THE BALTIC JOURNAL
OF ROAD
AND BRIDGE
ENGINEERING
2024/19(1)

ISSN 1822-427X/eISSN 1822-4288
2024 Volume 19 Issue 1: 88–113
https://doi.org/10.7250/bjrbe.2024-19.629

IDENTIFICATION OF ROAD BLACK SPOTS BASED ON THE SLIDING WINDOW OPTIMIZATION AND SAFETY PERFORMANCE FUNCTION DEVELOPMENT

SHAHIN SHABANI*, JALAL AYOUBINEJAD, NASSIR BARADARAN RAHMANIAN

Department of Civil Engineering, Payame Noor University, Tehran, Iran

Received 7 October 2023; accepted 11 January 2024

Abstract. The sliding window method is a road network screening approach commonly used for identifying black spots. Previous studies have indicated that the selection of window length significantly impacts the black spot identification process. This research proposes a new method that optimizes the sliding window framework by examining its characteristics. The optimization methodology employed in this study is as follows: Firstly, the road is segmented, and for each segment, different scenarios of window lengths are chosen using the Density-Based Spatial Clustering of Applications with Noise algorithm. Next, a Safety Performance Function is developed to calculate the predicted and expected number of crashes, as well as the Potential Safety Improvement, for each window movement across all selected scenarios within the segment. Subsequently, the average differences are calculated using the analysis of variance, and the window length with the lowest dispersion of difference values from the mean is identified as the optimal length for each segment. The case study yielded noteworthy results, indicating that the utilization of the sliding window with optimal lengths led to the identification of 122 high-risk black

* Corresponding author. E-mail: shabani@pnu.ac.ir

Shahin SHABANI (ORCID ID 0000-0001-9131-0597)
Jalal AYOUBINEJAD (ORCID ID 0000-0001-9969-0918)
Nassir BARADARAN RAHMANIAN (ORCID ID 0009-0001-1060-4487)

spot-candidates. These points exhibit a higher crash density, effective length, and greater value in quantitative evaluation tests compared to the results obtained using windows with common fixed lengths.

Keywords: black spots, density-based spatial clustering, network screening, optimization, sliding window, safety performance function.

Introduction

Network screening is part of a modern road safety management system and is a systematic review of hazard data across the network to identify locations with unusually high risks and to identify risk factors associated with a significant number of fatalities and injuries. Network screening yields a list of the locations with the highest risks and a list of major contributing risk factors (HSM, 2010). Although Road network screening models and methods have greatly improved recently, some barriers remain, specifically related to data quality, such as accurate crash location, which is mainly used to integrate crash data with other databases (Bonera et al., 2022). Currently, there are both methodological and practical barriers that together preclude the use of theoretically sound approaches for network screening as part of a road traffic safety management process. Methodological barriers include, among others, lack of a comprehensive framework for corridor-level network screening. In practice, screening and analysis/ranking at the corridor level for the entire network are required for various reasons, including the need to provide as homogeneous infrastructure conditions as possible across the roadway network to meet drivers' expectations (Veeramisti et al., 2020). Network screening techniques are widely used to identify road hot spots, and most of the research in this regard has focused on identifying road segments that are of concern to automobile collisions. In this regard, a wide range of frameworks for identifying accident hotspots as well-known network screening techniques is discussed. Sliding Window (SW), Peak Searching (PS), Continuous Risk Profile (CRP), and Spatial Clustering (SC) are some of the most commonly used techniques and these network screening methods were compared to each other to determine which one worked best in identifying high-risk points (Kwon et al., 2013; Green, 2018). A widely used technique for network screening is the sliding window approach. This method involves moving a fixed-length window along the road in small increments, adhering to a predetermined hot spot selection criterion. One such criterion is a critical crash count/rate threshold (Medury & Grembek, 2016). The objective of this paper is to introduce a novel approach utilizing an optimization framework for the sliding window technique as a road network screening method to identify road black spots. Furthermore, the article highlights significant limitations and weaknesses of the sliding window technique and proposes remedies to enhance the accurate identification of black spots.

1. Description of the sliding window method

In the sliding window method, two parameters need to be determined by the user: the window length and the minimum number of crashes per black spot. The method involves moving a fixed-length window along the entire road network to identify segments that meet the defined crash threshold. When a black spot is identified, the search for additional black spots continues from the next available crash that is not already part of a previously identified black spot. In a slightly modified version of the sliding window approach, the starting point of the window is always a crash. This modification, proposed by Medury and Grembek in 2016, ensures that each window includes at least one crash at its beginning. This adjustment can be beneficial for accurately identifying black spots and capturing crash patterns effectively. By incorporating this modification, the sliding window method can be further refined to enhance its performance in detecting black spots on the road network. In the research conducted by Elvik in 2007, a sliding window with a length of 250 m was employed. This window traversed along the road under observation and marked each location where the specified criterion was met. To be classified as a black spot, a minimum of four crashes within a five-year period was required, with a critical level of statistical certainty set at 5%. The fixed length of the sliding window is also calculated using the Poisson distribution and Equation (1):

$$\alpha = 1 - \sum_{x=0}^{x_{\min}-1} \frac{(\lambda_i T_i L)^X}{X!} e^{-(\lambda_i T_i L)}. \tag{1}$$

Here, α is the probability of observing a crash equal to or greater than the minimum critical value, i.e., 5%. X is the recorded number of crashes, T denotes time, L denotes length and λ denotes the expected number of crashes per kilometer of road. The value of the length of the sliding window will be determined so that for this length, the probability of observing a crash number equal to or greater than the minimum critical value (i.e., 4 crashes) equals 5%. The basic sliding window technique is not applicable to finding black spots located in junctions, because these crashes are usually close to each other, but dispersed along the roads (Szénási, 2016). Recently, Zhang et al. (2018) used quantile and graphic methods to determine the optimal window length and the appropriate increment lengths of the sliding window. Also, the effect of sliding window length and increment on the results of identifying hazardous road segments was discussed. The results showed that: 1) different lengths of the sliding window should correspond to different identification criteria; 2) the longer the sliding window length is, the greater possibility of exaggerating hazardous road segments will have; 3) the bigger the increment length is, the greater possibility of omitting hazardous road segments will have; 4) in respect of determining the optimal length of sliding window and increment by using the quantile method, the optimal combination is that the increment length is half of the window length. To show the effect of window length in identifying black spots, three lengths with sizes of 200, 400, and 800 m were used (Ghadi, 2020). Three different window lengths were chosen to measure the crash frequencies in every increment length (i.e., 100, 200, 400, respectively), in the case of the two selected roads (urban road and motorway). Generally, the consistency of the SW is apparently better for the motorway in comparison with the urban road. The results showed that the use of different segment lengths and increment length seemed to have a significant impact on the performance of the SW method. Recently, a novel network screening method for hot spot identification was proposed based on the optimization framework to maximize the hot spot selection criteria (Lee et al., 2020). The proposed method introduces the capability to dynamically determine the length of each hot spot, taking into account specific constraints. The calculation of the dynamic site length method utilizes Dynamic Programming, which has been demonstrated to effectively find a solution that is close to optimal while maintaining computational feasibility. To evaluate the performance of the proposed method, it was compared to other conventional screening methods, namely, the Sliding Window and Continuous Risk Profile, in terms of their optimal objective value. The findings of the study revealed that the proposed method surpassed the Sliding Window and Continuous Risk Profile approaches. It enables the investigation of a larger number of hotspots with enhanced accuracy and demonstrates improved spatial-temporal consistency. These results were achieved by utilizing the Empirical Bayes (EB) estimate and highlighted the superiority of the proposed method in identifying and analyzing hotspots within the road network. The sliding window method is utilized in numerous scientific disciplines, and several approaches have been suggested to determine the optimal window size. For instance, Baig et al. (2020) employed a deep learning-based adaptive method for window size selection. This method dynamically estimates the sliding window size to effectively capture the latest trends in resource utilization. In this section, we compare the results and findings of previous research to identify various features of the sliding window method, including its advantages and disadvantages/limitations. Several studies have highlighted limitations of this method based on the defined hypotheses. It is worth noting that altering the window length can significantly impact the process of identifying black spots. In other words, modifying the window properties, such as window length and sliding distance may yield different results. Based on the identified cases, we can summarise the advantages of the sliding window method as follows:

