

IMPACT ANALYSIS OF ACCIDENTS ON THE TRAFFIC FLOW BASED ON MASSIVE FLOATING CAR DATA

J. Golze^{1*}, U. Feuerhake¹, C. Koetsier¹, M. Sester¹

¹ Institute of Cartography and Geoinformatics, Leibniz University Hannover, Germany
(golze, feuerhake, koetsier, sester)@ikg.uni-hannover.de

Commission IV, WG 3

KEY WORDS: trajectory analysis, traffic accidents, floating car data, spatio-temporal impact, traffic flow

ABSTRACT:

The wide usage of GPS-equipped devices enables the mass recording of vehicle movement trajectories describing the movement behavior of the traffic participants. An important aspect of the road traffic is the impact of anomalies, like accidents, on traffic flow. Accidents are especially important as they contribute to the aspects of safety and also influence travel time estimations. In this paper, the impact of accidents is determined based on a massive GPS trajectory and accident dataset. Due to the missing precise date of the accidents in the data set used, first, the date of the accident is estimated based on the speed profile at the accident time. Further, the temporal impact of the accident is estimated using the speed profile of the whole day. The approach is applied in an experiment on a one month subset of the datasets. The results show that more than 72% of the accident dates are identified and the impact on the temporal dimension is approximated. Moreover, it can be seen that accidents during the rush hours and on high frequency road types (e.g. motorways, trunks or primaries) have an increasing effect on the impact duration on the traffic flow.

1. INTRODUCTION

Urban road networks are one of the most critical infrastructures used on a daily basis by many traffic participants in a city. These road networks are handling the traffic load of all participants like motorists, cyclists, pedestrians and more. Most of the time, the traffic flow is regulated automatically (e.g. via traffic signs or traffic lights) to prevent possible congestion or traffic jams. Congestion and traffic jams can lead to delays in the transportation system due to decreased flow rate in the traffic network, and also to increased fuel consumption which itself can cause negative environmental effects (Baykal-Gürsoy et al., 2009). However, there are some repetitive, as well as random, events that can cause a change of the predominant vehicle movement behavior. These events can be, for example, traffic accidents, construction works, weather conditions, temporal markets, parcel delivery, garbage collection, and many others. The effect of these events influences the vehicles movement behavior of other road users and also the traffic load of nearby roads. Traffic accidents are random events and often occur due to disregard or overlooking of traffic rules. Due to the nature of traffic accidents it is (often) impossible to predict their location and point in time (Baykal-Gürsoy et al., 2009). In this paper, we focus on an impact analysis of traffic accidents on the traffic flow. To this end, two large spatio-temporal datasets are combined. These contain on the one hand a collection of accident data (Accident Atlas) and on the other hand floating car trajectory data (FCD). Moreover, the influence of different accident categories on the spatial and temporal extend of the impact on the traffic flow is to be derived. In the Accident Atlas used, due to anonymisation, only year, month and day of the week is given, but not the exact date. In order to quantify the temporal impact of an accident on the surrounding traffic, first, the accident day is to be determined from a set of candidate days. Afterwards, the temporal impact can be approximated by the interval of the

subsequent affected speed profile compared to the daily mean speed value. Moreover, the impacts are analysed with respect to the road type and accident features.

The outline of this work is structured as followed. First, the related work is reviewed in Section 2. The two datasets, FCD and Accident Atlas, are presented in Section 3. Subsequently, the methodology is introduced in Section 4. An experiment and its results are presented in Section 5, their discussion in Section 6. Section 7 concludes this paper and gives an outlook on further aspects, which will be addressed in follow-up work.

2. RELATED WORK

Critical parts of the road network have already been highlighted in previous work based on simulation-based approaches utilizing the network topology by (Taylor, 2008) or road network structural features by (Zhang et al., 2015). (Taylor, 2008) focuses on the identification of critical locations in urban road networks which can be found using an accessibility-based assessment of vulnerabilities of the particular road network. The proposed method reaches out to cover the modelling of the overall impact on the network degradation as well as spatial variations of these impacts on the road network within a study area. (Zhang et al., 2015) are calculating different measures, e.g. average geodesic distance, network centrality and overall clustering coefficient, which represent the characteristics of the structure of a road network. They investigate the aspect of non-motorist accidents (e.g. pedestrians and cyclists) associated with the underlying road network structure.

In contrast to simulated data, real data is used for the detection of e.g. structural dependencies of the underlying road network in (Tempelmeier et al., 2019) and (Tempelmeier et al., 2021). In order to identify those dependencies the authors detect temporally co-occurring outliers, especially congestion, in the usual traffic flow. An additional data driven direction is the work

* Corresponding author

of (Feuerhake et al., 2018), which deals with the detection of similarities of structural road segments of a road network while also providing a supervised learning procedure for the prediction of unknown or missing features. For the structural similarity they use the freely available Open Street Map (OSM) data together with a floating car dataset. While utilizing an unsupervised method, the k-means algorithm, they estimate clusters describing the associated traffic behavior using previously calculated features.

