Research Review on a Spatial Analysis Article

Fifi Zhang
School of Mathematical and Computational Sciences, Massey University

158.741: Location Data: Mapping, Analysis and Visualisation

Kristin Stock

May 2nd, 2023

Le, K. G., Liu, P., & Lin, L.-T. (2022). Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: A case study. *International Journal of Crashworthiness*, *27*(2), 543–553. https://doi.org/10.1080/13588265.2020.1826800

1 Introduction

The research review focuses on a spatial analysis journal article entitled "Traffic Accident Hotspot Identification by Integrating Kernel Density Estimation and Spatial Autocorrelation Analysis: A Case Study." (Le et al., 2022a). The objective of this essay is to demonstrate a comprehensive understanding of the research problem and spatial analysis techniques, while concurrently presenting a critical perspective. The essay will be divided into five sections. Firstly, a summary of the main arguments, including an explanation of the research problem, its significance, and how the article addresses it. Secondly, the research problem is discussed in detail, along with an evaluation of the paper's scientific impact on addressing global issues. The third section focuses on the analysis methods used in the article, including an overview of commonly used methods and comments on the author's chosen approach, as well as suggestions for alternative methods that could have been employed. The fourth section explores other domains in which the same analysis methods could be applied, including different data types, research questions, and application areas. Finally, the essay concludes with an opinion on the paper, highlighting its strengths and weaknesses, as well as recommendations for future improvements.

2 Summary of the main arguments

The research problem of this study is the accurate identification of traffic accident (TA) hotspots using two spatial analysis methods, namely Kernel Density Estimation (KDE) and Spatial Autocorrelation Analysis. Identifying TA hotspots is important because the rise of private car ownership has increased the number of accidents, and a consequent escalation of property loss and casualties. Furthermore, identifying TA hotspots can help traffic authorities understand the underlying causes of these accidents and take effective measures to prevent them, thereby reducing the number of TAs.

The literature review section highlights four major gaps in the previous research on this topic. In the first place, instead of using spatial analysis methods, some of the previous studies relied on traditional statistical methods such as Poisson regression, Negative Binomial regression, and Empirical Bayesian method, which are time-consuming and lack visualization. In addition, the statistical significance of the hotspots identified by the KDE method was often ignored, leading to less accurate results. What's more, the datasets used in previous studies were inadequate, covering only one type of road or excluding urban areas and intersections of highways, which could result in less accurate findings. Finally, some vital parameters of the algorithm were inappropriate or ignored, such as the search radius value being too low (only 100m) and the severity index (SI) not being considered.

Therefore, this study aims to fill these gaps and improve upon previous research in four aspects. Firstly, the research applied two integrated spatial analysis methods by assessing the statistical significance of TA hotspots determined by the KDE method. Secondly, the study uses complete datasets that cover all types of roads, with a particular focus on urban areas. Thirdly, the main parameters of the algorithm have been adjusted based on previous research. For instance, the SI has been taken into consideration and the search radius value has been increased to 1000m. Lastly, the study employs a validation process to ensure the accuracy of the results. Overall, this study contributes to the existing literature by providing more accurate and comprehensive insights into TA hotspots.

The research process consists of several key steps. Firstly, a basic TA hotspots density map was created using the GIS-based KDE method, as shown in Figure 1. Secondly, the statistical significance of the hotspots clusters was evaluated using the Spatial Autocorrelation method, named Moran's I, as shown in Figure 2. Thirdly, the hotspot clusters were joined in the KDE map, and the High-High clusters were selected and ranked according to their significance levels using the natural-breaks (Jenks) classification method, as shown in Figure 3. Figure 3 highlights the hotspots identified by the KDE method (blue circles) that are not actually dangerous. Finally, a validation process was conducted using the Getis-Ord Gi statistics and Equivalent Property Damage Only (EPDO) methods to ensure the accuracy of the clusters and ranking.

Figure 1

Map representing TA hotspot locations in Hanoi (2015–2017)

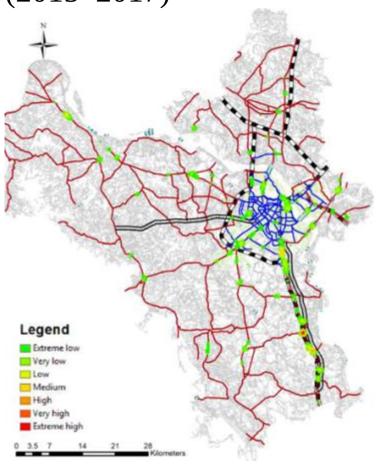


Figure 2

Map shows the significance of clusters with statistical meaning.