- *Simplicity*: The sliding window method is easy to understand and implement;
- *Method Development*: There is potential for further development and refinement of the method;
- *Compatibility*: The sliding window method is compatible with all performance measures used in road network screening methods;
- *Safe System Approach*: The method is known for its application in the safe system approach, making it effective in identifying critical sections and diagnosing safety issues.

The limitations of this method can be described as follows:

- *Sequential Selection*: The sliding window method adopts a first-come-first-serve approach in selecting black spots. This means that the first segment along the road that meets the selection criteria is always chosen, potentially overlooking black spots located further along the road;
- *Sensitivity to Window Parameters*: The length of the window and the increment length have a significant impact on the results of identifying black spots. Altering these window parameters can lead to different outcomes in the search for black spots;
- *Ineffective Length Coverage*: When using a fixed window length, often determined as the black spot length in this method, the end of an identified point may not adequately cover any crashes. As a result, there may be an ineffective length of the road included in the analysis, potentially missing relevant crash data;
- *User-Dependent Selection*: The determination of window length and increment length in the sliding window method relies on the user's engineering judgment or previous experiences. The selection process lacks a systematic utilization of statistical techniques or scientific methods, leading to potential subjectivity and variability in the chosen parameters;
- *Limitation in Intersections*: The sliding window technique is not applicable for identifying black spots located at intersections.

2. Research methodology

The aim of this research was to enhance the sliding window method, which is one of the road network screening techniques, to improve the accuracy of black spot identification. To achieve the aim, an optimized algorithm was developed that could estimate the optimal window length and optimal increment length. The proposed algorithm seeks to enhance the effectiveness and efficiency of the sliding window method in identifying black spots on road networks. The parameters and variables measured in this research include optimal length of the sliding window, optimal increment length, and length of black spots that are determined using the modified sliding window method.

The optimized sliding window methodology is implemented as follows:

1. *Segmentation and Window Length Scenarios*: The target road is segmented, and for each segment, multiple scenarios of window length are selected using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm.
2. *Safety Performance Function (SPF) Development*: An SPF is constructed for the target road segment, estimating the predicted and expected number of crashes for each window movement.
3. *Potential Safety Improvement (PSI) Calculation*: The *PSI* is computed by taking the difference between the predicted and expected crash values for each window movement in all window length scenarios within the segment.
4. *Optimal Length Determination*: The influence of the window length parameter on the mean differences is analysed for each segment. If it is found to be a significant factor, the optimal window length is estimated. If not, the window length is considered less influential in that segment.
5. *Mean Differences and Analysis of Variance*: The mean differences are calculated using analysis of variance, considering all window length scenarios. The optimal window length in each segment is determined based on the length with the lowest dispersion of difference values relative to the mean.
6. *Accuracy Assessment with Coefficient of Variation (CV)*: The estimated SPFs for the optimal window length in each segment are tested using the Coefficient of Variation (CV) as the accuracy index. If the CV value that is calculated using Equation (2) is less than or equal to 0.5 for at least one window movement to the end of the segment, it indicates that the *SPF* has achieved the desired level of accuracy, confirming the estimated optimal length (Kolody et al., 2022; HSM, 2010).