When focusing on the domain of traffic accidents, (Wong and Wong, 2016) investigate the impact of traffic incidents on traffic patterns based on vehicle trajectory data. They are estimating the index measures of the (a) kinematic wave speed, (b) percentage drop in capacity, (c) volume-capacity ratio, (d) incident duration, (e) period of influence and (f) total number of affected vehicles. Using these indices they are able to reconstruct the incident scene and investigate its impact mainly in the time before and after clearance of an incident side. (Pan et al., 2015) are using real-world transportation and incident datasets to make predictions on the spatio-temporal impact of classified traffic incidents. They define traffic incidents as "non-recurring events on road networks, such as accidents, weather hazard or road construction" (Pan et al., 2015). Their primary focus is set on the point in time and the impact on travel time estimations caused by the traffic incidents. The prediction of the speed changes and backlog length is done using a quantitative approach, rather than a qualitative approach, in order to receive a description on the impact of the surrounding areas using numerical measurements. Furthermore, abrupt speed changes and long-lasting propagation of speed changes are predicted using specialized models.

In the work by (Wang et al., 2016), traffic anomalies are detected based on map-matched raw trajectory data. Using a feature-based method, they identify candidate segments by the traffic flow acceleration value. Traffic anomalies are inferred by the density change ratio. The validation of their approach is done by extensive experiments. Non-recurrent traffic anomalies like accidents are detected in the work of (Liu et al., 2016). They are conducting a road similarity analysis on the basis of road traffic flow data derived from taxi GPS trajectories in Beijing. The abnormal score identifying the anomaly is based on historic traffic flow data, as well as the current traffic volume on the respective road. Furthermore, they are investigating the historic traffic flow data of neighbouring roads. Neighbouring roads are defined in terms of their geo-location and the traffic patterns.

3. DATASETS

3.1 Floating Car Data (FCD)

The trajectory dataset¹ has been compiled from the records of vehicle trajectories. It includes vehicles of different types (e.g. cars and trucks). The dataset covers the data from August, September and November 2019 and contains approximately 13M trajectories with 3.28B GPS point measurements in total. The records within the dataset contain the raw speed measurement (in km/h), capture time in Coordinated Universal Time (UTC) and GPS location. The center of the dataset is the state of Lower Saxony, Germany, while all included trajectories, at least, partially intersect this area. This way, the total extension of the dataset is exceeding the German borders (compare Figure 1).

¹ INRIX (<https://inrix.com/>)

For our experiment, however, we restrict the data to the region of Hanover, Lower Saxony. This reduces the remaining number of trajectories to 3.16M and the number of GPS points to approximately 317M (see Figure 1).

Overall, the quality of the trajectories varies in a sense that the sampling rate shows significant differences throughout the whole dataset. This way a part of the dataset's trajectories become partially impractical for the purpose of analyses. Moreover, the speed values partially exceed realistic expectations (e.g. 2075 km/h or more than 300 km/h in an inner-city region) and, according to the provided meta information, could not be used for further processing steps. Therefore, we only use the basic information on each trajectory which is the captured time and the GPS location.



Figure 1. Overview about the processed datasets: The considered accidents in the region of Hanover (Germany) are indicated by red markers. The black lines represent the trajectory data. (basemap: wmflabs OSM B&W)

3.2 Accident Atlas

The records within the accident dataset contain the reported and investigated traffic accidents in Germany from the years 2016 to 2019 with more than 800K entries. The dataset includes the location, hour in UTC+1, day of the week and month of the accidents. Moreover, detailed information about the category, kind and type of the accidents are included (compare Table 1), which describe the characteristics of the accidents, e.g. rear-end or turning collision with seriously or slightly injured participants. Furthermore, the travel mode of those involved in the accidents is included (e.g. car, bicycle or others). The dataset is provided free of charge by the Federal Statistical Offices² within the framework of OpenData. As the focus in this paper is restricted to the FCD data of the city of Hanover within the 3 month time span, the total number of accidents is reduced to 1118 records (see Figure 1).

While this dataset is precise with respect to the accident location, it provides only rough temporal time span for each accident instead of an accurate date and time period. This way an accident has multiple candidate days per month (3 - 5) in which it could have been happened depending on the number

² Statistische Ämter des Bundes und der Länder (www.statistikportal.de)

of weeks in the respective month. The accident time is defined by the combination of the month and the day of the week (in each distinct year) while only giving the full hour as an approximated accident time.

characteristic	description	
category	1	killed participant
	2	seriously injured participant
	3	slightly injured participant
kind	0	Collision with other vehicle starting, stopping or stationary traffic standing
	1	Collision with other vehicle driving ahead or waiting
	2	Collision with other vehicle driving side-ways in the same direction
	3	Collision with another vehicle coming from the opposite direction
	4	Collision with another vehicle turning or crossing the road
	5	Collision between vehicle and pedestrian
	6	Collision with an obstacle on the roadway
	7	Leaving the carriageway to the left
	8	Leaving the carriageway to the right
9	Accident of other type	

Table 1. Overview of the accident characteristics category and kind.