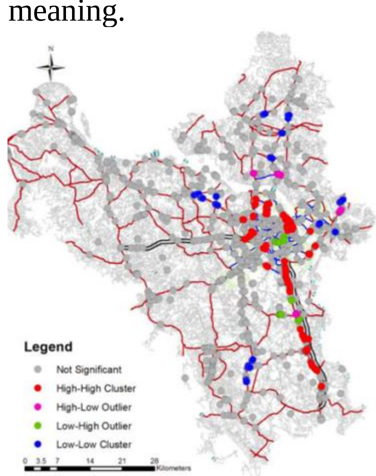
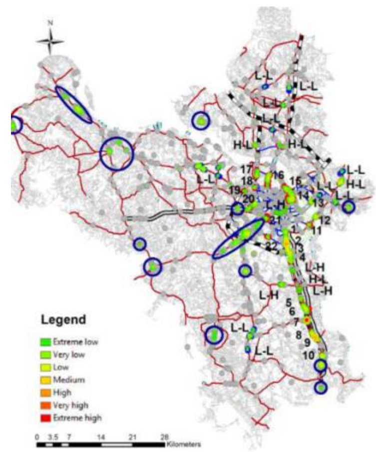


Figure 3

A hotspot priority map



3 Discussion about the research problem

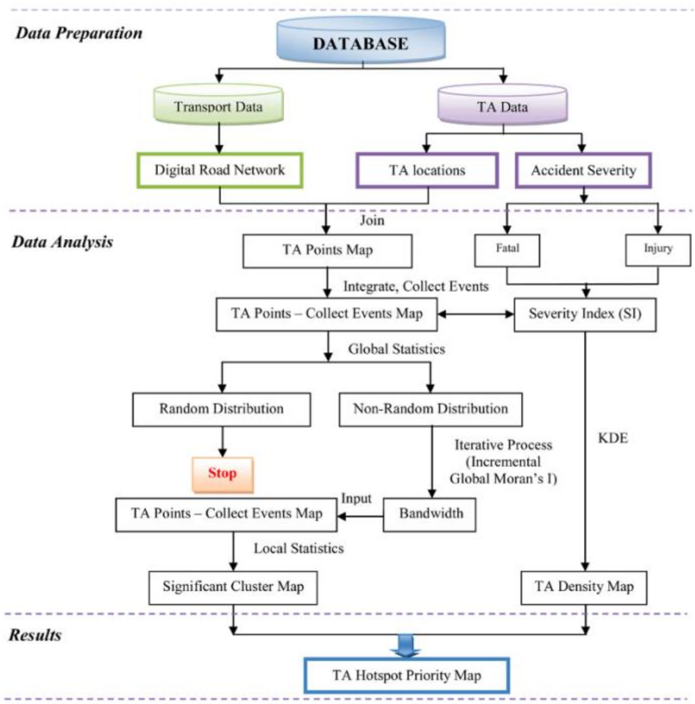
3.1 Literature Review of the Research Problem

Research on TA locations dates back to 1976 when a computer-animated film was used to visualize the geographical map of traffic crashes (Moellering, 1976). Over the past 40 years, research in this area has expanded beyond mapping and can be categorized into two groups: clustering analysis of TA and identification of contributory factors (Yao et al., 2016). With respect to the former, traditional statistical methods and spatial analysis methods are used to investigate TA clustering. Among traditional statistical methods, Bayesian algorithm is not only the best method for determining the TA hotspots (Dereli & Erdogan, 2017), but also for ranking them according to the risks (Dereli & Erdogan, 2017; Miaou & Song, 2005). On the other hand, spatial analysis methods are commonly based on the QGIS (Lloyd, 2010; Satria & Castro, 2016) and include three of the most commonly used methods: Kernel Density Estimation (KDE), Moran’s I statistic tool, and Getis-Ord. KDE is widely used for its effectiveness in the visualization (Plug et al., 2011; Shafabakhsh et al., 2017; Su et al., 2019), while Moran’s I and Getis-Ord can measure the extent of dependence between accident clusters. These methods can be used independently or in combination, such as combining KDE and Clusters (Anderson, 2009; Bíl et al., 2013; Erdogan et al., 2008), GIS and Moran's I (Li & Liang, 2018), NetKDE and Moran's I (Xie & Yan, 2013), NetKDE and Network Screening method (Harirforoush & Bellalite, 2019). In terms of the identification of contributory variables, various kinds of spatial statistical regression algorithms can be used such as the GWR model, RPNB model (Yao et al., 2016).