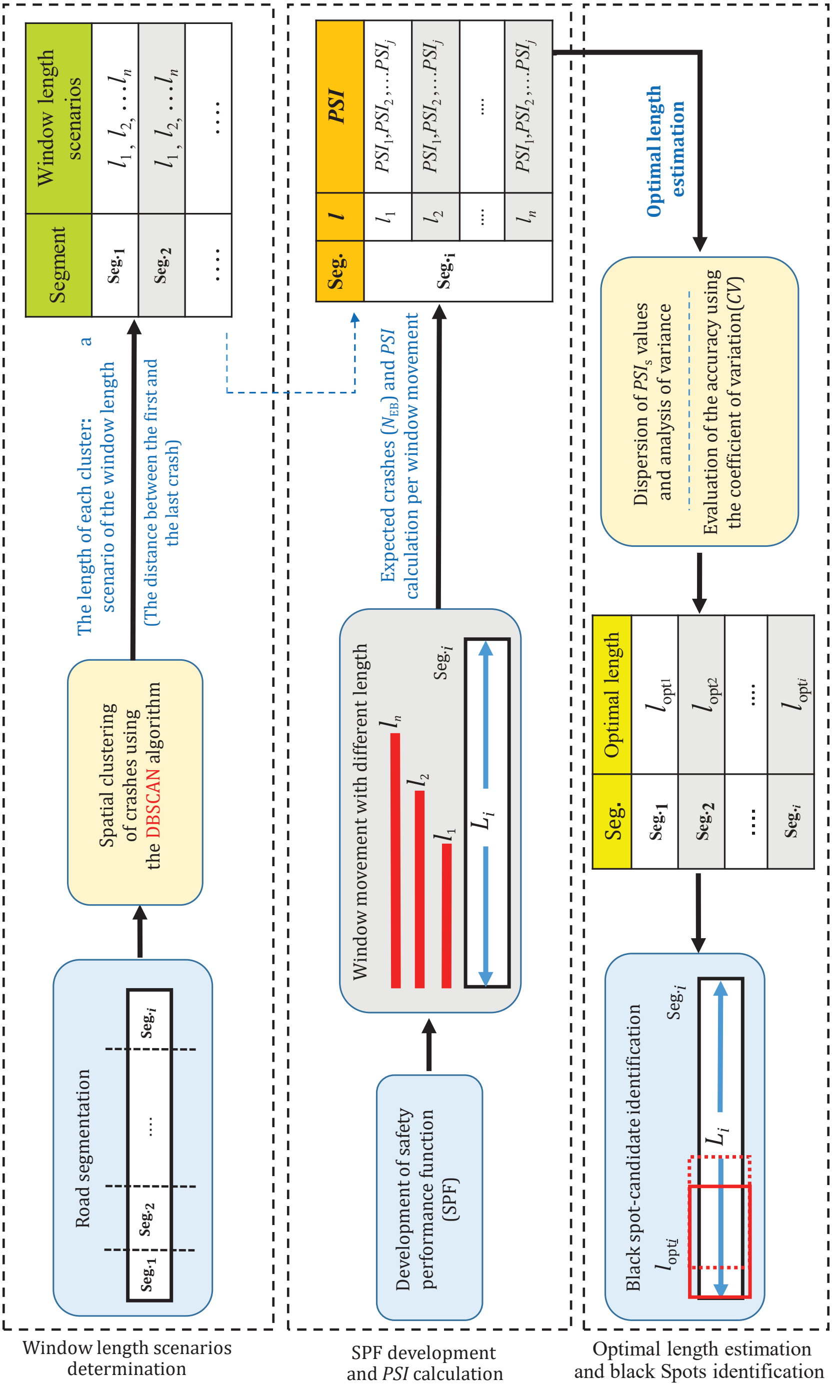

Figure 1. The proposed algorithm structure

$$CV = \frac{\sqrt{\mathrm{Var(SPF)}}}{\mathrm{SPF}}. \tag{2}$$

7. *Black Spot-candidate Identification*: After estimating and verifying the optimal window length, black spot-candidates on the road are identified. The window with the optimal length is moved using Python coding in ArcGIS software along the road. A dynamic-length window is employed, moving from the beginning to the end of each segment to identify points that meet the critical crash threshold criterion.
8. *Black Spot Criteria*: A black spot-candidate is identified when it starts and ends with at least one crash and its size does not exceed the length of the window determined by the proposed algorithm. In case multiple overlapping windows have the same maximum number of crashes, priority is given to the window with a shorter length, based on the start and end index, to be selected as the black spot.
9. *Dynamic Increment Length*: The increment length of the sliding window is also dynamic, depending on the position of crashes recorded along the road. The window is moved in a way that aligns the beginning of each window with the position of the next crash during the movement.

By implementing this optimized sliding window method, black spots can be identified accurately and efficiently, considering the dynamic nature of the window length and increment length.

A case study is conducted to implement the optimization algorithm of the sliding window method specifically for a divided road. The case study focuses on the Neishabur-Sabzevar highway in Khorasan Razavi province, which is a divided highway with comprehensive and accurate crash data spanning at least three years. The research results are analysed based on the values of relevant variables and defined indicators. In the optimization section of the sliding window method, the main variables used include the Potential Safety Improvement (*PSI*) and the dispersion rate of *PSI*s for each window length scenario. The *PSI* is calculated as the difference between the predicted and expected values of crashes for each window movement. The dispersion rate of *PSI*s provides insight into the variability of the results across different window lengths. The proposed algorithm, including the optimization steps, is illustrated in Figure 1, demonstrating the systematic process of determining the optimal window length for the sliding window method. By conducting this case study and following the proposed algorithm, the research aims to provide valuable insights and findings regarding the optimization of the sliding window method specifically for divided roads, using the Neishabur-Sabzevar highway as a representative

Table 1. Summary of the crash data

Year	2018	2019	2020	Total
Number of crash	223	297	342	862

example. The Neishabur-Sabzevar highway, a divided road, has a total length of 194 km. It consists of two lanes in each direction and features multiple intersections and access points. The average daily traffic volume on this highway is approximately 8000 vehicles per year. Table 1 presents the number of injury and death crashes that occurred on the Neishabur-Sabzevar highway from 2018 to 2020.

3. Results

3.1. Road segmentation

Success in developing and applying SPFs, apart from data quality and availability, fundamentally depends on two key factors: the validity of the statistical inferences for the available data and on how well the data can be organised into distinct homogeneous entities (Cafiso et al., 2018). Therefore, the roads should be divided into homogeneous segments. A new segment must be defined whenever any of the following variables, including average daily traffic, lane width, shoulder type, roadside hazard rate, the presence of an intersection, and the beginning or end of a horizontal arch changes. In other words, in the network screening approach, a road should be divided into homogeneous segments based on engineering judgment and using geometrical characteristics. Segment length can significantly affect SPF development, screening, and road network prioritization (Green, 2018). Also, the comparison of different segmentation methods indicates that the Highway Safety Manual (HSM) method is more suitable for separated highways (Ghadi & Török, 2019a). In the segmentation of the Neishabur-Sabzevar road using the HSM method, factors such as lane width and shoulder type remain consistent. Therefore, the segmentation analysis takes into account variables such as traffic volume, intersections and accesses, horizontal curves, and land uses. The summary of the segmentation results for the Neishabur-Sabzevar road is presented in Table 2.

Table 2. The summary of the segmentation results

Number of segments	The shortest length, km	The longest length, km
59	0.505	9.4

3.2. Spatial clustering of crashes using the DBSCAN algorithm and sliding window length scenario determination

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering technique introduced by Ester et al. (1996). This algorithm classifies points based on their density, grouping together points that exhibit similar properties or characteristics. Unlike other clustering methods, DBSCAN is not influenced by the shape of the data and is capable of identifying clusters with intricate and complex shapes. There are two parameters in the algorithm. The first parameter is the radius (ε), and the second is the minimum number of points in a cluster (μ). This algorithm first selects a sample and looks for a neighbour for this point according to the radius. If the algorithm can find a certain minimum number of points in that specific radius, then all those points together belong to a cluster. This continues until all the points are checked. In this algorithm, if a sample has less than the minimum number of samples in its radius, it is identified as an outlier (noise) and is not assigned to any cluster (El Bahi & Zatni, 2018; Deng, 2020). In this study, road crash clustering is performed using a distance measure to identify clusters where the distance between crash locations within each cluster is minimal. Figure 2 provides an example of the DBSCAN structure applied to the clustering of recorded crashes in a road segment sample. The red points represent crashes that have a distance less than the minimum radius (ε) from their neighbouring points. Since the minimum number (μ) requirement is met for these points, two clusters, C1 and C2, are formed. However, despite the distance between the two adjacent blue points being less than ε, they are considered noise points due to not satisfying the minimum number of μ. Furthermore, the blue point at the end of the segment is identified as noise because it is not located in the neighbourhood of the previous points. For this example, we assume μ to be 3. In this research, the length of a cluster is defined as the longitudinal distance between the first and last accident locations.

