from the intersection or passage of the central area).

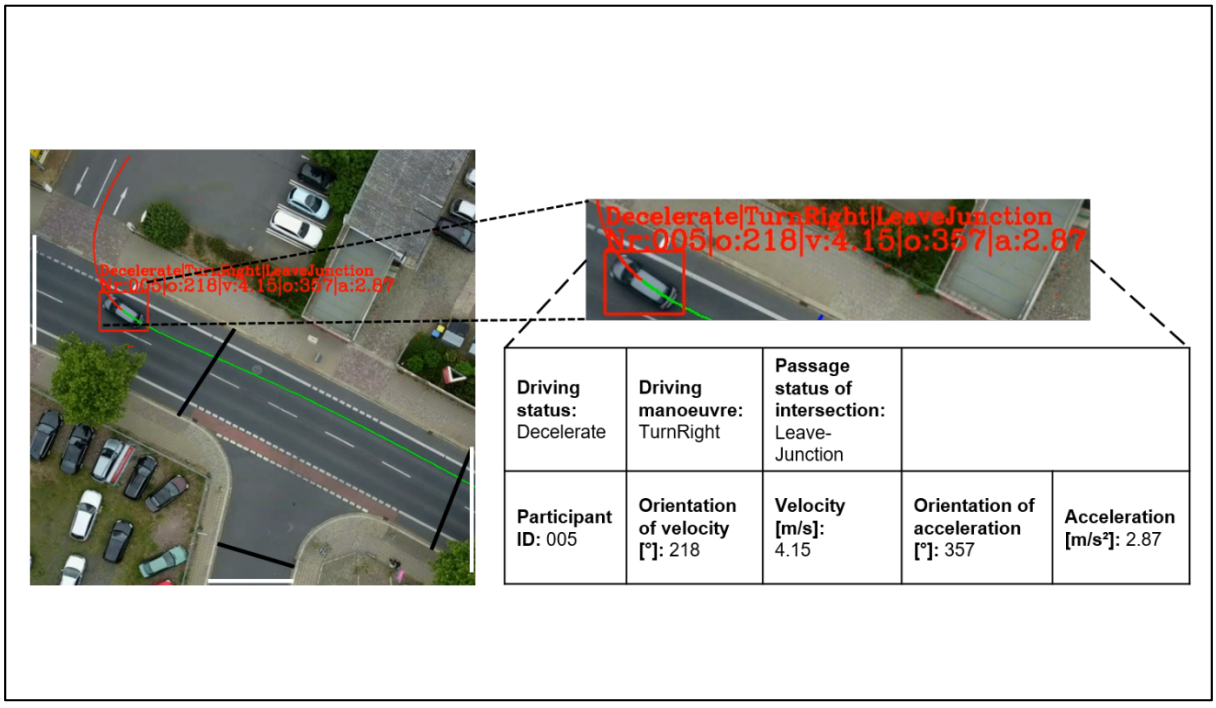


Figure 11. Analysis of a road user's manoeuvre.

To calculate TTC, the vehicles' geometries are first estimated entirely because the YOLOv4 detector (Bochkovskiy, Wang and Liao, 2020) uses only axis-aligned bounding boxes as outputs ("red box framing the vehicle in Figure 11"). The minimum size of the bounding box can be determined by rotating the image to estimate the size of the road users. The TTC for conflict situations, in which one car follows another car (i.e. situations corresponds to 3AT 601) can be calculated with Equation 1:

$$TTC_i(t) = \frac{\left(X_f(t) - X_l\right) - S_f}{v_f(t) - v_l(t)} \tag{1}$$

in which X is the vehicle position of the following (f) or leading vehicle (l) and v is the velocity of the vehicle and S the vehicle's length (Wang *et al.*, 2021).

By contrast, PET is calculated to determine the time elapsed between one road user leaving the potential collision zone (t_1) and another road user entering it (t_2). Following Allen, Shin and Cooper (1978), PET is defined as shown in Equation 2:

$$PET = t_2 - t_1 \tag{2}$$

Last, as mentioned, processing a large amount of data with tidv was only partly possible owing to the hardware's limited capacity. Thus, commercial provider DataFromSky (DFS) supported video analysis.[4] Although we assumed that DFS's video-processing method was similar to our developed procedure, in the data obtained using DFS, TTC and PET had to be calculated using estimated vehicle dimensions (i.e. width and length).

Determining accident type

We next analysed all four 3ATs that resulted in at least one crash between 2005 and 2021 (see Figure 7). Beyond that, we selected 12 additional 3ATs, also listed in Figure 7, to be identified in the VO data. However, due to time constraints, we did not analyse the 12 other theoretically possible 3ATs listed in Figure 7.

Ascertaining a 3AT requires the correct manoeuver classification and SSM calculation. Building upon the mined trajectories, the process consists of seven steps, as shown in Figure 12, beginning with (1) taking a pair of trajectories t_i, t_j and (2) checking whether the road users belonging to the trajectories are visible in the video at the same time. If so, then (3) the corresponding manoeuvers m_i, m_j need to be obtained, including the manoeuver direction (e.g. turning left from B to C), at which time the directions depend on the location observed (e.g. a four-way intersection).