3.2 Significant Impact on Solving the World's Problem

To determine the impact of a journal article on solving global issues, several factors can be evaluated, including the research problem, methodology, originality and significance of findings, and intended audience. To begin with, identifying TA hotspots is a crucial research problem that warrants attention. Moreover, the methodology is comprehensive and rigorous, including research design, sample size, and analysis methods. The research design is clearly described as shown in Figure 4, and the sample datasets contain complete geographic data and 3-year traffic accident data from the Transport Police Department. Also, four spatial analysis methods, including KDE, Moran's I, Gi* statistics method, and the EPDO method, have been used. Furthermore, the author has also provided a detailed explanation of the theory behind each analysis method used. Furthermore, the findings of this research are significant and original, and it provides an accurate and practical implementation method for quickly locating the TA hotspots. In addition, it was the first study to combine GI* statistics and the EPDO method, contributing to a more accurate result (Le et al., 2022). Lastly, the intended audience of this article, including traffic authorities, road design engineers, and property insurance companies, can easily identify statistically significant hazardous locations based on the detailed steps described in the paper. Overall, due to the nature of the problem researched, the methodology employed, the findings' significance and originality, and work's appropriateness in terms of its intended audience, this article makes a significant contribution to the resolution of the world's traffic accident problem.

Figure 4. The flowchart of the research design



4 Discussion about the analysis methods

The most commonly used analyses to identify TA hotspots are KDE, Spatial Autocorrelation measures, Hot Spot Analysis, and Spatial Regression models. KDE is a density-based method that shows the concentration and intensity of accidents in an area, but it does not consider the spatial relationship between accidents (Le et al., 2022b). Spatial Autocorrelation measures the degree to which accidents tend to be clustered or dispersed across space and whether they are randomly distributed. There are several methods used to measure spatial autocorrelation, including Moran's I, Geary's C, Local Indicators of Spatial Association (LISA), and Getis-Ord General G. Hot Spot Analysis identifies statistically significant clusters of high or low accident rates and considers both the spatial relationship between accidents and the statistical significance of clustering. Spatial Regression models the relationship between accident rates and predictor variables. It allows for the identification of the factors that contribute to accident occurrence and can help in predicting future accident hot spots.

As a result, an integrated method has been used in this research, which is a standard way of analysis. The study collected geographic information data in a shapefile format, and road traffic accident data. At first, a planar KDE was used to identify areas with high accident rates. Then Moran's I was used to evaluate the statistical significance of the hotspot clusters. These methods were chosen because they are suitable for analyzing spatial data and can provide insights into the clustering of road TAs. However, these methods may not be the most appropriate method. A more effective way is combining NetKDE and Hot Spot Analysis methods.

On one hand, NetKDE would likely be more suitable for evaluating road TA hotspots, as road accidents often exhibit complex spatial dependencies that cannot be adequately represented by a simple planar model. NetKDE can consider the spatial dependencies among the accidents, which can be useful in identifying clusters of accidents that are not well represented by a planar model. This can be especially important when dealing with road networks that have non-uniform spatial distributions or where there are multiple interacting factors that contribute to the occurrence of accidents. In addition, NetKDE can be used to incorporate additional data sources, such as traffic volumes, weather conditions, or other relevant factors, into the analysis, which can improve the accuracy of the hotspot identification.

On the other hand, while Moran's I measures the degree of spatial autocorrelation in the data, indicating areas with similar values, it may not be the most appropriate method for identifying TA hotspots. In contrast, the Hot Spot Analysis using the Gi* statistic is specifically designed to identify statistically significant hotspots or cold spots of high or low values in the data. This method is therefore better suited for identifying the specific areas where road traffic accidents are more likely to occur. Overall, combining NetKDE and Hot Spot Analysis can provide a comprehensive approach to analyzing TA, incorporating spatial dependencies and additional data sources while identifying hotspots with statistical significance.