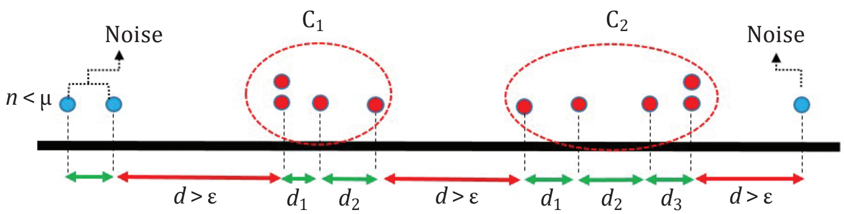

Figure 2. Schematic structure of DBSCAN algorithm for crash clustering

3.3. Choosing the values of DBSCAN algorithm parameters

Prior to performing the clustering, a preliminary analysis was conducted on the crash data of road segments to determine suitable parameter values for the DBSCAN algorithm, namely, the minimum points and radius. The analysis results are presented in Table 3, which includes the number of crashes (n) in each segment, the minimum longitudinal distance between consecutive crashes ($L_{\min}$) in the segment (in meters), the maximum longitudinal distance between consecutive crashes ($L_{\max}$) in the segment (in meters), and the average longitudinal distance between crashes (L_{ave}) in the segment (in meters). For the Neishabur-Sabzevar road, based on the analysis results, the average longitudinal distance between two crashes in all segments was found to be approximately 250 m, and each segment had a minimum of 3 crashes. Therefore, the radius parameter was set to 250 m, and the minimum number of points required to form a cluster was set to 3 crashes. Based on the principles of the DBSCAN algorithm and the crash data analysis of the segments, the algorithm was implemented using Python coding in the ArcGIS software. In each segment, the crashes were clustered linearly based on the longitudinal distance between the accidents that occurred on the road. The summary of the crash clustering results is presented in Table 4.

Table 4. The summary of the crash clustering results

Maximum No. of clusters in a segment	Minimum No. of clusters in a segment	Maximum cluster length, m	Minimum cluster length, m	No. of segments without clusters
8	1	979	34	9

3.4. Estimation of the sliding window length scenarios in each segment

The number of sliding window length scenarios in each segment depends on the number of clusters in that segment. Clustering of crashes on the studied road is performed using the DBSCAN algorithm, based on the longitudinal distance of the crashes index. The length of each cluster, which is the distance between the first crash and the last crash in the cluster, is considered a scenario for the sliding window length in each segment. According to the literature review, scenarios with a length less than 100 m are treated as if they were 100 m in subsequent calculations. Out of the total 59 segments on the Neishabur-Sabzevar road, 9 segments do not have any clusters. This can be attributed to two main reasons.

Table 3. The number of crashes and longitudinal distances between consecutive crashes, m

Seg.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
n	49	44	35	1	5	8	4	9	8	8	4	4	13	2	7
L_{min}	22	21.4	22.3	–	122	25.9	115.5	21.6	75.2	48.2	119.7	187.2	108	107.4	50.5
L_{max}	332	1530	1262	–	440	1137	228.8	582	844.5	1133	1249	630.6	1527	107.4	998.6
L_{ave}	115	236	351.8	–	229.4	360.4	190.6	205.9	373.3	427.6	497.6	483.5	550.8	107.4	329.6
Seg.	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
n	10	16	28	47	20	16	23	26	33	4	5	–	–	–	1
L_{min}	76.7	37.2	28.6	20.4	25.2	22.2	29.9	29.6	25.4	42.5	141.2	–	–	–	–
L_{max}	1382	680	1213	1101	652.1	326.3	1420	1150	669.9	294.2	531.9	–	–	–	–
L_{ave}	580	217	321	242	151.8	139.5	294.3	278.8	229.6	138.1	398.8	–	–	–	–
Seg.	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
n	1	25	7	18	22	3	10	8	16	16	18	11	13	33	5
L_{min}	–	22.4	33.5	20.3	21	178	37.7	24.4	30.7	47.4	26.1	43.6	49.3	20.4	78.6
L_{max}	–	365	386.3	114	364	196.7	380	258.7	444.4	200.8	619.8	912.3	1182	1310	655.9
L_{ave}	–	152	215.5	59.4	104.6	187.4	158.2	160.2	174.4	93.1	237.6	413.8	351.3	295.6	322.6
Seg.	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	
n	3	33	13	26	5	10	32	12	6	21	13	21	14	17	
L_{min}	20.2	26.1	21.2	22.3	58.8	31.5	20.9	21.8	219.4	29.1	39.1	26.4	35.6	21.5	
L_{max}	34	203	219.8	890	179.1	352.3	833	613.4	309.1	1070	409.2	1453	1117	1868	
L_{ave}	27	229	99.1	211	120.7	150.7	247.3	171.3	284.7	235.2	188.4	244.9	327.4	259.3	

Firstly, some segments have no recorded crashes (Segments 27, 28, and 29). Secondly, some segments do not meet the minimum parameter conditions for clustering in the DBSCAN algorithm, either due to the low number of crashes or the spatial distance between them (Segments 4, 12, 14, 30, 31, and 42). The clustering results reveal that the maximum number of sliding window length scenarios in a segment is 8, while the minimum is 1. The segment with the longest sliding window length is Segment 48, with a length of 979 m, whereas the shortest length is 34 m, which is observed in Segment 39. Table 5 presents the maximum ($l_{\max}$) and minimum ($l_{\min}$) estimated length scenarios, along with the corresponding number of scenarios (*n*), for each segment separately.

3.5. Safety performance function development

SPFs are crash prediction models. They are essentially mathematical equations that relate the number of crashes of different types to site characteristics. These models always include traffic volume but may also include site characteristics such as lane width, shoulder width, radius/degree of horizontal curves, presence of turn lanes, and traffic control (Srinivasan & Bauer, 2013). SPFs can be employed to predict crash counts for different roadway elements. Al-Omari et al. (2021) have developed multiple SPFs for various roadway elements using different classifications, such as functional classification and area type. The