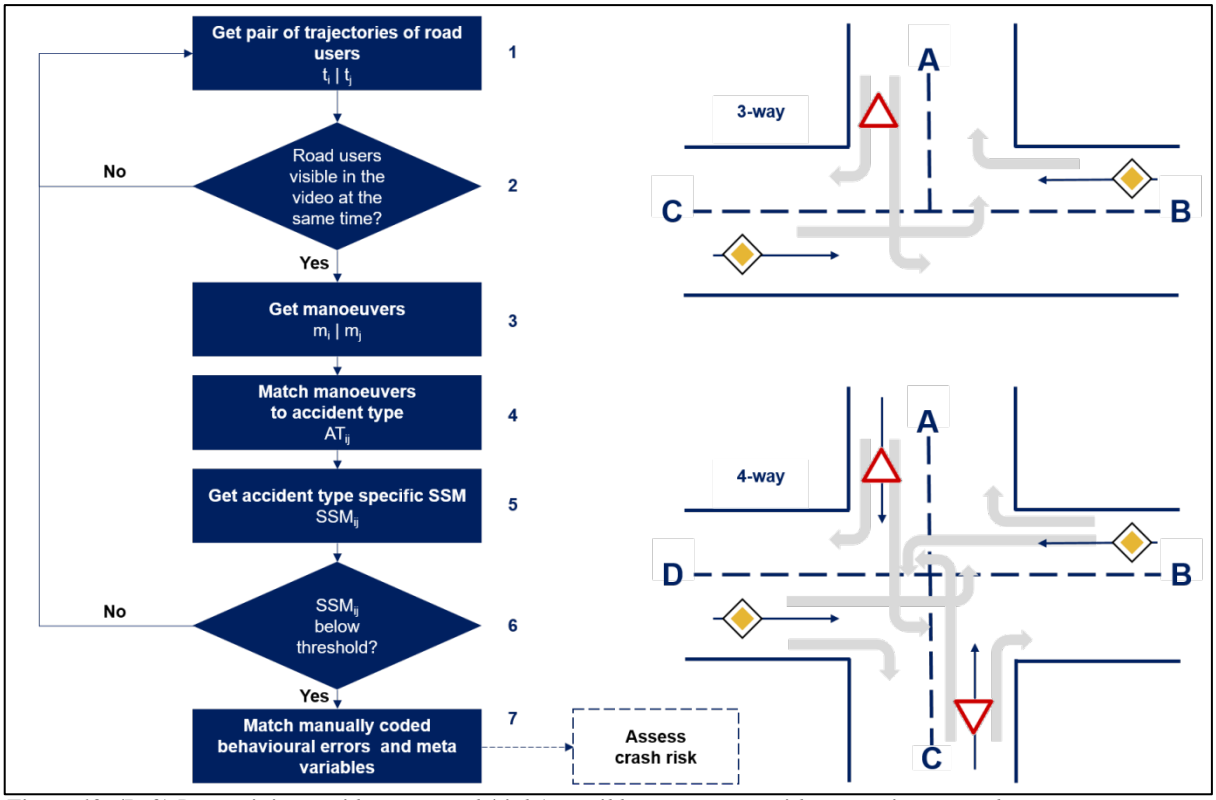


Figure 12. (Left) Determining accident type and (right) possible manoeuvres without turning around.

[4] www.ai.datafromsky.com/aerial

Next, (4) whether the combination of obtained manoeuvers matches a predefined 3AT AT_{ij} needs to be checked; for example, the manoeuver turning left from B to C and going straight from D to B could theoretically represent 3AT 211 (see Figure 7). After that, (5) the SSM_{ij} assigned to the previously determined AT_{ij} needs to be calculated. For 3AT 211, for instance, the PET is suitable for assessing conflicts between the two road users (see Figure 7). Once done, (6) whether the SSM_{ij} is in a predefined threshold needs to be determined in order to sort out irrelevant conflicts (e.g. $< 5s$ according to the length of the GIDAS pre-crash-matrix (Schubert, Erbsmehl and Hannawald, 2013)). Last, once the threshold is met, (7) the corresponding meta variables and manually coded behavioural errors can be matched to the determined AT_{ij} using the common timestamp (see Figure 5) between manual coding and the video-recording. As a result, the risk of a crash can be assessed.

Assessing the risk of a crash

Once the identified traffic conflicts were categorised according to the 3AT classification (see Figure 7), we examined the risk of a crash, or *crash risk*, of each 3AT to ensure that the 3AT populations obtained were relevant for fusion with PD. Because traffic conflicts can lead to crashes but do not have to (see Figure 3), we wanted to assess whether the detected traffic conflicts had at least an inherent risk of leading to a crash. If the risk was significantly greater than zero, then the traffic conflicts were suitable for subsequent fusion. Thus, we modelled crash risk using generalised extreme value (GEV) distributions (Zheng and Sayed, 2020) based on corresponding 3AT SSM distributions.

Zheng and Sayed (2020) have already predicted crash risk, R_C, for traffic conflicts in longitudinal traffic equivalent to 3AT 601 (see Figure 7), in which they used the modified TTC to assess conflicts and model the corresponding GEV distributions. For validation, in this paper we determine the crash risk for each 3AT, $R_{C,3AT}$, and predict the number of crashes, $N_{C,3AT}$, likely to occur in a year. Having a predicted number of crashes per year allows a comparison with PD within the scope of validation.

We applied the process from Coles (2013) and Zheng and Sayed (2020) to each of the observed locations (i.e. intersections without traffic signals) as follows:

1. **Extreme value determination:** Determining extreme values by data blocking, in which each detected traffic conflict represents a block represented by the corresponding SSM distribution (see Figure 13), and the SSM value from the block is taken to represent the maximum critical situation (i.e. SSM minimum);
2. **GEV modelling:** Estimating the corresponding GEV distribution $G_{3AT}(SSM)$ using maximum likelihood estimation,[5] particularly of the scale parameter σ, the location parameter μ and the shape parameter ξ;
3. **Risk calculation:** Making the corresponding SSM zero in the event of a crash, as shown in Equation 3;

$$R_{C,3AT} = G_{3AT}(0) = \begin{cases} exp\left\{-\left[1+\xi_i\left(-\frac{\mu_i}{\sigma_i}\right)\right]^{-\frac{1}{\xi_i}}\right\}, & \xi \neq 0 \\ exp\left[-exp\left(\frac{\mu_i}{\sigma_i}\right)\right], & \xi = 0 \end{cases} \quad (3)$$

4. **Extrapolation:** Predicting the number of crashes for a year by extrapolating the location-dependent observation time t_l, as shown in Equation 4; and

$$N_{C,3AT}\left(T = 1\ year\right) = R_{C,3AT} * \frac{60[min]}{t_l[min]} * 24[hours] * 365[days] \quad (4)$$

[5] We used the function "fevd" from the R package extRemes (version 2.1).

5. **Validation:** Comparing the predicted $N_{C,3AT}$ with the corresponding PD.

