GIS-Based Road Accident Analysis and Kernel Density Estimation: A Review of Hotspot Identification Methods, Applications, and Research Gaps

Abstract

Road traffic accidents are a major public-safety concern, and identifying accident-prone locations is essential for effective road-safety planning. Geographic Information Systems (GIS) provide an integrated platform for mapping accident locations, analysing spatial patterns, and supporting the prioritisation of road-safety interventions. Kernel Density Estimation (KDE) is widely used to transform individual accident locations into continuous density surfaces, enabling the visual identification of areas with high concentrations of crashes. This review examines the applications of GIS and KDE in road accident analysis, with particular attention to conventional KDE, network KDE, hotspot validation, and integration with statistical and machine-learning techniques. Anderson (2009) demonstrated the use of GIS-based KDE and K-means clustering to identify and characterise injury-accident hotspots in London. Xie and Yan (2008) developed a network-based KDE approach to account for the fact that traffic accidents occur along road networks. Thakali et al. (2015) compared KDE and kriging for crash hotspot identification, while Harirforoush and Bellalite (2019) combined network KDE with crash-rate screening to improve hotspot assessment. These studies show that KDE is useful for exploratory spatial analysis, but a high density of recorded crashes does not necessarily establish a statistically significant hotspot or a high exposure-adjusted risk. The reliability of KDE depends on accident-data quality, bandwidth selection, spatial resolution, road-network representation, and validation procedures. The review concludes that integrating KDE with network analysis, traffic-exposure measures, spatial statistics, and field investigations can provide a stronger evidence base for road-safety planning.

UNDER PEER REVIEW IN IJAR

Keywords: Road accidents; Geographic Information System; Kernel Density Estimation; accident hotspots; network KDE; spatial statistics; road safety.

1. Introduction

Road traffic accidents arise from interactions among human behaviour, vehicle characteristics, road geometry, traffic conditions, and the surrounding environment. Their spatial distribution is often uneven, with certain road sections and intersections experiencing greater concentrations of crashes than others. Identifying these locations is important for prioritising road-safety inspections, engineering improvements, traffic management, and enforcement activities.

Geographic Information Systems (GIS) support road accident analysis by combining spatial data, accident attributes, mapping, and statistical techniques. Accident records can be integrated with road networks, traffic volume, intersection locations, land use, population characteristics, and infrastructure inventories. This enables researchers to examine accident distributions and investigate their spatial relationships with road and environmental characteristics.

Kernel Density Estimation (KDE) is a commonly used technique for identifying concentrations of spatial events. It estimates a continuous density surface from discrete accident locations by assigning greater influence to nearby events and progressively less influence to more distant events. High-density areas can then be visualised as potential accident hotspots. Anderson (2009) applied KDE to injury-related road accidents in London and combined the resulting hotspot areas with environmental information and K-means clustering. The study demonstrated how density mapping could be extended to characterise different types of accident hotspots.

A limitation of conventional KDE is that it generally measures distance across a two-dimensional plane, whereas traffic accidents occur on a road network. Xie and Yan (2008) developed network KDE to estimate accident density within the network space. Subsequent research has investigated combinations of network KDE with spatial statistics and crash-rate screening to improve the identification of potentially hazardous locations (Yamada &Thill, 2007; Xie& Yan, 2008; Harirforoush&Bellalite, 2019).

KDE results also depend on the selected bandwidth, raster cell size, coordinate reference system, and quality of accident-location data. A high-density surface indicates a concentration of recorded accidents under the chosen analytical settings; it does not, by itself, establish a statistically significant hotspot or demonstrate a high accident rate relative to

UNDER PEER REVIEW IN IJAR

traffic exposure. KDE should therefore be interpreted alongside crash severity, traffic exposure, road characteristics, and field evidence.

This review examines the principles and applications of GIS-based accident analysis, the development of conventional and network KDE, the use of complementary hotspot identification methods, and the methodological challenges that affect the reliability of accident-density mapping.

2. Review Scope and Approach

This paper presents a narrative review of selected literature on GIS-based road accident analysis, conventional KDE, network KDE, spatial autocorrelation, and crash-rate screening. The reviewed studies include foundational methodological contributions and research comparing alternative approaches to hotspot identification.

The discussion focuses on four questions:

1. How has GIS-based KDE been applied to identify road accident hotspots?
2. How does network KDE differ from conventional planar KDE?
3. How can density-based hotspot identification be strengthened using spatial statistics and crash-rate analysis?
4. What methodological issues should be addressed in future GIS-based road-safety studies?