Table 5. The scenario of window length in Neishabur-Sabzevar, m

<table>
<tr><th>Seg.</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th></tr>
<tr><td>l_{max}</td><td>794</td><td>589</td><td>484</td><td>–</td><td rowspan="2">531</td><td rowspan="2">227</td><td rowspan="2">572</td><td rowspan="2">263</td><td rowspan="2">95</td><td rowspan="2">265</td><td rowspan="2">266</td><td>–</td><td>417</td><td>–</td><td rowspan="2">190</td></tr>
<tr><td>l_{min}</td><td>494</td><td>80</td><td>59</td><td>–</td><td>–</td><td>359</td><td>–</td></tr>
<tr><td>*n*</td><td>4</td><td>7</td><td>6</td><td>–</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>–</td><td>2</td><td>–</td><td>1</td></tr>
<tr><th>Seg.</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th></tr>
<tr><td>l_{max}</td><td rowspan="2">308</td><td>631</td><td>253</td><td>944</td><td>511</td><td>737</td><td>668</td><td>515</td><td>399</td><td rowspan="2">133</td><td rowspan="2">161</td><td>–</td><td>–</td><td>–</td><td>–</td></tr>
<tr><td>l_{min}</td><td>210</td><td>163</td><td>49</td><td>243</td><td>64</td><td>364</td><td>79</td><td>86</td><td>–</td><td>–</td><td>–</td><td>–</td></tr>
<tr><td>*n*</td><td>1</td><td>3</td><td>3</td><td>7</td><td>3</td><td>3</td><td>3</td><td>4</td><td>6</td><td>1</td><td>1</td><td>–</td><td>–</td><td>–</td><td>–</td></tr>
<tr><th>Seg.</th><th>31</th><th>32</th><th>33</th><th>34</th><th>35</th><th>36</th><th>37</th><th>38</th><th>39</th><th>40</th><th>41</th><th>42</th><th>43</th><th>44</th><th>45</th></tr>
<tr><td>l_{max}</td><td>–</td><td rowspan="2">559</td><td rowspan="2">302</td><td rowspan="2">915</td><td>960</td><td rowspan="2">417</td><td rowspan="2">766</td><td>265</td><td>270</td><td rowspan="2">918</td><td rowspan="2">535</td><td>–</td><td>114</td><td>664</td><td rowspan="2">255</td></tr>
<tr><td>l_{min}</td><td>–</td><td>203</td><td>37</td><td>34</td><td>–</td><td>112</td><td>172</td></tr>
<tr><td>*n*</td><td>–</td><td>1</td><td>1</td><td>1</td><td>2</td><td>1</td><td>1</td><td>2</td><td>3</td><td>1</td><td>1</td><td>–</td><td>2</td><td>3</td><td>1</td></tr>
<tr><th>Seg.</th><th>46</th><th>47</th><th>48</th><th>49</th><th>50</th><th>51</th><th>52</th><th>53</th><th>54</th><th>55</th><th>56</th><th>57</th><th>58</th><th>59</th><th></th></tr>
<tr><td>l_{max}</td><td rowspan="2">60</td><td>371</td><td rowspan="2">979</td><td>329</td><td rowspan="2">511</td><td>241</td><td>719</td><td>516</td><td rowspan="2">264</td><td>511</td><td rowspan="2">827</td><td>778</td><td>378</td><td>566</td><td></td></tr>
<tr><td>l_{min}</td><td>153</td><td>195</td><td>64</td><td>224</td><td>147</td><td>150</td><td>66</td><td>254</td><td>167</td><td></td></tr>
<tr><td>*n*</td><td>1</td><td>8</td><td>1</td><td>4</td><td>1</td><td>3</td><td>4</td><td>2</td><td>1</td><td>4</td><td>1</td><td>2</td><td>2</td><td>2</td><td></td></tr>
</table>

prediction of crash frequency, based on the Negative Binomial regression model (NB), is expressed as Equation (3). This equation is a function of various influential factors, including traffic, geometric design, and other characteristics (Farid et al., 2018, 2019).

$$N_{\mathrm{SPF}i} = \exp\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i\right), \tag{3}$$

where

$\exp(\varepsilon_i) \sim \Gamma[1, k_i]$;

$N_{\mathrm{SPF}i}$ – a predicted number of crashes;

β – a regression vector of parameter estimation;

X – crash contributing factors.

The term, k_i, is referred to as the over-dispersion parameter, which allows for the NB model to accommodate over-dispersed crash data. As $k_i \to 0$, the NB model reduces to the Poisson model (Lord & Mannering, 2010). When sufficient data are available, it is recommended that users develop jurisdiction-specific SPFs. In addition to better crash forecasts, developing state-specific SPFs can help in network screening and evaluation of engineering treatments at a site or project level (Wali et al., 2018). For the safety performance function development, the SPF model is created for the road based on the parameters of the Observed Crashes Number, Average Annual Daily Traffic (AADT), and Access Density according to Equation (4) and the table of values in each segment (Table 6).

$$Y = \exp\left(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon\right), \tag{4}$$

where

Y – the number of segment crashes in the three years;

X_1 – length of the segment, km;

X_2 – average annual daily traffic (vehicles/day) in the segment;

X_3 – access density in the segment;

β –regression vector of parameter estimation;

ε – Gamma-distributed error term with mean of one and variance of α.

Based on the table of values and the method mentioned in HSM, the road safety performance function as a crash frequency prediction model is obtained as follows:

$$Y = \exp(0.248X_1 + 0.0002X_2 + 0.1618X_3),\ \alpha = 0.3496. \tag{5}$$

3.6. Estimation of the predicted and expected number of crashes

After developing the SPF and estimating the predicted number of crashes, the expected number of crashes is estimated while accounting for possible bias due to Regression to the Mean (RTM). One way to do this is to use the EB method. To implement the EB, the first step is to estimate the weight, W, by the following equation (Srinivasan & Bauer, 2013; Ghadi & Török, 2019b):

$$W = \frac{1}{\left(1 + \alpha N_{SPF}\right)}, \tag{6}$$

where
N_{SPF} – the predicted number of crashes using the SPF;
α – over-dispersion parameter of the SPF.

Table 6. Values of parameters in each segment of Neishabur-Sabzevar

Seg.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	49	44	35	1	5	8	4	9	8	8	4	4	13	2	7
X_1	4.3	9	9.4	0.7	1.5	2.7	0.7	1.4	2.5	3.3	1.6	2.2	7.6	1.7	2.4
X_2	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077	8077
X_3	0.9	1	0.7	0	0.7	0.7	0	0.7	0.8	0.6	0	0.5	0.5	0	0.4
Seg.	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Y	10	16	28	47	20	16	23	26	33	4	5	0	0	0	1
X_1	6.1	4.5	8.5	8.8	2.4	2.7	6.7	6.8	6.1	1	2.2	1.5	1.7	0.5	0.6
X_2	8077	8077	8077	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307
X_3	0.7	0.4	0.6	0.6	1.2	1.1	0.9	0.6	1.3	1	1.8	2	2.9	0	3.2
Seg.	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
Y	1	25	7	18	22	3	10	8	16	16	18	11	13	33	5
X_1	1.2	2.1	1.3	1.1	2	0.6	1.3	1.3	1.6	1.8	3.9	5.8	4.3	8.1	1.7
X_2	7307	8077	8077	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307	7307
X_3	2.5	1.4	0.8	1.8	1	1.7	0.8	0.8	1.2	0.6	0.5	0.7	1.2	0.5	0
Seg.	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	
Y	3	33	13	26	5	10	32	12	6	21	13	21	14	17	
X_1	0.7	7	1.1	5.3	1.1	1.8	7.4	2.9	1.1	4.2	2.2	4.3	4.1	6.2	
X_2	7307	8077	8077	7307	7307	7307	7307	7307	7575	7575	7575	7575	7575	7575	
X_3	0	0.1	1.8	0.8	0.9	1.1	0.5	0	0.9	0.2	2.3	0.5	0.5	0.3	

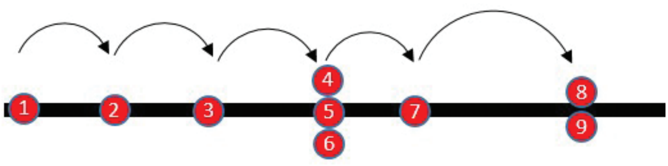

Figure 3. The modified of the sliding window

Then, the expected number of crashes N_{EB} is calculated using Equation (7), where N_{obs} is the observed number of crashes

$$N_{EB} = N_{SPF} W + N_{obs} \left(1 - W\right). \tag{7}$$

In this section, the sliding window is moved along the way using the Geographic Information System (GIS) through Python coding, and using the safety performance function and empirical Bayes method, for each movement of the window with different length scenarios, the predicted number (N_{SPF}) and the expected number (N_{EB}) of crashes are calculated. During the movement of the modified sliding window, as shown in Figure 3, in each segment separately, the starting point is always a crash so that the window along the way to the end is moved on the crash locations, and in each movement, the predicted number and the expected number of crashes are estimated.