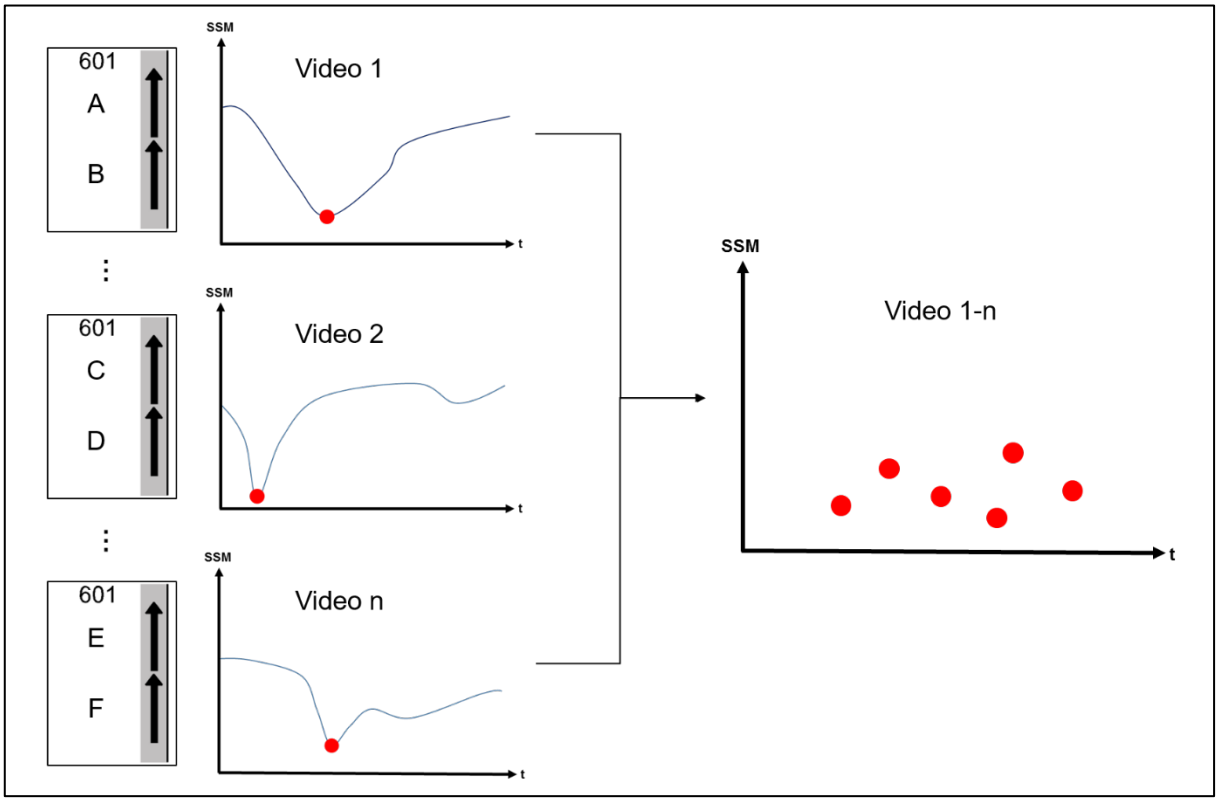


Figure 13. Determination of extreme values dependent on the 3AT.

RESULTS

Collected data

Table 2 introduces the data collected by two drone pilots: Pilot A and Pilot B. We assigned each pilot to one intersection depending on their route of approach to the intersection; Pilot A observed the three-way intersection for 790 min (13.2 h) and Pilot B the four-way intersection for 855 min (14.3 h), both in similar weather conditions. Among the results, Pilot A marked 2.4 times more POIs at the three-way intersection (i.e. 1827) than Pilot B at the four-way intersection (i.e. 774). Pilot A also marked eight near crashes and 241 POIs associated with behavioural errors, whereas Pilot B observed no near crashes and associated only 10 POIs with behavioural errors. At the same time, Pilot B associated 8 of 11 behavioural errors with visual causes (e.g. obstacles obstructing the drivers' view), whereas Pilot A recorded no such errors.

Table 2.
Overview of the data collected.

	Tharandter Straße / Frankenbergstraße	***Dorfhainer Straße / Kohlenstraße***
Drone pilot	A	B
Intersection type	3-way	4-way
Observation duration	1 June–31 August 2022	
Number of days observed	11	12
Recording time (in minutes)	790	855
Average temperature (in median °C)	21	22
Number of points of interest	1827	774
Number of crashes	0	0
Number of near crashes	8	0
Number of vehicle–vehicle interactions	1154	523
Number of behavioural errors	241	11
Number of vehicles with behavioural errors	175	11
Number of causes in total	0	8
Number of visual causes	0	8
Number of technical causes	0	0
Number of environmental causes	0	0

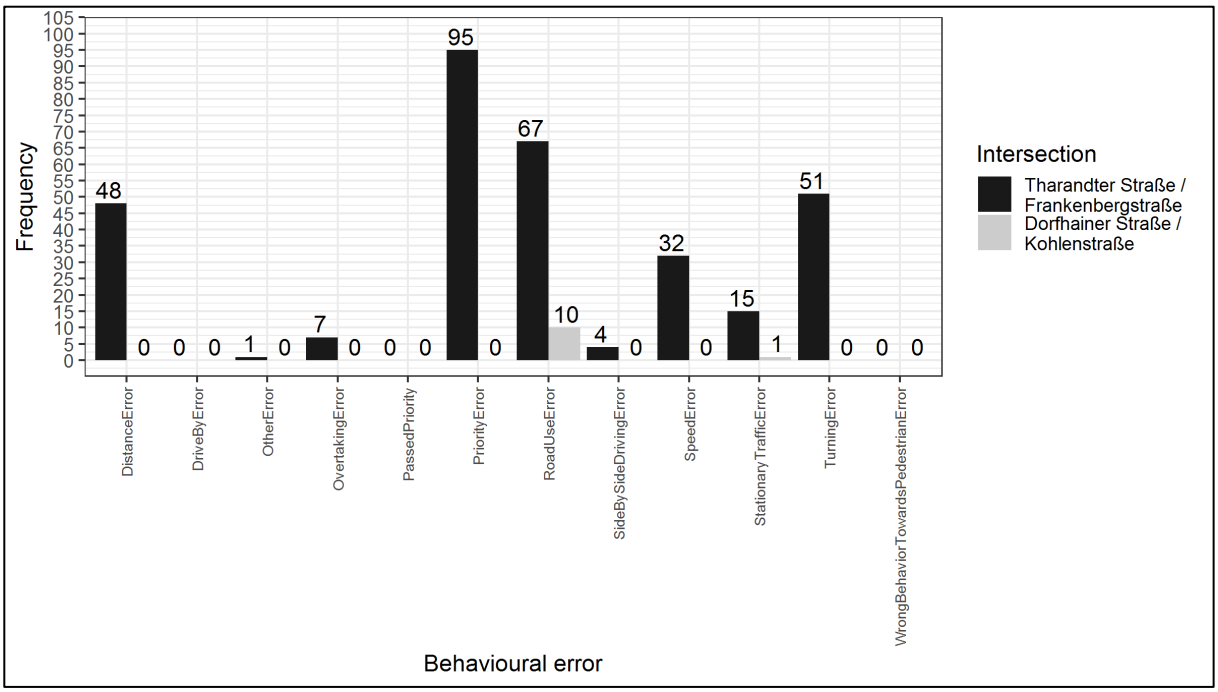


Figure 14. Type of behavioural errors.

Figure 14 highlights the different behavioural errors observed by the drone pilots. Overall, Pilot A observed nine types of errors at the three-way intersection, among which priority errors (95×), road use errors (67×) and turning errors (51×) were the most frequent. Meanwhile, Pilot B observed only two types of behavioural errors: 10 road use errors and one stationary traffic error. The aspect of "PassedPriority", meaning when a road user voluntarily gives up the right of way to another road user, was not logged during the observation and introduced only after the observation had ended; as we found this aspect to be important during the observation. Moreover, neither pilot observed the behavioural errors "DriveByError" or "WrongBehaviourTowardsPedestrianError".