The review is a narrative synthesis, not a fully reproducible systematic review. A systematic-review submission would require a documented database search, search strings, publication-period limits, inclusion and exclusion criteria, screening records, and a study-selection flow diagram.

3. Role of GIS in Road Accident Analysis

GIS provides a spatial framework for integrating accident records with information about the transport network and surrounding environment. Accident datasets may contain coordinates, date and time, severity, collision type, and vehicle category. These records can be linked to road-centreline data, junctions, traffic counts, land use, and infrastructure information.

UNDER PEER REVIEW IN IJAR

The major applications of GIS in accident analysis include:

- **Accident mapping:** Displaying the spatial distribution of recorded crashes.
- **Hotspot identification:** Detecting concentrations using density estimation and spatial statistical techniques.
- **Spatial integration:** Combining accidents with traffic volume, road geometry, intersection density, land use, and other explanatory variables.
- **Temporal analysis:** Examining changes by year, month, weekday, or time of day when reliable timestamps are available.
- **Road-safety planning:** Supporting prioritisation of road-safety audits, infrastructure improvements, and enforcement.

The review of GIS applications by *Traffic Collisions in Space: Four Decades of Advancement in Applied GIS* describes the development of event-based approaches, including planar and network-constrained analysis, as well as methods based on road-segment attributes (see the reference by Xie and Yan, 2008, and the broader GIS review literature). GIS is therefore most useful when mapping is combined with appropriate analytical methods and reliable supporting data.

4. Kernel Density Estimation in GIS

4.1 Concept and mathematical formulation

KDE is a non-parametric technique used to estimate the spatial intensity or density of point events. In road-safety research, accident locations are treated as spatial events whose influence is distributed across a surrounding area.

$$f(x) = \frac{1}{nh^2} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right)$$

where:

f(x) = estimated accident density at location x;
n = total number of accident points;
h = bandwidth/search radius;
x_i = location of the ith accident; and

UNDER PEER REVIEW IN IJAR

K = kernel function.

The exact normalisation and units depend on the KDE implementation, kernel definition, weighting method, and output settings. Researchers should therefore report the software, parameter values, and density units used in the analysis.

4.2 Bandwidth and raster resolution

Bandwidth determines the spatial extent over which each accident contributes to the estimated density. A small bandwidth can reveal localised concentrations but may produce fragmented patterns and isolated peaks. A larger bandwidth creates a smoother surface and may reveal broader accident concentrations, although it can merge distinct hotspots.

The output raster cell size determines the spatial resolution at which the density surface is represented. Bandwidth and cell size are separate parameters: bandwidth controls smoothing, whereas cell size controls the spacing of the output estimates.

Xie and Yan (2008) examined the effects of bandwidth, kernel functions, and linear-unit length in network KDE. Their findings highlighted the substantial influence of bandwidth on the smoothness and spatial pattern of estimated accident density. Their study also demonstrated that the representation of the road network can affect the level of local detail in the results.

For robust analysis, researchers should document the bandwidth, kernel function, raster cell size, coordinate reference system, weighting scheme, and density units. Sensitivity analysis using multiple reasonable bandwidths can help determine whether the principal hotspots remain stable.

4.3 Interpretation of density surfaces

KDE surfaces are often classified into low, moderate, high, and very-high density categories. These categories can help communicate spatial patterns to transport planners. However, the resulting classifications depend on the selected method, such as equal intervals, quantiles, or natural breaks.

UNDER PEER REVIEW IN IJAR

A high KDE value indicates a concentration of recorded accidents relative to the estimated density surface. It does not automatically establish that a location is statistically significant or that its accident rate is higher than that of another location after accounting for traffic exposure. KDE should therefore be treated primarily as an exploratory method for identifying candidate hotspots.

5. Applications of KDE in Accident Hotspot Identification

5.1 Conventional planar KDE

Conventional KDE estimates accident density using Euclidean distance across a two-dimensional plane. It is relatively straightforward to implement in GIS and provides an intuitive visual representation of accident concentrations.

Anderson (2009) used five years of police-recorded injury-accident data from London, covering 1999–2003, to develop KDE maps and identify hotspot areas. Environmental attributes were subsequently attached to hotspot cells, and K-means clustering was used to classify similar hotspot groups. The analysis produced five groups and fifteen clusters, illustrating how KDE can be combined with additional techniques to characterise locations with different accident and environmental profiles.

The principal strength of planar KDE is its ability to transform individual accident locations into a continuous density surface. Its main limitation is that Euclidean distance does not necessarily represent road connectivity. Accidents on nearby but disconnected roads can contribute to the same density surface even where the roads are separated by barriers or do not share a direct network connection.