3.7. Calculation of the difference between predicted and expected values of crashes (PSI)

In each segment along the road, there is an optimal length for the sliding window. Initially, the difference between the predicted and expected values of crashes is calculated. Subsequently, the dispersion of Potential for Safety Improvement (*PSI*) values from the mean is analysed using the analysis of variance to estimate the optimal window length in each segment. This estimation process is then evaluated. *PSI* represents the potential for safety improvement, which is determined by the difference between the predicted number of crashes (N_{SPF}) and the expected number (N_{EB}) based on Equation (8). *PSI* plays a crucial role in estimating the optimal length of the window in each segment, as highlighted by Kwon et al. (2013).

$$PSI = N_{EB} - N_{SPF}. \tag{8}$$

Certainly, in the process of calculating the predicted number of crashes (N_{SPF}) and the expected number of crashes (N_{EB}) for each sliding window movement with window length scenarios, it is important to note that each window movement may yield a different *PSI* value. The *PSI* value is determined by three key parameters: access density, the number of crashes, and traffic, which vary for each window movement. Hence, the values of the necessary parameters, namely, the predicted number of crashes (N_{SPF}), the expected number of crashes (N_{EB}), and *PSI*, are computed separately for each window movement and each length scenario across all segments.

3.8. Investigation of PSI values dispersion from the mean and the optimal window length estimation

In this section, the influence of the window length parameter on the difference in mean values within each segment is investigated using analysis of variance (Fisher, 1992). The objective is to determine whether the window length parameter significantly affects the mean difference or not. Among all the segments that have multiple window length scenarios, the analysis of variance was conducted in 26 segments to examine the effectiveness of the window length parameter on the variation in mean *PSI* values. The results indicate that in 14 out of the 26 segments, the difference in mean values is influenced by the sliding window length parameter. For these segments, the optimal length is determined by selecting the scenario with a lower dispersion (variance) compared to the mean value. Consequently, each segment has a unique and dynamically optimized length for the sliding window. In the remaining 12 segments, where the difference in mean PSIs is attributed to error factors, no significant difference is observed among the window length scenarios. In such cases, the average of the scenarios is considered to be the final length for identifying black spots. Some segments may have only a single scenario, while others may lack a window length due to either the absence of crashes or the crashes not meeting the minimum conditions for clustering. The results of this analysis, including the estimated optimal length for each segment, are presented in Table 7.

Table 7. The estimated optimal length for each segment, m

Seg.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
l_{opt}	679	579	484	–	531	227	572	263	100	265	266	–
Seg.	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
l_{opt}	388	–	190	308	376	204	308	337	737	379	250	306
Seg.	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
l_{opt}	133	161	–	–	–	–	–	559	302	915	203	417
Seg.	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
l_{opt}	766	100	158	918	535	–	113	208	255	100	247	979
Seg.	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	
l_{opt}	255	511	153	574	331	264	428	827	778	316	566	

3.9. Evaluation of the estimated optimal length

For the optimal lengths estimated in each segment using the coefficient of variation index (CV), the estimated SPFs (predicted number of crashes resulting from the safety performance function) are subjected to testing. The CV value is determined based on Equation (2), and if it is less than or equal to 0.5 in at least one window movement towards the end of the segment, it indicates that the window SPF value has achieved the desired level of accuracy, thus confirming the estimated optimal length. However, if none of the window SPFs meet the desired level of accuracy in all movements, it implies that the estimated length is not optimal. In such cases, the length with the smallest difference in mean among the other defined scenarios should be considered, and this step should be repeated. This evaluation process is conducted solely for the 14 segments where the analysis of variance has revealed the effectiveness of the window length parameter in the difference of mean weights of PSIs. The optimal length is determined as the scenario with a lower dispersion of PSI values compared to the mean. The investigation has demonstrated that, in all segments, there is at least one window movement towards the end where the SPF value of the window meets the desired level of accuracy. Consequently, the estimated optimal lengths are confirmed. The results of this evaluation, including the CV values and the corresponding number of the first window increment where the CV value is equal to or less than 0.5, are presented separately for each segment in Table 8.

Table 8. The evaluation results of the optimal length

Seg.	1	2	3	19	21	22	24	24	35	38	44	52	57	59
$CV \leq 0.5$	0.48	0.5	0.29	0.28	0.07	0.15	0.28	0.28	0.5	0.5	0.5	0.5	0.25	0.37
No. of Increment	1	1	1	1	1	1	2	2	1	1	12	2	1	1

3.10. Identification of black spot-candidates

After estimating the optimal lengths, the identification of black spot-candidates or potential black spots on the road is carried out by moving the window with the determined dynamic lengths in each segment. Points that meet the criterion of a minimum critical value, set as 3 crashes in 3 years in this research, are identified as potential black spots. In this research, only the identification of candidate points was investigated and to determine the final black spots among the identified candidate points, further examination should be done using statistical models. The black spot-candidate is defined as the point with the highest number of crashes among the overlapping windows. In this method, only the windows that overlap with each other are compared based on the number of covered crashes. If multiple overlapping windows have the same number of crashes, the window with the shortest length is selected as the critical window. The length of the black spot-candidate is calculated as the distance between the first and last crash positions within the window. Once a black spot-candidate is detected, the search for other points continues from the next crash that does not overlap with any black spot-candidate. In this research, the possibility of identifying black spot-candidates on the border between two segments is investigated. While moving the window, if there are crashes in the second segment within the optimal length of the first segment, and provided that these points do not independently meet the black spot criteria within the optimal length of the second segment, they are identified as joint black spot-candidates. If the criteria are not met, the starting points of the second segment are recognized separately as black spot-candidates. The results of black spot-candidate identification along the studied road using the sliding window movement with optimal lengths, segmented by segment, are presented in Table 9. The “L_{bs}” column represents the length of the black spot-candidate, and “N_{acc}” represents the number of crashes within the black spot-candidate. A total of 122 potential black spots were identified, with the highest number of points (8) observed in Segment 19. The maximum length of a black spot-candidate on the road is 888 m, while the minimum length is 30 m. The average length of potential black spots on the road is 290 m, and the average number of crashes per point is 5.18.