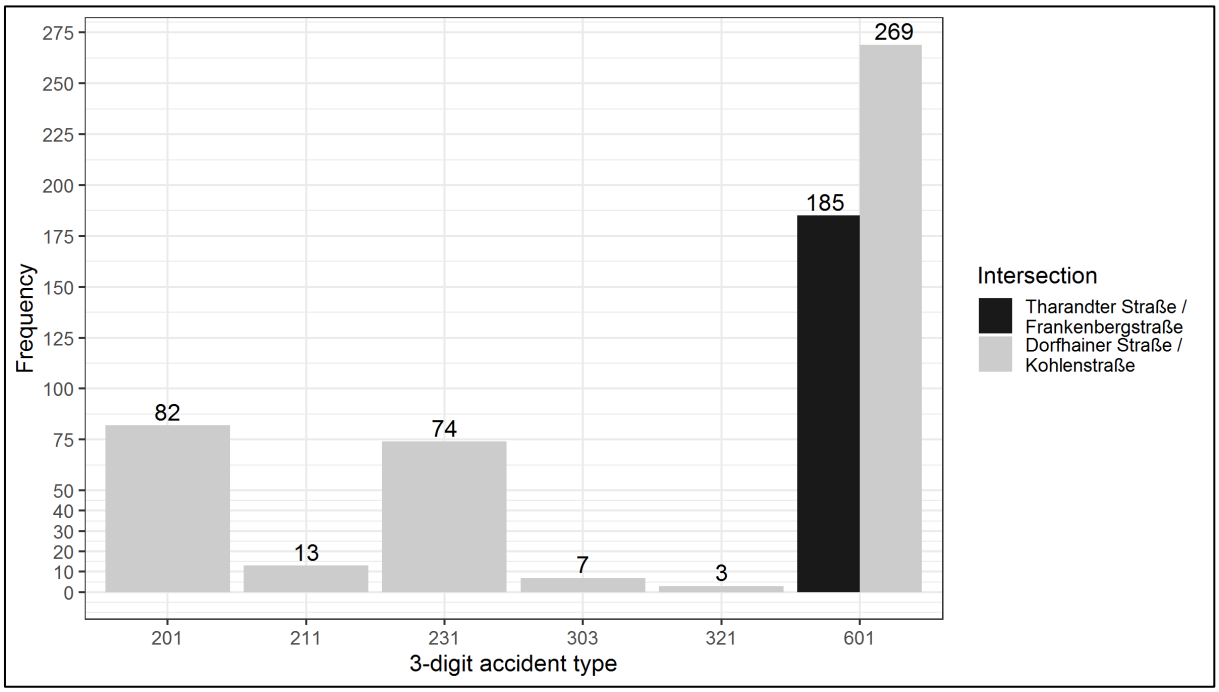


Figure 15. Detected conflict situations according to 3AT.

Determined conflict situations

Figure 15 introduces the automatically detected conflict situations between two cars, categorised according to the 3AT classification shown in Figure 7. Overall, 185 conflicts were detected at the three-way intersection, which represented 15.9% of all POIs marked by Pilot A, versus 448 conflicts detected at the four-way intersection, which represented more than 85.7% of the POIs marked by Pilot B. Remarkably, only 3AT 601 appeared in the three-way intersection, whereas that type and five additional 3ATs (i.e. 201, 211, 231, 303 and 321) appeared in the four-way intersection. Furthermore, as Table 3 reveals, only one type of behavioural error, "DistanceError", could be matched with the detected conflicts using the timestamp. An example of matching for DistanceError appears in Figure 16; although the white car with the red dot keeps a distance of 5.78 m while going 50 km/h, in Germany a distance of 15 m has to be maintained at that speed in urban areas (Autowelt, 2022). By comparison, the black car in the middle maintains a distance of 13.32 m, which is closer to the prescribed 15 m. The other behavioural errors listed in Figure 14 could not be assigned. Table 3 also introduces the conflict-related SSM distributions based on the extreme values revealed by data blocking. For example, 3AT 601, detected 178 times at the three-way intersection, had an extreme value distribution $G_{601}(TTC)$, with an average TTC of 1.53 s (median), a 25th percentile of 0.86 s and a 75th percentile of 2.25 s. To determine the extreme values, we removed SSM values less than 0.2 s if they did not lead to a crash, which was always the case. Therefore, the conflicts detected and the sample size, n, of the respective distributions differed (see Table 3). Beyond that, because the corresponding sample sizes for 3AT 601 (i.e. at the three-way intersection), 303 and 601 were quite small for GEV modelling, the distributions determined are not especially meaningful.

Table 3.
Detected conflict situations at the observed locations.

	No.	***Frequency***	***3AT***	***Behavioural error***	***Cause***	***SSM***	***n*** *(0.2 s > SSM < 5.0 s)*	***SSM [s]*** *(25th percentile | median | 75th percentile)*
Tharandter Straße/ Frankenbergstraße (3-way)	1	178×	601	/	/	TTC	178	0.86 | 1.53 | 2.25
	2	7×	601	Distance error	/	TTC	7	1.42 | 1.74 | 1.92
Dorfhainer Straße/ Kohlenstraße (4-way)	3	82×	201	/	/	TTC	27	0.86 | 1.08 | 1.60
	4	13×	211	/	/	PET	13	1.07 | 1.50 | 1.80
	5	74×	231	/	/	TTC	67	0.60 | 1.22 | 1.62
	6	7×	303	/	/	PET	7	2.20 | 2.64 | 2.64
	7	3×	321	/	/	PET	3	2.24 | 2.24 | 2.45
	8	269×	601	/	/	TTC	263	1.32 | 1.72 | 2.12