4. METHODOLOGY

The approach we are presenting in this paper consists of the following primary steps shown in Figure 2.

(i) **Determination of Spatio-temporal Region of Interest.** Due to the different spatial and temporal extensions of the two datasets, the first step is the identification of the overlapping area in the spatial and temporal dimension. Therefore, the study area is identified as the region with the largest accumulation of GPS trajectories. In addition, further restrictions are made with regard to accident characteristics as well as the removal of inappropriate sampled trajectories.

(ii) **Spatio-temporal Fusion.** Subsequently, the next step is the assignment of trajectories to the accidents. Due to the incomplete temporal information available for each accident we obtain every trajectory for all of the candidate days and within a certain spatial range, e.g. 500 m, around the accident location. Together with step (i), this steps can be seen as the spatial and temporal fusion of the datasets.

(iii) **Identification of Accident Date.** One major step of this approach is the identification of the accident date from the previous data selection in order to be able to analyse its temporal impact. For this process, we calculate speed profiles for each of the candidate days. The idea is that among all trajectories, those relating to the accident will show a distinct behaviour, e.g. drop in speed. In the temporal dimension, all speed measurements are aggregated to 15 minute intervals using the mean as the representative value (see Equation 1).

$$v_{bin_x} = \frac{1}{n} \sum_{i=1}^n v_i, \quad (1)$$

where v_{bin_x} = mean speed in interval bin_x
 n = total number of speed values
 v_i = i -th speed value

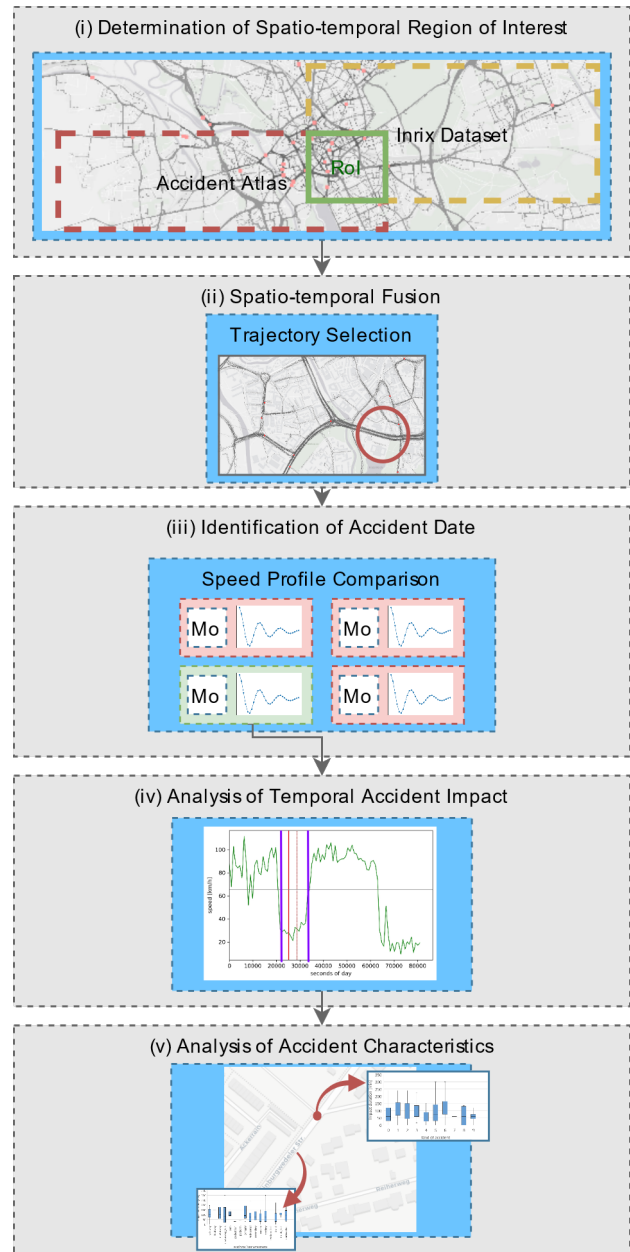


Figure 2. Workflow of our approach. At first the region of interest is identified based on the available data (i). Secondly, the two datasets are fused in the spatial and temporal domain (ii). Afterwards, the actual day of the accident is identified (iii) followed by the temporal impact analysis of the accident (iv) and the investigation of the accident characteristics (v)