4. Discussion of results

In order to evaluate the proposed method, the results of the optimization algorithm for the sliding window technique are compared with the results obtained using a fixed window length approach. Previous studies suggest considering three fixed window lengths of 300, 500, and 1000 m for the road, and black spot-candidates are identified using each of these lengths. To compare the results, appropriate Key Performance Indicators (KPIs) are utilized. Two indicators are calculated for different scenarios: the average length of the point and the average number of crashes at each point. The ratio of the average number of crashes to the average length of the point is a measure that positively correlates with crash density and effective length. The results of the comparison of these indicators, as shown in Table 10, demonstrate that the proposed algorithm utilizing the window with optimal dynamic length outperforms the fixed window lengths in all three cases. The proposed algorithm shows better performance in terms of the evaluated KPIs. To further evaluate and compare the results obtained from the use of optimal and fixed lengths, two quantitative evaluation tests are employed: the sensitivity test and the specificity test (Elvik, 2008a). These tests are calculated for each movement of the window along the segment. Among the four evaluated modes (dynamic optimal length, and fixed lengths of 500, 300, and 1000 m), the length that achieves the highest values in these tests across all 59 segments indicates better performance. In these tests, sensitivity (T1) and specificity (T2) are determined for each of the four-window length values used in black spot-candidate identification, according to Equations (9) and (10). The length with the highest values in these tests signifies the best performance.

$$\text{Sensitivity} = \frac{\text{number of correct positive}}{\text{total number of positive}}. \tag{9}$$

$$\text{Specificity} = \frac{\text{number of correct negative}}{\text{total number of negative}}. \tag{10}$$

Concept of correct positive and correct negative is defined as follows (Elvik, 2008b):

$$\text{Correct positive: if } E \geq c \text{ and } R \geq c,$$

$$\text{Correct negative: if } E < c \text{ and } R < c,$$

where E is the expected number of crashes, R is the observed crashes, and c denotes a selected critical value (critical value is 3 crashes during 3 years). Correct and false are defined based on the comparison of the expected number of crashes and selected critical value, while positive and negative are defined based on the comparison between the observed number of crashes and the critical value.

Table 9. The results of black spot-candidate identification (The number of crashes within the points and the length of the points, m)

Seg.		**1**						**2**						**3**		
L_{bs}	555	627	667	460	202	553	538	539	393	414	184	330	316	56	244	452
N_{acc}	15	9	8	13	3	8	4	5	5	9	4	6	5	3	3	5
Seg.	**3**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**13**		**15**	**16**		**17**		**18**
L_{bs}	445	477	205	572	238	92	248	243	375	323	177	290	270	247	197	150
N_{acc}	5	4	3	4	7	3	4	3	3	3	4	3	5	4	3	4
Seg.	**18**					**19**						**20**		**21**		**22**
L_{bs}	181	58	105	236	231	44	275	278	223	192	332	307	235	597	420	301
N_{acc}	4	3	5	6	4	3	4	3	8	3	8	4	5	8	5	4
Seg.	**22**			**23**					**24**			**25**	**26**		**32**	
L_{bs}	329	251	150	198	71	218	200	77	97	275	126	120	150	550	304	352
N_{acc}	5	3	4	5	3	3	3	3	4	10	4	3	3	16	6	3
Seg.	**33**	**34**		**35**			**36**	**37**	**38**		**39**		**40**		**41**	
L_{bs}	272	827	121	117	158	182	374	689	33	30	142	98	859	404	481	380
N_{acc}	4	18	7	4	3	6	3	8	3	5	4	4	15	4	9	3
Seg.	**43**				**44**			**45**	**46**			**47**				**48**
L_{bs}	101	104	160	157	156	188	45	232	54	190	138	164	193	232	236	888
N_{acc}	3	3	6	5	4	4	4	3	3	3	3	6	3	4	4	12
Seg.		**49**			**50**	**51**			**52**			**53**		**54**	**55**	
L_{bs}	183	178	192	63	482	58	105	570	410	487	522	306	132	237	229	394
N_{acc}	4	6	3	5	5	3	3	5	8	7	6	5	5	3	3	6
Seg.	**55**	**56**			**57**		**58**		**59**							
L_{bs}	248	758	780	698	509	677	228	269	508	150						
N_{acc}	5	10	3	13	3	4	5	5	12	3						

Table 10. The results of the comparison of the performance indicators

Window length, m	No. of black spot-candidates	Total length of black spot-candidates, m	Total No. of crashes in black spot-candidates	Average length of black spot-candidates, m	Average No. of crashes at each point	KPI (Ratio)
L_{opt}	122	36.35	632	0.29	5.18	17.86
300	130	39	579	0.3	4.45	14.83
500	120	60	636	0.5	5.3	10.6
1000	100	100	731	1	7.31	7.31

The results of these two tests are presented in Table 11. In both T1 and T2 tests, the optimal dynamic length mode has the highest value compared to the other 3 fixed lengths and, therefore, has a better performance in identifying black spot-candidates.

Table 11. The results of evaluation tests

Window length, m	T1 Test	T2 Test
L_{opt}	0.87	0.60
300	0.66	0.48
500	0.80	0.36
1000	0.80	0.28

Conclusion

The weaknesses of the sliding window technique were investigated, and the following aspects were determined:

- The influence of window length and window increment on research results for black spots.
- The possibility of ineffective road length (no crashes) at the end of defining a black spot due to a constant window length.
- Selection of window length and increment length based on engineering judgment rather than statistical techniques.
- Inability to use the sliding window technique to find black spots located at intersections.

To address these weaknesses, an optimization algorithm was developed to measure and estimate four parameters: the optimal length of the sliding window, the optimal window increment, the length of the black spot, and the number of crashes covered.

The optimization section of the sliding window method used the Potential Safety Improvement (*PSI*) and the dispersion rate of *PSI*s in each window length scenario as the main variables. The proposed method selected different window length scenarios for each segment using the DBSCAN algorithm. The safety performance function was developed, and the PSI was calculated for each window movement in all window length scenarios. The optimal window length for each segment was determined as the length with the lowest dispersion of the difference value from the mean.

The results showed that out of 59 segments, 26 segments had multiple sliding window length scenarios. For 14 of these segments, the optimal window length was chosen based on the scenario with the lowest dispersion of the difference value from the average *PSI*. In the remaining 12 segments, where there was no significant difference between the window length scenarios, the average length of the scenarios was considered as the final window length.

The starting point of the window movement in each segment was set as the location of the first crash in that segment. The optimal window increment was determined to be equal to the distance between the crashes, ensuring that the window moved along the crash points until the end.

The evaluation of the optimal length confirmed that in all segments, at least one window movement met the desired level of accuracy in terms of the Safety Performance Function (SPF). Therefore, the estimated optimal lengths were validated.

The proposed method also accounted for the identification of black spots on the border between two segments.

Using the optimal dynamic window lengths, a total of 122 potential black spots with a length of 36.35 km were identified. The results of the optimization algorithm were compared to using fixed window lengths (300, 500, and 1000 m). The findings demonstrated that the points identified using the optimal dynamic window lengths had higher crash density and useful length compared to the fixed length approach.