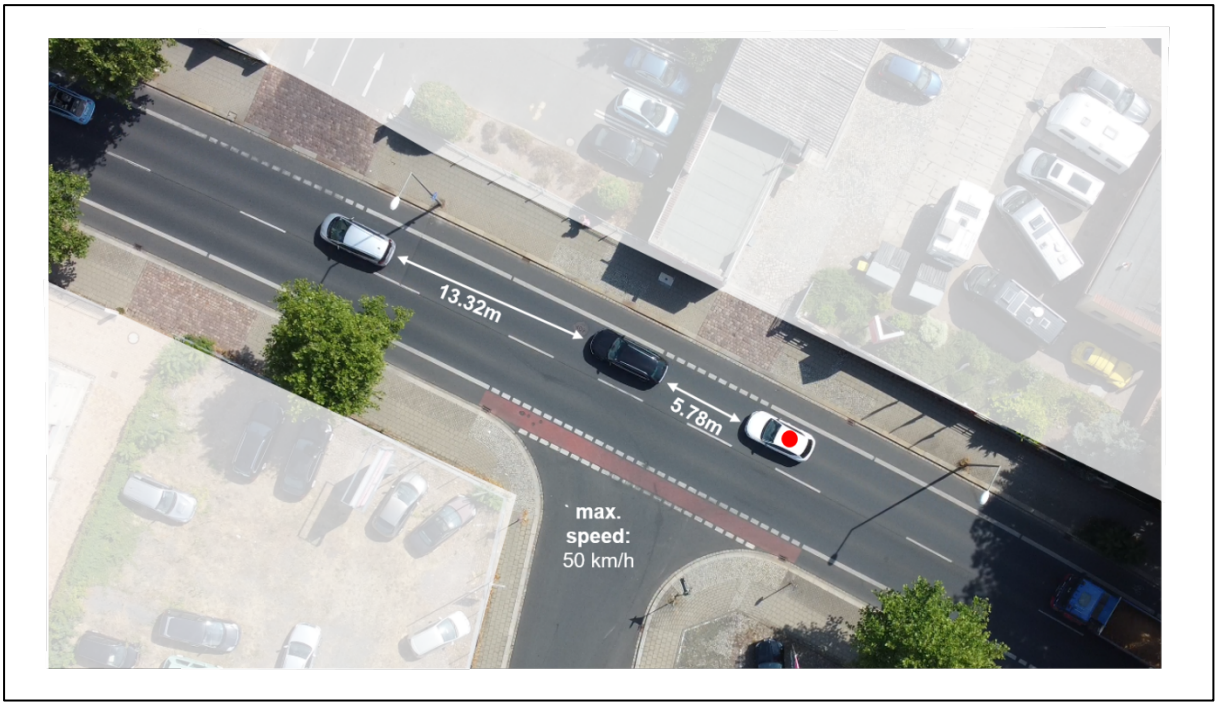


Figure 16. Manually coded distance error of car with red dot at Tharandter Straße/Frankenbergstraße.

Modelled crash risk

Table 4 lists the results of the modelled crash risk compared with official PD for the observed locations. Overall, the predicted number of crashes per year was rounded to 16 for the three-way intersection and to one for the four-way intersection. For the three-way intersection, the 3ATs 302 and 201, which led to crashes between 2005 and 2021, could not be predicted due to missing data (see Figure 15). As for the four-way intersection, the predicted results and official PD seem to coincide for 3AT 601; whereas 3AT 601 never led to a crash between 2005 and 2021, the predicted number of crashes was 0.012 and thus close to 0. The graphical analysis in Figure 17 (right) confirms that positive result, for the modelled distribution and empirically determined data coincide well. However, 3AT 211 occurred only once between 2005 and 2021 but was predicted to occur 1.3 times a year. At that rate, over 16 years, 3AT 211 would have occurred approximately 20 times, which is 20 times more than captured in the PD. Added to that, 3AT 601 also occurred only once between 2005 and 2021 at the three-way intersection but was predicted to occur 16 times a year, or 240 times more than in the PD across the 16-year period. Those results are confirmed in Figure 17 (left), which shows that the modelled and empirical distributions coincide poorly. In total, the predicted crash risks illustrate that the identified traffic conflicts according to the 3AT 601 for both intersections and the 3AT 211 have an inherent crash risk and seem suitable for subsequent fusion with PD.

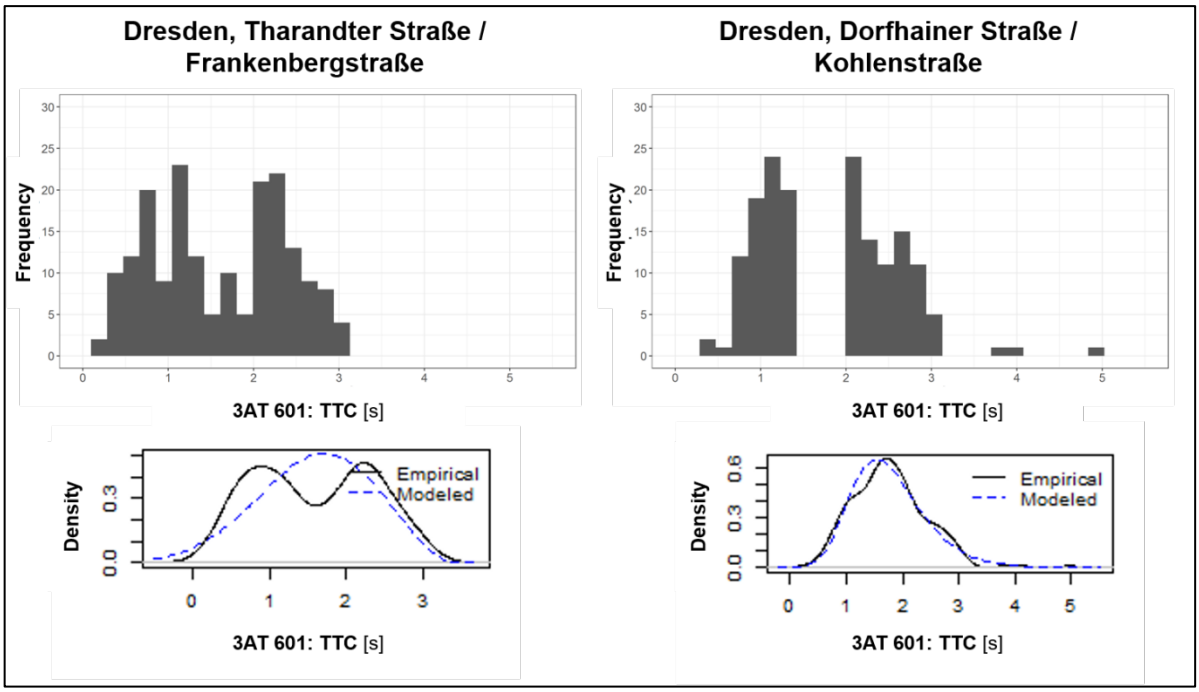


Figure 17. Histogram of TTC and GEV modelling results for 3AT 601.

Table 4.
Results of the modelled risk of a crash.

	3AT	*Number of car–car crashes, 2005–2021*	*n*	*μ*	*σ*	*ξ*	*RC*	*Predicted crashes per year*
Tharandter Straße / Frankenbergstraße (3-way)	601	1	185	1.363	0.788	−0.393	0.0237	15.74
	302	1	Not detected in traffic observation					
	201	2	Not detected in traffic observation					
Dorfhainer Straße/ Kohlenstraße (4-way)	601	0	263	1.491	0.564	−0.084	1.882e-05	0.012
	211	1	13	1.284	0.557	−0.249	0.0021	1.268

DISCUSSION

This paper reveals that the design of VO studies and the information to be collected during VOs should be planned to accommodate subsequent fusion with PD. By adopting the FUSE4Rep process model to VO using drones, we have demonstrated how to choose observation locations, derive sampling plans and collect relevant information, especially on behavioural errors. Along those lines, the sampling plan should be oriented towards an unobserved, superordinate population with overlap with the PD to be fused. The information to be collected should also coincide with information contained in the PD, information needed to describe test scenarios according to the 6LM and information derived from the adapted causality model of crashes. Due to space constraints, we have published only the information necessary for conducting drone-based VO in the ListDB Codebook. Even so, we have also shown how to mine trajectories efficiently, classify traffic conflicts according

to the 3AT classification and how to match the traffic conflicts with behavioural errors. Last, we have demonstrated how to assess the inherent crash risk of the identified traffic conflicts using extreme value theory.