Furthermore, the *daily mean* of the speed profile is calculated for each of the candidate days. In order to identify the most probable day for the accident, a *score* is calculated for each candidate day based on the *daily mean* and the respective *accident mean* (see Equation 2). The *accident mean* is calculated solely on the subset of the speed profile within the time interval $[t_{accident} - 1, t_{accident} + 1]$, one hour before till one hour after the time of the accident $t_{accident}$. The *score* describes the reduction of the speed during this time interval in which the accident should have happened.

$$score = daily\ mean - accident\ mean, \quad (2)$$

where $score = score\ value^*$
 $daily\ mean = daily\ mean\ value^*$
 $accident\ mean = accident\ mean\ value^*$
 $*$ for the respective candidate day

The most likely candidate day is identified by the largest $score$ value, which can be seen as the candidate with the largest negative offset compared to the $daily\ mean$. Nevertheless, the best candidate $score_{best}$ needs to be validated respect to the remaining $score$ values. The most probable candidate day is only approved if there is a significant difference of $score_{best}$ and the remaining $score$ values. To this end, a significance threshold Θ_{sig} is introduced.

(iv) **Analysis of Temporal Accident Impact.** Once the accident day has been identified and approved, the complete time series for the accident day is analyzed. Therefore, the time-related impairment (reduction) of the average speed of the aggregated trajectories near the spatial position of the accident is analyzed. As it is already known, from the previous step, that there is an impact, the goal now is to estimate the impacting time span or interval $[t_{left}, t_{right}]$. Here, first, the time interval $[t_{accident} - 1, t_{accident} + 1]$ of the accident is considered again to determine the time of the largest impact t_{impact} . Based on t_{impact} a bi-directional search is performed in order to find the points in time t_{left} and t_{right} , where the speed profile is reaching the $daily\ mean$ (compare Figure 3). The total impact duration Δt_{impact} is the absolute difference of the time interval limits t_{left} and t_{right} .

$$\Delta t_{impact} = |t_{left} - t_{right}| \quad (3)$$

where $\Delta t_{impact} = impact\ time\ span$
 $t_{left} = left\ interval\ limit$
 $t_{right} = right\ interval\ limit$

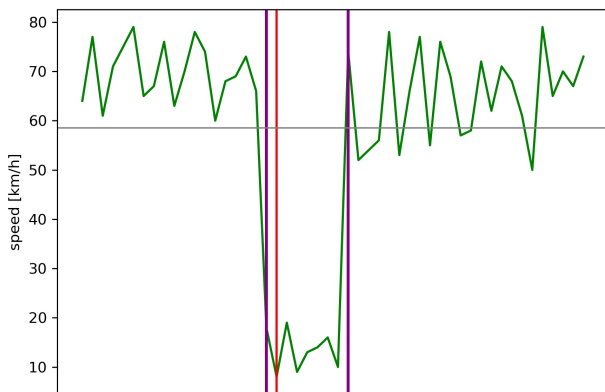


Figure 3. Schematic example of a bi-directional search. The search starts from the point with the largest negative value (red line) and simultaneously expands in both temporal directions until the green line intersects the overall mean (gray). Those points (purple lines) mark the interval of the resulting time span - which is considered as duration of the accident.

(v) **Analysis of Accident Characteristics.** After the accidents have been assigned to space and time, several analyses can be conducted to investigate certain behaviour and dependencies.

5. EXPERIMENT AND RESULTS

For this experiment, the study area is defined to be the region with the largest accumulation of GPS trajectories. This area is the the state Lower Saxony, Germany, while the densification is maximized in the regional district of Hanover. Another limitation is the time constraint imposed by the FCD dataset to the three month period (August, September, November) in the year 2019.

In order to proof the concept of the method presented in Section 4 only a subset of the original data is retrieved as described in step (i). Here, a single month, November 2019, and the weekdays Monday to Friday are selected. This results in a subset of 303 accidents. Figure 4 shows their temporal distribution over the time of the day.

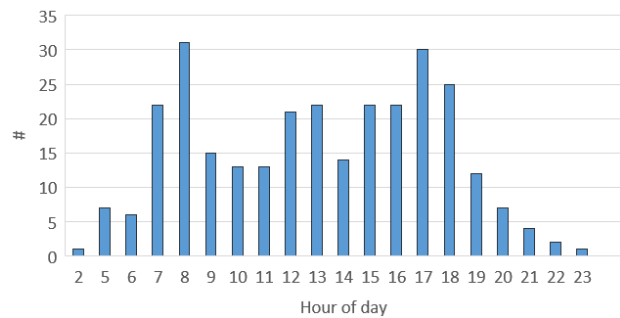


Figure 4. Distribution of the number of accidents over time of the day.

In step (ii), which consists of the spatio-temporal fusion of the datasets, spatial queries are used to intersect the information of both datasets. In table 2 the statistics of the resulting subset is shown. Overall, it can be seen that, on average, a large amount of trajectories is crossing each accident location (3071). But, furthermore, the average number of GPS points of each trajectory is quit low (8).