5.2 Network Kernel Density Estimation

Network KDE (NKDE) estimates accident density along the road network rather than across an unrestricted plane. It is particularly useful when the purpose is to identify potentially hazardous road sections rather than broad geographical areas.

Xie and Yan (2008) proposed a network KDE approach in which the road network is represented by linear units of equal network length, known as *lixels*. This allows density to be estimated within network space and accounts for the connectivity of the road system. Their

UNDER PEER REVIEW IN IJAR

case study used accident records from 2005 and a road network in Bowling Green, Kentucky. The authors found that conventional planar KDE could extend density beyond the event network and overestimate density values in locations that did not correspond to the road network.

Network KDE is especially relevant to highways, expressways, and urban road corridors. However, its reliability depends on accurate road-network topology, appropriate network segmentation, and consistent accident geocoding. It can also require more specialised processing than conventional raster-based KDE.

5.3 Integrated KDE and crash-rate screening

A high concentration of accidents does not necessarily mean that a road section is unusually hazardous relative to its traffic exposure. Harirforoush and Bellalite (2019) proposed a two-stage approach for identifying accident hotspots in Sherbrooke, Canada. Network KDE was first used to identify locations with high crash density. The candidate locations were then screened using critical crash rates based on the Highway Safety Manual approach.

Their study used accident data from 2011–2013 and compared different approaches to aggregating the crash data. The results indicated that the merged-data approach performed better according to the study's prediction accuracy index. The authors also emphasised that crash clustering alone does not establish that a location is a true hotspot; additional factors such as exposure, intersection type, and crash variability should be considered.

This integrated approach illustrates an important principle: KDE is useful for screening and visualisation, while exposure-aware crash-rate analysis can provide additional evidence for prioritising locations.

5.4 KDE and clustering methods

KDE can be combined with clustering algorithms to classify accident hotspots according to their spatial and environmental characteristics. Anderson (2009), for example, used K-means clustering after identifying hotspot areas through KDE.

Such approaches can help distinguish locations according to road type, accident severity, surrounding land use, traffic conditions, and infrastructure. However, clusters produced by a

UNDER PEER REVIEW IN IJAR

clustering algorithm should not automatically be interpreted as statistically significant spatial clusters. Density estimation, machine-learning clustering, and statistical hotspot testing address different analytical questions.

6. Comparison of GIS-Based Hotspot Identification Methods

KDE is one of several techniques used in road accident hotspot analysis. Alternative methods can provide complementary information.

Method	Primary purpose	Main advantage	Important limitation
Planar KDE	Estimate continuous accident density	Easy to interpret and map	Sensitive to bandwidth and ignores network connectivity
Network KDE	Estimate accident concentration along roads	Represents road-network structure	Requires reliable network topology
Global Moran's I	Assess overall spatial autocorrelation	Tests broad clustering or dispersion patterns	Does not identify all local hotspots
Local Moran's I	Identify local clusters and spatial outliers	Distinguishes local spatial relationships	Sensitive to spatial weights and multiple testing
Getis–OrdGi*	Identify statistically significant high- or low-value clusters	Provides a statistical hotspot measure	Depends on neighbourhood definition and input variable
Crash-rate analysis	Relate crashes to traffic exposure	Helps distinguish frequency from exposure-related risk	Requires reliable exposure data
Kriging	Estimate a spatially continuous variable	Can model spatial dependence when assumptions are appropriate	May not be the most suitable standalone method for raw event counts
Empirical Bayes methods	Estimate expected crash frequency using statistical models	Can account for random variation in crash counts	Requires suitable safety-performance models and supporting data

UNDER PEER REVIEW IN IJAR

Thakali et al. (2015) compared KDE and kriging for crash hotspot identification using historical crash data from Hennepin County, Minnesota. The comparison demonstrated the importance of evaluating alternative geostatistical approaches rather than assuming that a single technique is suitable for all accident datasets.

Xie and Yan (2008) and subsequent network-space research also highlight the distinction between mapping event concentration and establishing statistically meaningful hotspots. Yamada and Thill (2007) examined network-constrained spatial analysis, while integrated work combining network KDE and local spatial statistics has sought to strengthen the statistical assessment of crash clusters.

The method selected should therefore reflect the research objective, available data, road-network structure, and whether the study requires exploratory mapping, statistical inference, or exposure-adjusted risk assessment.

7. Integration of KDE with Other Analytical Techniques

7.1 Traffic volume and exposure

Traffic volume is an important contextual variable in accident analysis. High accident density may partly reflect high traffic exposure rather than unusually high risk per vehicle or trip. Where reliable traffic counts are available, researchers should compare accident frequency with an appropriate exposure measure, such as vehicle-kilometres travelled or another defensible denominator.