The efficient, accurate interpretation of VO data and subsequent generation of test scenarios require combining human knowledge with automatic video analysis. While video analysis leveraging computer vision algorithms can help to mine road users' trajectories and detect traffic conflicts, it cannot contextualise the detected conflicts due to relying only on the drone's bird's-eye view. Even so, the drone pilot can help to detect barely visible parking crashes or behavioural errors, including informal agreements with other road users about yielding the right of way; identify causes of behavioural errors such as visual obstacles or technical problems; and note any major traffic-impacting events (e.g. football games in a nearby stadium that are not visible in the video-recording). Nevertheless, some of the information collected from pilots is highly subjective, especially when it concerns behavioural errors such as "DistanceError" or "SpeedError". Another challenge is the real-time synchronisation of points of interest and information collected by pilots with the traffic conflicts detected in the video-recordings made by drones. Especially at busy intersections, delays of only one or two seconds in the synchronisation can result in incorrect assignments, e.g. one second delay corresponds already to 13.8m distance travelled at a permitted speed of 50 km/h.

Categorizing traffic conflicts according to the 3AT classification leads to a good fit with PD according to the FUSE4Rep process model. At the same time, the approach detects only known constellations of traffic conflicts, whereas conflicts that do not fall into the 3AT classification cannot be sufficiently considered. Furthermore, the choice of the appropriate SSM determines the quality with which the conflict is described. For example, the TTC does not take acceleration behaviour into account.

GEV modelling allows analysing the inherent risk of traffic conflicts found separately according to 3AT classification and independently of other simulation methods such as stochastic traffic simulations (Siebke *et al.*, 2023). However, the modelled GEV distributions heavily depend on outliers and thus accurately calculated SSM values. Evaluating the data of the calculated SSMs has revealed that, especially at the three-way intersection, low values often result (e.g. TTC < 0.2 s), which typically precipitate crashes. A more detailed video analysis of traffic conflicts that stand out due to low TTC values suggests that the reason may be poorly fitting bounding boxes and, in turn, wrongly calculated vehicle dimensions. Figure 18 illustrates that problem; whereas the following car with ID 204 is very close to the leading car in front, a crash is highly unlikely because the vehicles do not overlap in their dimensions, and, thus, the following car could drive past. However, if the object detector incorrectly determines the vehicle dimensions and positions, then the TTC will be miscalculated. Furthermore, GEV modelling should be refined to take the overall scene embedding the traffic conflicts into account. For example, traffic flow can significantly affect the overall safety and crash risk at intersections (Oh, Washington and Choi, 2004), which can consequently affect the distribution of extreme values. A next step is therefore to include corresponding covariates in GEV modelling (Zheng and Sayed, 2020). Last, quantitative measures, including the coefficient of determination R^2 and the Akaike information criterion, to assess the quality of the GEV modelling, need to be introduced.

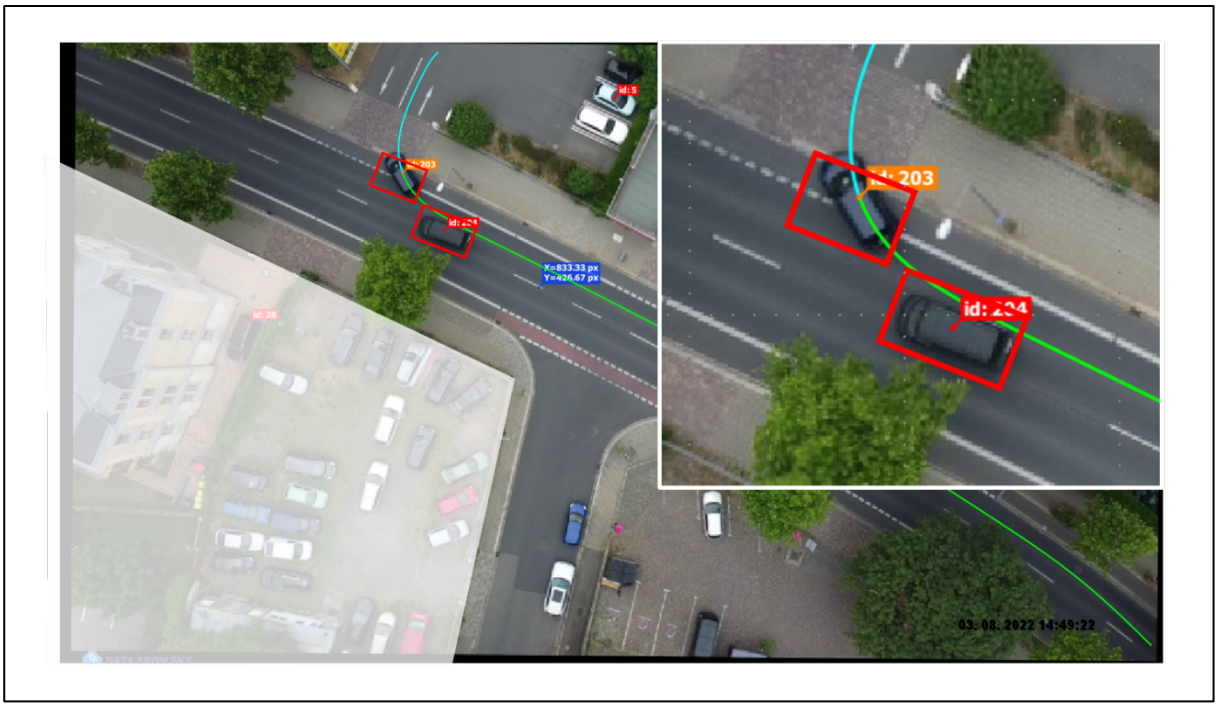


Figure 18. Traffic conflict detected with miscalculated TTC due to poorly fitting bounding boxes.

Limitations
In the data collected by the drone pilots, the unknown inter-rater reliability is a major limitation. Although both pilots received the same intensive training and supervision during observations, each pilot was assigned an observation site in which they specialise. Thus, the results of the pilots at the different survey locations were not compared with each other.

The validation of GEV modelling results using PD was also limited in its informative value. On the one hand, the number of accidents with five relevant crashes between 2005 and 2021 was quite small for comparison. On the other, the prediction referred to future traffic events, whereas the accident data referred to past traffic events.

For practitioners
Independent of applying the FUSE4Rep process model, practitioners can use the proposed process for conducting VO studies to collect and analyse data for use in generating test scenarios (Nalic *et al.*, 2020). As a result, practitioners can also transfer the information collected into corresponding OpenX formats (ASAM e.V., 2022) such as OpenScenario and OpenDrive. Practitioners can also improve the survey process by conducting drone pilot training that is more technical and aligned with the Swedish Traffic Conflict Technique (Polders and Brijs, 2018). Also, the mapping of the POIs in the ListDB app should be synchronised electronically in real time with the drone recoding. Additionally, the observation locations could be divided into spatial zones so that for each POI, spatial information is also available for matching with the conflicts detected by the video analysis.