Attribute	Number
Trajectories	930.750
Points	8.089.192
avg. Points / Trajectory	8
Accidents	303
avg. Trajectories / Accident	3071

Table 2. Overview of the characteristics the fused subsets of both datasets.

Due to the fact that the raw speed measurements in the FCD dataset are corrupted, they need to be recalculated based on the available positional and temporal information. The speed is calculated based on the following Equation 4:

$$v_i = \frac{d(p_i, p_{i-1})}{t_i - t_{i-1}} \quad (4)$$

where $v_i = speed$
 $d(p_i, p_{i-1}) = euclidean\ distance$
 $p_i = position\ of\ i-th\ point$
 $t_i = time\ of\ i-th\ point$

In step (iii) all possible candidate days' speed profiles are calculated for each of the accidents using the intersected FCD data with the recalculated speed values. Furthermore, the daily mean

speed values (Equation 1) and the scores (Equation 2) are determined. The latter are used to identify the most probable accident day by applying the significance threshold Θ_{sig} .

To estimate the required value for Θ_{sig} a parameter study is conducted on the Monday data of the subset considered in the experiment. Table 3 shows the corresponding results. The higher value for Θ_{sig} the less accident dates are identified. However, using a low value, e.g. $\Theta_{sig} = 1km/h$, provides a high count of identified accidents. A visual inspection of those identifications shows that still most of them (26) are plausible.

Θ_{sig}	identifications	validated
1 km/h	42	26
2 km/h	26	16
3 km/h	19	10
4 km/h	15	9
5 km/h	13	7
6 km/h	12	7
7 km/h	10	5
8 km/h	8	5
9 km/h	7	5
10 km/h	5	4

Table 3. Results of the conducted parameter study for the significance threshold Θ_{sig} .

In the following, the results of the temporal impact analysis step (iv) are presented. They are based on the previously determined significance value $\Theta_{sig} = 1km/h$. Our approach identified more than 72% (219) of the accident dates. Nevertheless, due to the lack of groundtruth information this can not be validated. This means, that approx. 28% of the accidents could not be identified as there is no significant change (drop) in the usual traffic flow during the accident time.

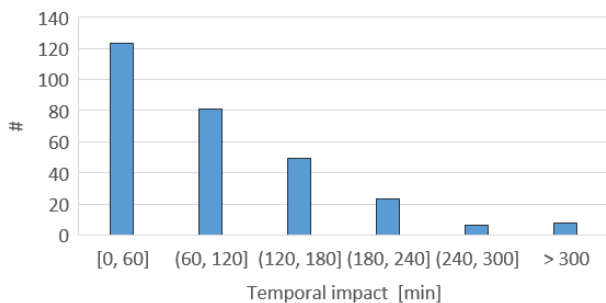


Figure 5. Distribution of the temporal impacts of the 303 accidents.

Figure 5 shows the overall distribution of the temporal impacts by the accidents. Approximately 75 % of the accidents have an impact of a maximum of two hours.

To get a deeper insight into the influencing factors for the temporal impact, further investigations on the individual accident and road characteristics are performed (v). In Figure 6 the dependence on the daytime is pointed out. During the typical daily rush hours, i.e. 8 - 9 am and 4 - 6 pm, the temporal impact is usually longer. This is comprehensible, due to the higher traffic density in those time spans.

A further analysis studies the influence of the road type, which is obtained from OpenStreetMap (OpenStreetMap contributors, 2017). As shown in Figure 7 the temporal impact on larger roads like motorways, trunks or primaries and the corresponding links are more extensive.

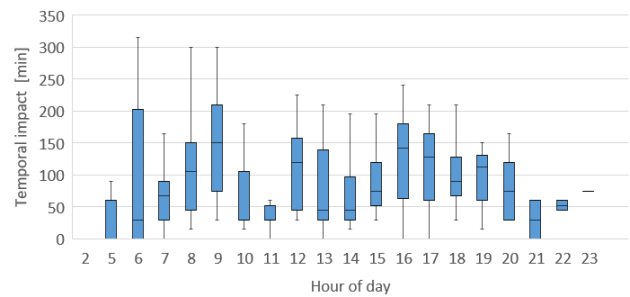


Figure 6. Temporal impact by hour of the day.

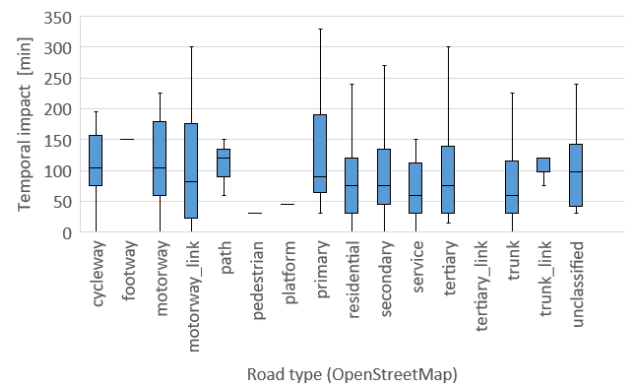


Figure 7. Temporal impact by OSM road type.