5 Other application domains

5.1 Different data types

Spatial Analysis methods, such as KDE, Spatial Autocorrelation, Hot spot analysis, and Spatial Regression, are applicable to various data types, including point, polygon, line, and raster data. Spatial Analysis enables the examination of geographic data in the form of a set of points with their corresponding coordinates, such as crime or disease outbreak locations. It also enables the analysis of data represented as a set of polygons, such as census tracts or administrative boundaries, as well as data represented as lines or networks, such as road networks or river networks. Moreover, Spatial Analysis can be used to analyze data represented as a grid of cells, such as satellite imagery or elevation data. In general, Spatial Analysis methods can be applied to any type of data that contains a geographic component, such as latitude and longitude coordinates or address information.

5.2 Different research problems

These methods can also be applied to a wide range of research questions related to geography, environmental science, public health, urban planning, and other fields that involve spatial data. For example, where are the areas with the highest concentration of a particular phenomenon, such as crime, pollution, or disease outbreaks? (KDE and Hot spot analysis), are there spatial relationships between different phenomena, such as the correlation between air pollution and respiratory disease rates in different neighborhoods? (Spatial Correlation), what factors are associated with spatial patterns, such as the relationship between land use and crime rates in different areas? (Spatial Regression) and so on. Overall, Spatial Analysis methods provide powerful tools for investigating spatial relationships and patterns, which can help researchers and policymakers make informed decisions and develop effective interventions.

5.3 Different application areas

These methods have a wide range of applications in various fields, including health science, environmental science, urban planning, and emergency management. In health science, Spatial Analysis can help in studying the geographic distribution of diseases(Rex et al., 2020; Thakar, 2020; Xu et al., 2022), identifying targeted patients(Karim et al., 2013; Taal et al., 2022), analyzing DNA diversity (Bertorelle & Barbujani, 1995), and understanding the social and environmental factors that contribute to health disparities, such as the correlation between PM2.5 and lung cancer patients(Zhang & Tripathi, 2018), and the relationship between deprivation and mortality (Lorant et al., 2001). In environmental science, Spatial Analysis can be used to analyze the spatial distribution of different species(Fischer & Getis, 2010, p. 685) such as spatial autocorrelation analysis in plant population(Mathur, 2015) and land birds(Koenig, 1998). In urban planning, Spatial Analysis can help inform decisions about zoning(Dayyani et al., 2019; Sadiq et al., 2022). Lastly, in emergency management, Spatial Analysis can be used to analyze the spatial distribution of hazards such as crime(Andresen, 2011; Anselin et al., 2008; Johansson et al., 2015; Kalinic & Krisp, 2018) and network intrusion(Mashuri et al., 2021). Overall, Spatial Analysis methods provide powerful tools for analyzing spatial data and understanding the complex relationships between geography, environment, and human activities in many different application areas.

6 Opinion about the paper

In my opinion, this research paper has three notable strengths. Firstly, this study offers a significant academic contribution in two ways. It achieves high accuracy in identifying TA hotspots, which is a major accomplishment. Additionally, it saves time in pinpointing implementation locations due to the combined use of spatial analysis methods during both the analysis and validation processes. Secondly, it exhibits a sound and rigorous research methodology with a clear research design, authorized sample data, and multiple uses of analysis methods. Lastly, the article is well-written and organized. For instance, the research question, methodology, and results sections are clearly explained, and the figures are easy to understand.

Despite its strengths, the research paper also exhibits three notable weaknesses. Firstly, the literature review is insufficient and lacks clarity. For instance, important insights such as the most frequent use of spatial autocorrelation methods, like Moran's I and Getis-Ord G*, are not covered, and the earliest research on visualizing traffic crashes using a computer-animated film in 1976 is not mentioned. Secondly, the traffic accident dataset used in the study is relatively small, consisting of only 1,132 data points over three years. Lastly, the introduction section is unclear and fails to explain some key points. For example, while the authors emphasize the importance of evaluating the statistical significance of TA hotspots identified by KDE, they do not provide a clear explanation as to why this is necessary or what the consequences of not doing so may be.