For scientists
Scientists can directly use the proposed application of the FUSE4Rep process model for VO for data to fuse with German PD and derive representative test scenarios. In that process, scientists should also improve the calculation of SSM values, especially TTC; because conflicts in which one car follows another car ideally are only one part of the conflict situations, in which the TTC has to be determined. Thus, the calculation of TTC should be more generalised to be able to take angular situations into account for assessment as well. According to Laureshyn, Svensson and Hydén (2010) and Tarko (2019), generalised TTC calculation should be used to allow a possible collision angle between two road users. In addition, new object detectors should be used to directly detect rotated bounding boxes (e.g. Yang *et al.*, 2021) and thereby deliver more precise vehicle dimensions. Moreover, scientists should study how correcting factors of observation altitude (i.e. approx. 80 m above the ground) affects the determination of the position of objects (Kruber *et al.*, 2020) and might increase the accuracy of the entire video analysis. Last, scientists should investigate the additional automatic detection of behavioural errors and improve GEV modelling by taking appropriate covariates (e.g. traffic flow) into account.

CONCLUSION

Strictly applying the FUSE4Rep process model in collecting and analysing VO using drones supports the generation of representative test scenarios by applying data fusion. In that process, not only VO analysis but also study planning (e.g. sample plan) and the information to be collected need to be aligned in advance with the PD to be fused later on. In data collection, drone pilots can set observed traffic conflicts into context and record behavioural errors of road users, which then can be linked to the traffic conflicts. After that, GEV modelling can help to assess the inherent crash risk of the traffic conflicts identified.

Drone-based VOs for subsequent scenario generation are already publicly available as data sets about location-specific trajectories. Such data sets seldom follow a comprehensible sampling plan, and only a few metavariables are used to describe them. As our results show, it makes sense to embed the drone-based VOs in a systematic in-depth survey, which we describe using the 122 variables of the ListDB Codebook. In addition, the consistent alignment of the traffic conflict analysis with PD and the German 3AT classification can help to ensure that test scenarios can be found quickly and that fusion can be performed. We anticipate that only with the subsequent application of the FUSE4Rep process model will it become possible to find representative test scenarios for the evaluation of ADSs in the future.

ACKNOWLEDGEMENTS

A big thank-you goes to the most important supporters, the drone pilots, Kai Hua and Weikai Wang, who carefully and reliably collected VO data and helpful feedback from the observation sites. We also wish to thank Zheng Feng, Zhenyu Pan, Qilin Tang, Tom Wolf, Shuai Jiang, Tianyi Wang, Zihao Jin and Kevin Gowinkowski for their preliminary work, support and fruitful discussions; Lena Otto for her extensive and helpful feedback; the local government, police and German air traffic control for supporting the drone flights; Susanne Arndt and Matthias Fuchs for promoting and supporting the ListDB project; and DataFromSky (ai.datafromsky.com) for

supporting the video analysis. Our research was conducted within the MOTUS project (19FS2015A) and supported by the mFUND grant by Germany's Federal Ministry for Digital and Transport.

BIBLIOGRAPHY

Allen, B. L., Shin, B. T. and Cooper, P. J. (1978) 'Analysis of Traffic Conflicts and Collisions', pp. 67–74.

ASAM e.V. (2022) *ASAM Standards*. Available at: https://www.asam.net/standards/ (Accessed: 28 December 2022).

Autowelt, A. (2022) *Abstandsmessung - Wann bin ich zu dicht dran?* Available at: https://www.allianz-autowelt.de/bussgeld/abstandsmessung/# (Accessed: 28 December 2022).

Bäumler, M. *et al.* (2020) 'Use Information You Have Never Observed Together : Data Fusion as a Major Step Towards Realistic Test Scenarios', in *IEEE 23rd Intelligent Transportation Systems Conference (ITSC), Proceedings*, pp. 672–679. doi: 10.1109/ITSC45102.2020.9294649.

Bäumler, M. and Prokop, G. (2022) 'Validating automated driving systems by using scenario-based testing: The Fuse4Rep process model for scenario generation as part of the "Dresden Method"', *Zeitschrift für Verkehrssicherheit*, July, pp. 226–230. doi: 10.53184/ZVS3-2022-11.

Ben-Kiki, O., Evans, C. and Ingerson, B. (2021) *YAML Ain't Markup Language (YAML™) version 1.2. Revision 1.2.2 (2021-10-01)*. Available at: https://yaml.org/spec/1.2.2/ (Accessed: 28 December 2022).

Bergmann, L. T. (2022) 'Ethical Issues in Automated Driving—Opportunities, Dangers, and Obligations', *Studies in Computational Intelligence*. Springer Science and Business Media Deutschland GmbH, 980, pp. 99–121. doi: 10.1007/978-3-030-77726-5_5.

Bischoff, M. (1995) 'Zur Problematik der Repräsentativität in der empirischen Sozialforschung'. Bochum: Fakultät für Sozialwissenschaft, Ruhr-Universität Bochum.

Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y. M. (2020) 'YOLOv4: Optimal Speed and Accuracy of Object Detection', *arXiv preprint*.

Bock, J. *et al.* (2019) 'The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections', *arXiv preprint*. Available at: http://arxiv.org/abs/1911.07602.

Coles, S. (2004) *An Introduction to Statistical Modeling of Extreme Values*. 3rd edn. London: Springer-Verlag.

D'Orazio, M., Di Zio, M. and Scanu, M. (2006) *Statistical Matching: Theory and Practice*. Edited by R. M. Groves et al. West Sussex: John Wiley & Sons Ltd.

Destatis (2021) *Fachserie 8 Reihe 7: Verkehr (Verkehrsunfälle)*. Statistisches Bundesamt, Wiesbaden. Available at: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700217004.pdf?__blob=publicationFile.

Erbsmehl, C. *et al.* (2017) 'Analysis and Investigation Method for All Traffic Scenarios (AIMATS)'.

Erbsmehl, C., Lich, T. and Mallada, J. (2019) 'How to Link Accident Data and Road Traffic Measurements to Enable ADAS/AD Simulation?', *Journal of Mechanics Engineering and Automation*, 9(6), pp. 177–184. doi: 10.17265/2159-5275/2019.06.001.

Gabler, S. and Häder, S. (2015) *Stichproben in der Theorie (Version 1.1)*. (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. doi: 10.15465/gesis-sg_009.

Hackeloeer, A. *et al.* (2014) 'Georeferencing : a review of methods and applications', *Annals of GIS*. Taylor & Francis, 20(1), pp. 61–69. doi: 10.1080/19475683.2013.868826.

Hayward, J. C. (1972) 'Near-Miss Determination Through Use Of A Scale of Danger', *Highway Research Board*, pp. 24–35. Available at: http://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf.