Another analysis investigates, if the accident specific features show that the kind of the accident has an effect on the temporal impact (compare Figure 8). More specifically, the kinds 1 (collision with other vehicle driving ahead or waiting) and 6 (collision with an obstacle on the roadway) show an increase of approx. 45 minutes impact duration compared to the others. Furthermore, comparing the different accident categories show a higher impact duration, in case that the accident is of a specific category, here 3, lightly injured (see Figure 9). Unfortunately, not all accident categories are included in the selected subset of the Accident Atlas.

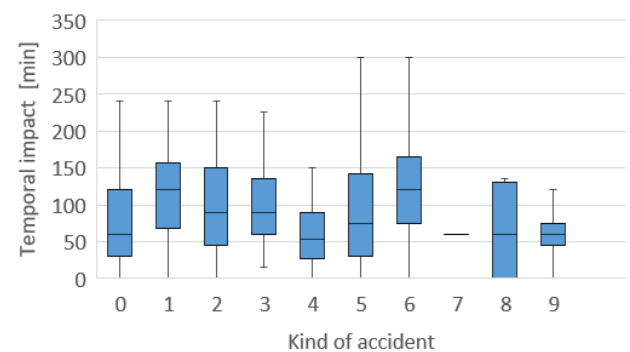


Figure 8. Temporal impact by the kind of the accident.

The process of steps (iii and iv) is illustrated based on the analysis of an example accident (#87191). Its speed profile can be seen in Figure 10. It shows the speed profiles for all four candidate days, together with the accident time interval extracted from the Accident Atlas. The speed profile is expected to show a sharp decrease in the relevant time interval around the time of the accident $[t_{accident} - 1, t_{accident} + 1]$ (dotted red lines). Therefore, from a visual inspection, the most probable candidate in this example, seems to be the day 11. The automatic

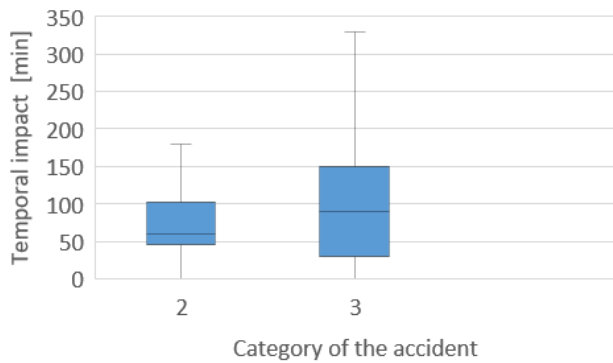


Figure 9. Temporal impact by the category of the accident.

analysis of step (iii) provides the *score* values shown in table 4. The score for day 11 is significantly higher, the difference to the second highest score (day 4) is above the empirically determined $\Theta_{sig} = 1km/h$. Thus, the analysis provides the same day, which is selected for the subsequent impact analysis.

candidate	day	score
1	04	-2.67
2	11	12.36
3	18	-3.23
4	25	-0.61

Table 4. Resulting scores for each candidate day for the example accident #87191.

The exemplary result of the temporal impact analysis step (iv) for the derived accident day is illustrated in Figure 11. The maximum impact is identified in the peak of the speed profile in the accident time interval. Performing the bi-directional search, both points in time are found where the speed profile is reaching the daily mean speed. For this example accident (#87191) the temporal time span of the impact is approximately 150 minutes (purple interval).

6. DISCUSSION

The presented results show that the approach is generally capable of determining the temporal impact of an accident on the traffic flow.

However, due to the lack of ground truth data for the actual accident dates, the estimated dates cannot be validated. Of course, there is the possibility that there are other aspects that can cause a drop in the daily mean speed. But the probability that this happens exactly at the accident time is quite low. Because of that the other aspects are not considered in this work, e.g. construction works, weather conditions, temporal markets, parcel delivery, garbage collection.

Due to its statistical nature, the approach requires a sufficient amount of reasonably sampled trajectory data. The analyses in steps (iii) and (iv) are based on identifying anomalies (in negative direction) to the usual traffic flow, which again is modelled as an average speed profile. Since the latter are determined by the trajectories close to the accident, a sufficient number of those trajectories is required to represent the usual daily, weekly or seasonal changes in the traffic flow. Thus, fewer data would lead to less reliable results. Further, if the sampling rates of the trajectories are too low, the speed values are not able to represent the actual driving behavior close to the accidents. For in-

stance, small accident impacts, which would lead to quite short changes in the traffic flow, cannot be found.