While this is an academic article of high quality, I believe it would be improved if it provided a better understanding of the relationship between accident rates and predictor variables. Therefore, I suggest that the spatial regression method should be used in this research. Identification of the TA hotspots is the first step; however, it is more important to detect the underlying contributing factors. Only by doing this can the relevant departments take effective preventive measures. Overall, the use of Spatial Regression would enhance the quality and impact of this study.

7 Conclusion

The article presents an innovative approach to identifying TA hotspots by integrating KDE and Spatial Autocorrelation Analysis. The paper's scientific impact lies in its potential to inform policy and decision-making processes aimed at reducing traffic accidents and improving transportation safety. Although the methods used in the article have some limitations, they represent a significant improvement over traditional methods and offer a systematic and rigorous methodology foundation. These methods can also be used for different data types, research questions, and application areas. The article is a valuable contribution to the field, highlighting the importance of integrating multiple data sources and analytical techniques to tackle complex problems in transportation safety.

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

Andresen, M. A. (2011). Estimating the probability of local crime clusters: The impact of immediate spatial neighbors. *Journal of Criminal Justice*, *39*(5), 394–404. https://doi.org/10.1016/j.jcrimjus.2011.05.005

Anselin, L., Griffiths, E., & Tita, G. (2008). Crime mapping and hot spot analysis. In *Environmental Criminology and Crime Analysis*. Willan.

Bertorelle, G., & Barbujani, G. (1995). Analysis of DNA diversity by spatial autocorrelation. *Genetics*, *140*(2), 811–819. https://doi.org/10.1093/genetics/140.2.811

Bíl, M., Andrášik, R., & Janoška, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*, *55*, 265–273. https://doi.org/10.1016/j.aap.2013.03.003

Dayyani, L., Pourtaheri, M., Eftekhari, A. R., & Ahmadi, H. (2019). The identification and zoning of areas having rural deteriorated textures in the Tehran province by using KDE and GIS. *Human and Ecological Risk Assessment: An International Journal*, *25*(1–2), 475–504. https://doi.org/10.1080/10807039.2018.1523675

Dereli, M. A., & Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. *Transportation Research Part A: Policy and Practice*, *103*, 106–117. https://doi.org/10.1016/j.tra.2017.05.031

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention*, *40*(1), 174–181. https://doi.org/10.1016/j.aap.2007.05.004

Fischer, M. M., & Getis, A. (Eds.). (2010). *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer. https://doi.org/10.1007/978-3-642-03647-7

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

Johansson, E., Gåhlin, C., & Borg, A. (2015). Crime Hotspots: An Evaluation of the KDE Spatial Mapping Technique. *2015 European Intelligence and Security Informatics Conference*, 69–74. https://doi.org/10.1109/EISIC.2015.22

Kalinic, M., & Krisp, J. M. (2018). *Kernel Density Estimation (KDE) vs. Hot-Spot Analysis—Detecting Criminal Hot Spots in the City of San Francisco*.

Karim, I., Abdul, W., & Kamaruddin, N. (2013). Classification of dyslexic and normal children during resting condition using KDE and MLP. *2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M)*, 1–5. https://doi.org/10.1109/ICT4M.2013.6518886

Koenig, W. D. (1998). Spatial Autocorrelation in California Land Birds. *Conservation Biology*, *12*(3), 612–620. https://doi.org/10.1111/j.1523-1739.1998.97034.x

Le, K. G., Liu, P., & Lin, L.-T. (2022a). Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: A case study. *International Journal of Crashworthiness*, *27*(2), 543–553. https://doi.org/10.1080/13588265.2020.1826800

Le, K. G., Liu, P., & Lin, L.-T. (2022b). Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: A case study. *International Journal of Crashworthiness*, *27*(2), 543–553. https://doi.org/10.1080/13588265.2020.1826800

Li, Y., & Liang, C. (2018). The Analysis of Spatial Pattern and Hotspots of Aviation Accident and Ranking the Potential Risk Airports Based on GIS Platform. *Journal of Advanced Transportation*, *2018*, e4027498. https://doi.org/10.1155/2018/4027498

Lloyd, C. (2010). *Spatial Data Analysis: An Introduction for GIS Users*. OUP Oxford.