KDE can visualise accident concentrations, while crash-rate analysis and statistical models can assess accident occurrence relative to exposure. These techniques should be interpreted as complementary rather than interchangeable.

7.2 Road geometry and intersections

Road curvature, junction configuration, road width, lane arrangement, access density, and other geometric characteristics can influence crash occurrence. Integrating these variables with KDE maps can help researchers examine the physical context of accident hotspots.

UNDER PEER REVIEW IN IJAR

Intersections are particularly important in urban areas because turning movements, conflicting traffic streams, pedestrian crossings, and traffic-control arrangements can affect accident patterns. However, a spatial overlap between a high KDE value and an intersection does not establish that the intersection design caused the crashes. Further analysis and field verification are required.

7.3 Integration with AHP and weighted overlay

GIS-based KDE can be incorporated into a broader multi-criteria road-safety assessment. For example, accident density may be considered alongside traffic volume, road density, intersection density, population density, and existing signage information.

The Analytic Hierarchy Process (AHP) can derive relative criterion weights through pairwise comparisons, while GIS can standardise and combine the resulting spatial layers. A weighted overlay can then generate a composite index for prioritising locations.

However, the purpose of each layer must be clearly defined. KDE describes the spatial concentration of recorded accidents, whereas traffic volume and road geometry represent other dimensions of the transport environment. Their combination requires justified normalisation, defensible weights, consistency checks, and sensitivity analysis.

Existing signage density also requires careful interpretation. A high sign count does not necessarily demonstrate adequate signage, because signs may be damaged, obscured, poorly placed, redundant, or unsuitable for the road conditions. Field inspection should assess sign type, condition, visibility, placement, and functional adequacy.

7.4 Integration with machine learning

Machine-learning methods can support crash-frequency prediction, severity classification, and the identification of relationships between accidents and explanatory variables. Common approaches include Random Forest, Support Vector Machines, Artificial Neural Networks, gradient-boosting methods, and decision trees.

KDE and machine learning have different roles. KDE describes the spatial concentration of observed events, whereas machine-learning models estimate an outcome from selected input

UNDER PEER REVIEW IN IJAR

variables. KDE surfaces may be used as exploratory predictors, but model development should avoid data leakage and inappropriate spatial validation.

If a model is trained to reproduce a KDE-derived classification, its performance indicates how well it reproduces that constructed target. It does not independently demonstrate that the model can predict future crashes or identify actual safety risk. Independent validation requires an appropriate outcome, such as crashes from a separate time period or independently verified field observations.

8. Limitations and Methodological Challenges

Despite its usefulness, GIS-based KDE has several limitations.

Bandwidth sensitivity: Different search radii can produce different hotspot patterns. An arbitrarily selected bandwidth may not capture both local and regional accident concentrations.

Spatial resolution: Raster cell size affects the detail of the density surface. A fine raster cannot compensate for inaccurate accident coordinates or insufficient data.

Data quality and geocoding: Inaccurate coordinates, duplicate records, incomplete reporting, and inconsistent location descriptions can distort density estimates. Accident records should be checked for spatial accuracy, missing attributes, and temporal consistency.

Lack of exposure adjustment: KDE generally describes event concentration rather than risk per unit of travel. Traffic exposure must be assessed separately.

Road-network representation: Planar KDE can smooth density across barriers or between disconnected roads. Network KDE can reduce this problem where suitable road-network data are available.

Temporal aggregation: Combining multiple years of accidents into one density surface may conceal changes in road design, traffic conditions, construction activity, or enforcement. Time-specific analysis is preferable when sufficient data are available.

UNDER PEER REVIEW IN IJAR

Statistical significance: A high-density region is not automatically a statistically significant hotspot. Appropriate spatial statistical tests or crash-rate screening can provide complementary evidence.

Transferability: Bandwidths, classification thresholds, and model parameters selected for one road network may not be suitable for another. Sensitivity analysis and local validation are necessary.

9. Research Gaps and Future Directions

The reviewed literature indicates several opportunities for further research.

First, future studies should compare planar KDE and network KDE using the same accident dataset and clearly documented parameters. Comparisons should evaluate hotspot stability, spatial overlap, and the ability to identify locations associated with subsequent crashes.

Second, more attention should be given to bandwidth selection and reporting. Researchers can test multiple bandwidths, examine changes in hotspot rankings, and identify locations that remain high priority under alternative parameter settings.