Evaluation and comparison of the results using two quantitative evaluation tests (T1 and T2) indicated that the optimal dynamic length outperformed the three fixed lengths in terms of identifying black spots candidates and had the highest value.

Overall, the proposed method utilizing optimal dynamic window lengths showed significant improvements in identifying black spots compared to the conventional fixed length approach.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Al-Omari, M., Abdel-Aty, M., & Cai, Q. (2021). Crash analysis and development of safety performance functions for Florida roads in the framework of the context classification system. *Journal of Safety Research*, *79*, 1–13. https://doi.org/10.1016/j.jsr.2021.08.004

Baig, S., Iqbal, W., Berral, J.L., & Carrera, D. (2020). Adaptive sliding windows for improved estimation of data center resource utilization. *Future Generation Computer Systems*, *104*, 212–224. https://doi.org/10.1016/j.future.2019.10.026

Bonera, M., Barabino, B., & Maternini, G. (2022). A straightforward framework for road network screening to Lombardy region. *Sustainability*, *14*(19), Article 12424. https://doi.org/10.3390/su141912424

Cafiso, S., D'Agostino, C., & Persaud, B. (2018). Investigating the influence of segmentation in estimating safety performance functions for roadway sections. *Journal of Traffic and Transportation Engineering*, *5*(2), 129–136. https://doi.org/10.1016/j.jtte.2017.10.001

Deng, D. (2020). DBSCAN clustering algorithm based on density. *7th International Forum on Electrical Engineering and Automation (IFEEA)*, Hefei, China, 949–953. https://doi.org/10.1109/IFEEA51475.2020.00199

El Bahi, H., & Zatni, A. (2018). Document text detection in video frames acquired by a smartphone based on line segment detector and DBSCAN clustering. *Journal of Engineering Science and Technology*, *13*(2). https://www.researchgate.net/publication/322941007

Elvik, R. (2007). *State-of-the-art approaches to road accident black spot management and safety analysis of road networks*. European Commission, National Research Council of Norway, and Institute of Transport Economics. https://www.toi.no/getfile.php/139022-1210254330/Publikasjoner/TØI%20rapporter/2007/883-2007/883-2007-nett.pdf

Elvik, R. (2008a). Comparative analysis of techniques for identifying hazardous road locations. *Transportation Research Record: Journal of the Transportation Research Board*, *2083*(1), 72–75. https://doi.org/10.3141/2083-08

Elvik, R. (2008b). A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis & Prevention*, *40*(6), 1830–1835. https://doi.org/10.1016/j.aap.2008.08.001

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231. https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf

Farid, A., Abdel-Aty, M., & Lee, J. (2018). Transferring and calibrating safety performance functions among multiple states. *Accident Analysis & Prevention*, *117*, 276–287. https://doi.org/10.1016/j.aap.2018.04.024

Farid, A., Abdel-Aty, M. & Lee, J. (2019). Comparative analysis of multiple techniques for developing and transferring safety performance functions. *Accident Analysis & Prevention*, *122*, 85–98. https://doi.org/10.1016/j.aap.2018.09.024

Fisher, R.A. (1992). Statistical methods for research workers. In S. Kotz, & N.L. Johnson (Eds.), *Breakthroughs in Statistics*. *Springer Series in Statistics*. Springer-Verlag, New York, Inc. https://doi.org/10.1007/978-1-4612-4380-9_6

Ghadi, M., & Török, Á. (2019a). A comparative analysis of black spot identification methods and road accident segmentation methods. *Accident Analysis & Prevention*, *128*, 1–7. https://doi.org/10.1016/j.aap.2019.03.002

Ghadi, M., & Török, Á. (2019b). Comparison of different road segmentation methods. *Promet – Traffic and Transportation*, *31*(2), 163–172. https://doi.org/10.7307/ptt.v31i2.2937

Ghadi, M. (2020). *Methods of segmenting and analyzing of road accident data* [Doctoral dissertation, Budapest University of Technology and Economics]. https://repozitorium.omikk.bme.hu/server/api/core/bitstreams/ca41b13b-6b0a-4ede-a9fc-73a1b1dbdb19/content

Green, E. R. (2018). *Segmentation strategies for road safety analysis* [Doctoral dissertations, University of Kentucky]. https://uknowledge.uky.Edu/ce_etds/62

HSM. (2010). *Highway safety manual* (1st ed.). American Association of State Highway and Transportation Officials, Washington, DC.

Kolody, K., Perez-Bravo, D., Zhao, J., & Neuman, T.R. (2022). *Highway safety manual user guide*. Transportation Research Board (NCHRP Project 17-50). https://doi.org/10.17226/26552

Kwon, O., Park, M., Yeo, H., & Chung, K. (2013). Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. *Accident Analysis & Prevention*, *51*, 141–149. https://doi.org/10.1016/j.aap.2012.10.019

Lee, J., Chung, K., Papakonstantinou, I., Kang, S., & Kim, D. (2020). An optimal network screening method of hotspot identification for highway crashes with dynamic site length. *Accident Analysis & Prevention*, *135*, Article 105358. https://doi.org/10.1016/j.aap.2019.105358

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A*: *Policy and Practice*, *44*(5), 291–305. https://doi.org/10.1016/j.tra.2010.02.001

Medury, A., & Grembek, O. (2016). Dynamic programming-based hot spot identification approach for pedestrian crashes. *Accident Analysis & Prevention*, *93*, 198–206. https://doi.org/10.1016/j.aap.2016.04.037

Srinivasan, R., & Bauer, K. (2013). *Safety performance function development guide: developing jurisdiction specific SPFs*. The University of North Carolina Highway Safety Research Center (Report FHWA-SA-14-005). https://safety.fhwa.dot.gov/rsdp/downloads/spf_development_guide_final.pdf

Szénási, S. (2016). Variable sized planar sliding window technique for searching accident hot spots. *16th International Multidisciplinary Scientific GeoConference SGEM2016*, Bulgaria. https://www.researchgate.net/publication/312395090_VARIABLE_SIZED_PLANAR_SLIDING_WINDOW_TECHNIQUE_FOR_SEARCHING_ACCIDENT_HOT_SPOTS

Veeramisti, N., Paz, A., & Baker, J. (2020). A framework for corridor-level traffic safety network screening and its implementation using Business Intelligence. *Safety Science*, *121*, 100–110. https://doi.org/10.1016/j.ssci.2019.08.042

Wali, B., Khattak, A., Waters, J., Chimba, D., & Li, X. (2018). Development of safety performance functions: Incorporating unobserved heterogeneity and functional form analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(30), 9–20. https://doi.org/10.1177/0361198118767409

Zhang, D., Peijie, W., & Xianghai, M. (2018). The optimal sliding window length and increment length for identifying hazardous road segments. *Transportation Research Board 97th Annual Meeting*, Washington, DC, USA.