In the current version of this approach the trajectory information used to generate the daily speed profiles is obtained by applying an intersection with a corresponding buffer region around an accident's location. However, the driving direction of the trajectories is not considered. This, of course, leads to the problem that on streets with distinct lanes for each direction, also trajectories are considered which might not be impacted by the accident at all, and would hardly show a change in their speed profiles. Merging them with the actual affected trajectories leads to less significant results in steps (iii) and (iv).

During the estimation of the accident dates the threshold Θ_{sig} is used to determine the significance of the change in the traffic flow. In this work, an absolute value is used. This might lead to the problem that the significance is not the same when looking at roads with different mean speeds. For instance, on a road with lower maximum speed, e.g. 30 km/h, the change in speed might be also lower. Because of that reason a relative value for this parameter, e.g. a change of 10 % of the usual mean speed, might lead to even more stable results.

7. CONCLUSION AND OUTLOOK

In this work an approach for analysing the temporal impact of accidents on the traffic flow is presented. To this end, two different datasets are used. While the first (Accident Atlas) contains the accident related data, the second (FCD) consists of floating car data. For fusing these two heterogeneous datasets overlapping spatial and temporal information is required. Since the accident dataset lacks of precise temporal information in terms of the accidents' dates, a method for generating this missing information has been developed. This method is based on the analysis of the candidate days' speed profiles and aims to detect significant changes in the mean speeds during the accident's time of the day. After identifying the actual accident day the related trajectory information is used to determine changes in the usual traffic flow and, in this way, also to quantify the temporal impact of the accidents. An experiment based on a subset of the available data has been conducted to validate the presented approach. The evaluation of the results, which are mainly depending on the introduced significance threshold Θ_{sig} , shows that the accidents' actual dates have been identified in 219 of the cases (75%). The analysis of their temporal impacts reveals dependencies on factors like the time of the day and the underlying road type.

Although the presented work provides reliable results, their discussion shows that there are some open aspects which will be addressed in follow-up work.

In order to increase the identification rate of the accidents (currently 75%), a deeper investigation into the failure cases is needed. The approach expects that an accident has an influence on the traffic flow, in terms of a drop in mean speed. There might be, however, cases, where this assumption does not hold and therefore an additional model is needed. One aspect is to reject the night time, including the early morning, in order to reduce noise in the data and exclude the time span with the lowest trajectory density. This way, a better representation of the traffic flow can be expected.

As the quality of the results may be negatively influenced due to ignoring the trajectories' driving directions, there will be an

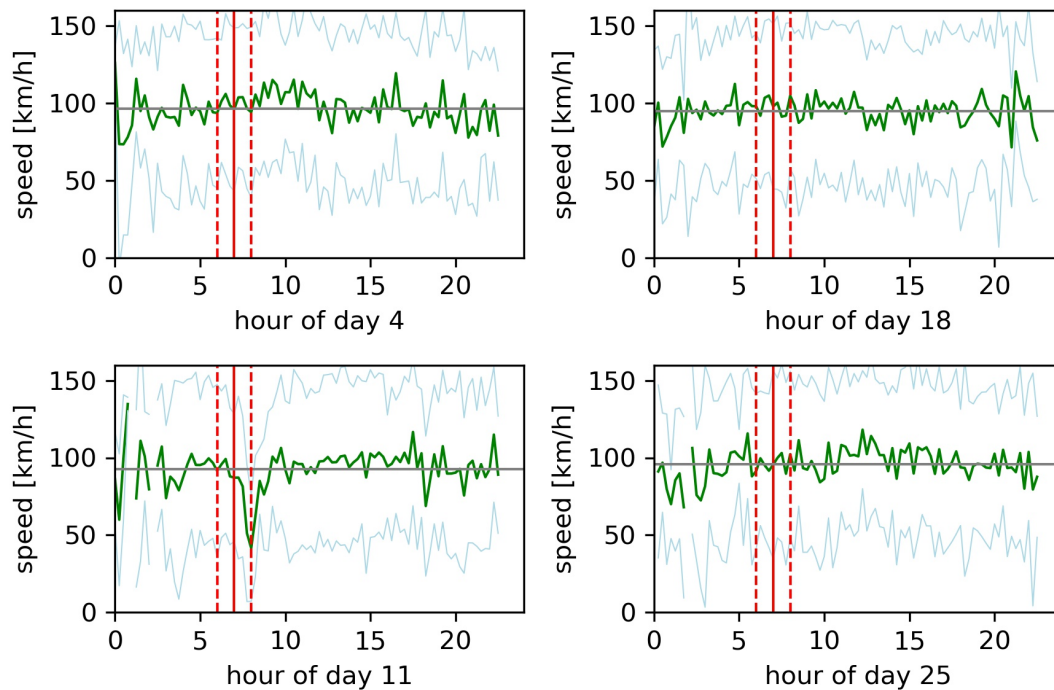


Figure 10. Speed profile (green) and standard deviation (blue) of the four candidate days of the example accident #87191. The approximate accident time plus minus 1 h is highlighted (red).