Hohm, A. (2022) 'The Second Wave of Automated Driving', *ATZ worldwide 2022*. Springer, 124(5), pp. 64–64. doi: 10.1007/S38311-022-0815-X.

Khan, M. A. *et al.* (2017) 'Unmanned Aerial Vehicle–Based Traffic Analysis: Methodological Framework for Automated Multivehicle Trajectory Extraction', *Transportation Research Record: Journal of the Transportation Research Board*, 2626, pp. 25–33.

Krause, S. (2019) 'Development of a database for the description and analysis of use cases for automated driving', *98th Annual Meeting of the Transportation Research Board (TRB 2019)*.

Kruber, F. *et al.* (2020) 'Vehicle position estimation with aerial imagery from unmanned aerial vehicles', in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 2089–2096.

Laureshyn, A., Svensson, Å. and Hydén, C. (2010) ‘Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation’, *Accident Analysis and Prevention*. Elsevier Ltd, 42(6), pp. 1637–1646. doi: 10.1016/j.aap.2010.03.021.

Lehmann, M. *et al.* (2019) ‘Use of a criticality metric for assessment of critical traffic situations as part of SePIA’, in Bargende, M. et al. (eds) *19. Internationales Stuttgarter Symposium: Automobil- und Motorentechnik*. Wiesbaden: Springer Vieweg, pp. 9–10. doi: 10.1007/978-3-658-16988-6.

Nalic, D. *et al.* (2020) ‘Scenario Based Testing of Automated Driving Systems: A Literature Survey’, in *Proceedings of the FISITA Web Congress 2020*.

Oh, J., Washington, S. and Choi, K. (2004) ‘Development of Accident Prediction Models for Rural Highway Intersections’, *Transportation Research Record: Journal of the Transportation Research Board*, 1897, pp. 18–27.

Orsini, F. *et al.* (2021) ‘A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models’, *Accident Analysis and Prevention*. Elsevier Ltd, 161(May), p. 106382. doi: 10.1016/j.aap.2021.106382.

Ortlepp, J. and Butterwegge, P. (2016) ‘Unfalltypen-Katalog: Leitfaden zur Bestimmung des Unfalltyps’. Berlin: Gesamtverband der Deutschen Versicherungswirtschaft e.V. | Unfallforschung der Versicherer. Available at: https://www.udv.de/resource/blob/80022/89b4d80028aacf8cab649d3a3c6157a0/unfalltypenkatalog-data.pdf.

Pfeiffer, M. S. J. (2006) ‘Statistical and Methodological Foundations of the GIDAS Accident Survey System’, *2nd International Conference on ESAR „Expert Symposium on Accident Research“*, pp. 81–87.

Polders, E. and Brijs, T. (2018) *How to analyse accident causation? A handbook with focus on vulnerable road users*. (Deliverable 6.3). Horizon 2020 EC Project, InDeV. Hasselt, Belgium: Hasselt University.

Rässler, S. (2002) *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York: Springer Science and Business Media New York.

Richards, M. and Ford, N. (2020) *Fundamentals of Software Architecture: An Engineering Approach*. O’Reilly Media.

Scholtes, M. *et al.* (2021) ‘6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment’, *IEEE Access*, 9, pp. 59131–59147. doi: 10.1109/ACCESS.2021.3072739.

Schubert, A., Erbsmehl, C. and Hannawald, L. (2013) ‘Stadardized pre-crash-scenarios in digital format on the basis of the VUFO simulation’.

Siebke, C. *et al.* (2023) ‘Predicting the impact on road safety of an intersection AEB at urban intersections. Using a novel virtual test field for the assessment of conflict prevention between cyclists/pedelecs and cars’, *Transportation Research Interdisciplinary Perspectives*. Elsevier Ltd, 17(June 2022), p. 100728. doi: 10.1016/j.trip.2022.100728.

Tarko, A. P. (2020) *Measuring Road Safety Using Surrogate Events*, *Measuring Road Safety with Surrogate Events*. Amsterdam: Elsevier Inc. doi: 10.1016/C2016-0-00255-3.

United States - Defense Mapping Agency (1987) ‘Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems’, 8350.

Wang, C. *et al.* (2021) ‘A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling’, *Accident Analysis and Prevention*. Elsevier Ltd, 157(April), p. 106157. doi: 10.1016/j.aap.2021.106157.

Yang, X. *et al.* (2021) ‘R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object’, *Proceedings of the AAAI conference on artificial intelligence*, 35(4), pp. 3163–3171.

Zheng, L. and Sayed, T. (2020) ‘A novel approach for real time crash prediction at signalized intersections’, *Transportation Research Part C: Emerging Technologies*. Elsevier, 117, p. 102683. doi: 10.1016/j.trc.2020.102683.

APPENDIX

Table 5.
Excerpt from the official ListDB Codebook (https://w3id.org/listdb/) for video-based traffic observations (VO).

Category	*Variable*	*Values*	*Explanation*	*Example*
Time *(Static)*	**TimeStamp**	YYYYMMDD_HHMMSS	Start of VO	20221210_185120
	RecordingTime [minutes]	Integer	Duration of VO	25
	Weekday	Monday | Tuesday | Weekday | Thursday | Friday | Saturday | Sunday	/	Monday
	PublicHoliday	Yes | No | Unknown	/	Yes
Environment *(Static)*	**Temperature** [°C]	Integer	Outdoor temperature	15
	RoadCondition	Dry | Wet | Icy/Snow-covered | Slippery | Unknown	/	Dry
	RoadSurfaceTemperature [°C]	Integer	/	9
	Sunshine	No | Light | Strong | Not applicable | Unknown	/	No
	Rain			Light
	Fog			Strong
	Snow			No
	Wind			Light
	WindSpeed [km/h]	Integer	/	10
	Light	Day | Night | Not applicable	/	Day
Point of interest (POI) *(Time-based)*	**POI**	Crash | NearCrash | SpecialOperationVehicle | ObstacleOnRoad | VehicleVehicle | VehicleCycle | VehicleBike | VehiclePed | CycleCycle | CycleBike | CyclePed | BikePed | SingleObject | MultiObjects | TurnAround | Other	Detected POIs describing events or interactions saved with timestamp HHMMSS	VehicleCycle, 185230
	BehaviouralError	PassedPriority | RoadUseError | PriorityError | TurningError | SideBySideDrivingError | WrongBehaviourTowardsPedestrianError | OvertakingError | DriveByError | DistanceError | SpeedError | StationaryTrafficError | OtherError	Detected wrong behaviour in relation to a prior detected POI Road user showing the wrong behaviour is	Vehicle, PriorityError
	CauseError	VisualCause | EnvironmentalCause | TechnicalCause	Detected cause of wrong behaviour in relation to a previously detected POI	VisualCause
Special remarks *(Static)*	**Remarks**	String	Free text describing events affecting the traffic situation	Football game between 3 and 6 p.m. in the nearby stadium (stadium not visible in the video)