Third, accident density should be integrated with traffic exposure and road characteristics. This can help distinguish locations with high crash counts due to heavy traffic from locations with unusually high crash occurrence relative to exposure.

Fourth, spatial and temporal analyses should be combined where data quality permits. Comparisons before and after road modifications, construction, or traffic-management interventions may help assess whether hotspot patterns change. Such comparisons require appropriate study designs and should not attribute observed changes to a single intervention without supporting evidence.

Fifth, future research should strengthen field-based validation. GIS-derived hotspots can guide inspections of road geometry, signage, lighting, pedestrian facilities, traffic control, and visibility. This is particularly important where accident databases lack detailed infrastructure information.

UNDER PEER REVIEW IN IJAR

Finally, integrated GIS–AHP–machine-learning frameworks should clearly distinguish descriptive mapping, composite-index development, and independent prediction. Transparent reporting of model targets, training and test data, validation procedures, and uncertainty can improve the reliability of road-safety research.

10. Proposed Framework for Future GIS-Based Accident Studies

A practical framework for future urban road-safety research can be organised into the following stages:

1. Collect police-reported accident records, road-network data, traffic counts, road geometry, intersection information, and relevant environmental variables.
2. Clean and geocode the accident records, check coordinate accuracy, remove confirmed duplicates, and document missing data.
3. Conduct exploratory GIS mapping and examine the temporal distribution of crashes.
4. Apply planar KDE and, where feasible, network KDE using documented and justified parameters.
5. Conduct bandwidth and raster-resolution sensitivity analysis.
6. Compare density maps with complementary methods, such as Getis–OrdGi*, Local Moran's I, or crash-rate analysis.
7. Integrate the spatial evidence with road characteristics and traffic exposure.
8. Use AHP or another justified weighting method if a composite prioritisation index is required.
9. Conduct field inspections at selected locations to verify road conditions, signage, visibility, and other potential safety concerns.
10. Evaluate results using independent data or a validation procedure appropriate to the research objective.

This framework treats KDE as one component of a wider evidence-based road-safety process rather than a complete accident-risk model.

11. Conclusion

GIS and Kernel Density Estimation are useful tools for examining the spatial distribution of road accidents and identifying candidate hotspots for further investigation. KDE transforms

UNDER PEER REVIEW IN IJAR

discrete accident locations into continuous density surfaces that support the visual interpretation of crash concentrations. Network KDE provides a road-oriented alternative where connectivity is important, while spatial autocorrelation methods and crash-rate analysis provide complementary information.

Foundational work by Anderson (2009) demonstrated the application of KDE and clustering to characterise accident hotspots. Xie and Yan (2008) extended KDE to network space, while Thakali et al. (2015) compared KDE with kriging. Harirforoush and Bellalite (2019) further demonstrated the value of combining network KDE with crash-rate screening to distinguish high-density locations from locations requiring more substantive safety assessment.

The main limitation of KDE is that its output depends on data quality and analytical parameters and primarily describes spatial concentration rather than exposure-adjusted risk or causality. Reliable hotspot identification therefore requires careful data preparation, justified bandwidth selection, sensitivity analysis, consideration of road-network structure, and appropriate validation.

Integrating KDE with traffic exposure, road geometry, spatial statistics, field inspections, and well-designed predictive models can improve the evidence base for road-safety planning and support the prioritisation of interventions.

References

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots.*Accident Analysis & Prevention, 41*(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014

Harirforoush, H., &Bellalite, L. (2019). A new integrated GIS-based analysis to detect hotspots: A case study of the city of Sherbrooke. *Accident Analysis & Prevention, 130*, 62–74. https://doi.org/10.1016/j.aap.2016.08.015

Thakali, L., Kwon, T. J., & Fu, L. (2015). Identification of crash hotspots using kernel density estimation and kriging methods: A comparison. *Journal of Modern Transportation, 23*, 93–106. https://doi.org/10.1007/s40534-015-0068-0

UNDER PEER REVIEW IN IJAR

Xie, Z., & Yan, J. (2008).Kernel density estimation of traffic accidents in a network space.*Computers, Environment and Urban Systems*, *32*(5), 396–406. https://doi.org/10.1016/j.compenvurbsys.2008.05.001

Yamada, I., &Thill, J.-C.(2007). Local indicators of network-constrained clusters in spatial point patterns.*Geographical Analysis*, *39*(3), 268–292. https://doi.org/10.1111/j.1538-4632.2007.00703.x

Traffic collisions in space: Four decades of advancement in applied GIS.(2015). *Geo-spatial Information Science.*https://doi.org/10.1080/19475683.2015.1085440

UNDER PEER REVIEW IN IJAR