Lorant, V., Thomas, I., Deliège, D., & Tonglet, R. (2001). Deprivation and mortality: The implications of spatial autocorrelation for health resources allocation. *Social Science & Medicine*, *53*(12), 1711–1719. https://doi.org/10.1016/S0277-9536(00)00456-1

Mashuri, M., Ahsan, M., Lee, M. H., Prastyo, D. D., & Wibawati. (2021). PCA-based Hotelling's T2 chart with fast minimum covariance determinant (FMCD) estimator and kernel density estimation (KDE) for network intrusion detection. *Computers & Industrial Engineering*, *158*, 107447. https://doi.org/10.1016/j.cie.2021.107447

Mathur, M. (2015). Spatial autocorrelation analysis in plant population: An overview. *Journal of Applied and Natural Science*, *7*(1), Article 1. https://doi.org/10.31018/jans.v7i1.639

Miaou, S.-P., & Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention*, *37*(4), 699–720. https://doi.org/10.1016/j.aap.2005.03.012

Moellering, H. (1976). The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. *Accident Analysis & Prevention*, *8*(4), 215–227. https://doi.org/10.1016/0001-4575(76)90007-5

Plug, C., Xia, J. (Cecilia), & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention*, *43*(6), 1937–1946. https://doi.org/10.1016/j.aap.2011.05.007

Rex, F. E., Borges, C. A. de S., & Käfer, P. S. (2020). Spatial analysis of the COVID-19 distribution pattern in São Paulo State, Brazil. *Ciência & Saúde Coletiva*, *25*, 3377–3384. https://doi.org/10.1590/1413-81232020259.17082020

Sadiq, J., Dent, T., & Wysocki, D. (2022). Flexible and fast estimation of binary merger population distributions with an adaptive kernel density estimator. *Physical Review D*, *105*(12), 123014. https://doi.org/10.1103/PhysRevD.105.123014

Satria, R., & Castro, M. (2016). GIS Tools for Analyzing Accidents and Road Design: A Review. *Transportation Research Procedia*, *18*, 242–247. https://doi.org/10.1016/j.trpro.2016.12.033

Shafabakhsh, G. A., Famili, A., & Bahadori, M. S. (2017). GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *Journal of Traffic and Transportation Engineering (English Edition)*, *4*(3), 290–299. https://doi.org/10.1016/j.jtte.2017.05.005

Su, J.-M., Wang, Y.-M., Chang, C., & Wu, P.-J. (2019). Application of a Geographic Information System to Analyze Traffic Accidents Using Nantou County, Taiwan, as an Example. *Journal of the Indian Society of Remote Sensing*, *47*(1), 101–111. https://doi.org/10.1007/s12524-018-0874-z

Taal, A. T., Blok, D. J., Handito, A., Wibowo, S., Sumarsono, Wardana, A., Pontororing, G., Sari, D. F., van Brakel, W. H., Richardus, J. H., & Prakoeswa, C. R. S. (2022). Determining target populations for leprosy prophylactic interventions: A hotspot analysis in Indonesia. *BMC Infectious Diseases*, *22*(1), 131. https://doi.org/10.1186/s12879-022-07103-0

Thakar, V. (2020). Unfolding Events in Space and Time: Geospatial Insights into COVID-19 Diffusion in Washington State during the Initial Stage of the Outbreak. *ISPRS International Journal of Geo-Information*, *9*(6), Article 6. https://doi.org/10.3390/ijgi9060382

Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. *Journal of Transport Geography*, *31*, 64–71. https://doi.org/10.1016/j.jtrangeo.2013.05.009

Xu, G., Jiang, Y., Wang, S., Qin, K., Ding, J., Liu, Y., & Lu, B. (2022). Spatial disparities of self-reported COVID-19 cases and influencing factors in Wuhan, China. *Sustainable Cities and Society*, *76*, 103485. https://doi.org/10.1016/j.scs.2021.103485

Yao, S., Loo, B. P. Y., & Yang, B. Z. (2016). Traffic collisions in space: Four decades of advancement in applied GIS. *Annals of GIS*, *22*(1), 1–14. https://doi.org/10.1080/19475683.2015.1085440

Zhang, H., & Tripathi, N. K. (2018). Geospatial hot spot analysis of lung cancer patients correlated to fine particulate matter (PM2.5) and industrial wind in Eastern Thailand. *Journal of Cleaner Production*, *170*, 407–424. https://doi.org/10.1016/j.jclepro.2017.09.185