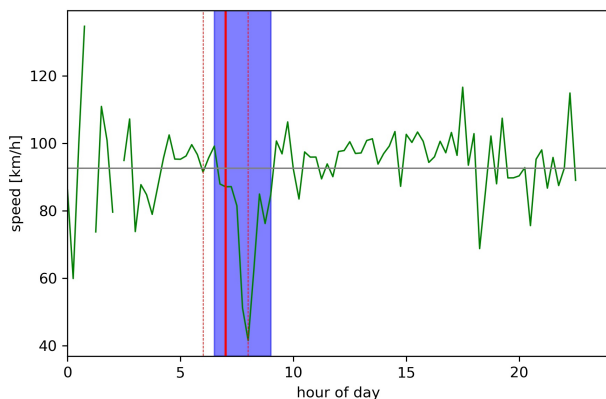


Figure 11. Speed profile (green) of the most probable candidate day of the example accident #87191. The approximate accident time is highlighted (red) with +/- one hour. The estimated temporal impact interval of this accident is highlighted (purple).

adjustment to the data fusion step (ii) in the next version of this approach. For this purpose, the accidents' locations have to be matched to the corresponding lane or driving direction of the street.

In terms of determining the deviations of the traffic flow, a more general approach is the calculation of an overall daily mean speed based on the basis of the complete FCD dataset. This would result in a more robust and generalized *daily mean*.

Furthermore, another aspect will be adjusted within the analysis of accident impact step (iv). This will account for the spatial impact of an accident on the same road segment. Therefore, the spatial impact of an accident will be investigated in a fixed spatial interval to account for accumulation effects. Closely connected to the previous aspect is the subsequent impact on the

traffic flow of the neighbouring road segments. In order to address the effect on the road network, speed profiles of adjacent roads will be investigated to determine the impact propagation of the accidents not only on the same road.

Another aspect, as mentioned in Section 6, is the replacement of the significance value Θ_{sig} with a relative parameter. This way, smaller speed changes on roads with an overall lower maximum speed can result in a better performance.

Moreover, the accident related characteristics can further be investigated with respect to their spatial context, e.g. rural versus urban areas. In general, these investigations will provide insight into the resilience of road networks with respect to accidents.

ACKNOWLEDGEMENTS

This work is funded by the Federal Ministry for Economic Affairs and Energy (BMWi), Germany under the project "d-E-mand" (grant ID 01ME19009B).

REFERENCES

- Baykal-Gürsoy, M., Xiao, W., K., O., 2009. Modeling traffic flow interrupted by incidents. *European Journal of Operational Research*, 195(1), 127-138.
- Feuerhake, U., Wage, O., Sester, M., Tempelmeier, N., Nejd, W., Demidova, E., 2018. Identification of Similarities And Prediction of Unknown Features in an Urban Street Network. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4, 185-192.
- Liu, X., Liu, X., Wang, Y., Pu, J., Zhang, X., 2016. Detecting anomaly in traffic flow from road similarity analysis. B. Cui, N. Zhang, J. Xu, X. Lian, D. Liu (eds), *Web-Age Information*

Management, Springer International Publishing, Cham, 92–104.

OpenStreetMap contributors, 2017. Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>.

Pan, B., Demiryurek, U., Gupta, C., Shahabi, C., 2015. Forecasting spatiotemporal impact of traffic incidents for next-generation navigation systems. *Knowledge and Information Systems*, 45(1), 75–104.

Taylor, M. A. P., 2008. Critical Transport Infrastructure in Urban Areas: Impacts of Traffic Incidents Assessed Using Accessibility-Based Network Vulnerability Analysis. *Growth and Change*, 39(4), 593-616.

Tempelmeier, N., Feuerhake, U., Wage, O., Demidova, E., 2019. St-discovery: Data-driven discovery of structural dependencies in urban road networks. *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '19, Association for Computing Machinery, New York, NY, USA, 488–491.

Tempelmeier, N., Feuerhake, U., Wage, O., Demidova, E., 2021. Mining Topological Dependencies of Recurrent Congestion in Road Networks. *ISPRS International Journal of Geo-Information*, 10(4). <https://www.mdpi.com/2220-9964/10/4/248>.

Wang, Y., Xu, J., Xu, M., Zheng, N., Jiang, J., Kong, K., 2016. A feature-based method for traffic anomaly detection. *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, UrbanGIS '16, Association for Computing Machinery, New York, NY, USA.

Wong, W., Wong, S. C., 2016. Evaluation of the impact of traffic incidents using GPS data. *Proceedings of the Institution of Civil Engineers - Transport*, 169(3), 148–162.

Zhang, Y., Bigham, J., Ragland, D., Chen, X., 2015. Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*, 42(9), 34–47.