THE APPLICABILITY OF GEOMATICS TOOLS IN THE SPATIAL ANALYSIS OF ROAD CRASHES: A CASE STUDY OF CRASHES AT ROAD INTERSECTIONS IN CAPE TOWN

S VIEIRA[1]*, S HULL[1] and R BEHRENS[2]**

[1]Geomatics Division, School of Architecture, Planning and Geomatics, University of Cape Town; Email: *vrxsim001@myuct.ac.za; **simon.hull@uct.ac.za
[2]Centre for Transport Studies, Department of Civil Engineering, University of Cape Town; Email: roger.behrens@uct.ac.za

ABSTRACT

Obtaining complete and accurate road crash data, particularly in relation to the spatial distribution of crashes, is an ongoing challenge. Deficiencies in crash data pose significant challenges to road safety initiatives. To address this problem, it is imperative to understand both its scale and where it exists. Through spatio-temporal analysis, this paper analyses Road Traffic Crash (RTC) locations and crash prone sites in the City of Cape Town (CoCT) municipality for the years 2017, 2018, 2019 and 2021. Only RTCs occurring at road intersections have been geospatially analysed to uncover spatial and temporal patterns, as these RTCs were able to be spatially defined. The geospatial findings in the paper, therefore, may not be generalised to explain all RTCs in the city. The spatio-temporal analyses revealed that RTC hotspots tended to occur in the CoCT's central regions (i.e. city centre and surrounds), with fewer crashes in the periphery. However, some high-low RTC outliers also tended to occur at road intersections along the CoCTs peripheries which indicated road intersections where high crash counts occurred. The findings underscore the value of having crash location information.

1. INTRODUCTION

While many countries maintain road crash data systems, overseen by entities like the national police or hospitals, data completeness and accuracy vary (Hazen & Ehiri, 2006). Road Traffic Crash (RTC) data completeness and accuracy are frequently compromised due to prevalent under-reporting. In Low- and Middle-Income Countries (LMICs), for instance, crash data is under-reported by up to 50 percent, mainly due to victims' lack of medical treatment (Nantulya & Reich, 2002; WHO, 2004) and inadequate police surveillance (Sokourenko, Wakim & Bassil, 2019). This under-reporting and resultant data incompleteness underestimates the magnitude of road traffic crashes and undermines their associated economic costs (Younis et al., 2019). Consequently, RTCs receive less attention from political authorities than they should (Jadaan et al., 2018) and thus less countermeasures are adopted. It is crucial to have ample information about a problem for its recognition (Elvik, 2010) and subsequent action.

While South Africa is one of the foremost LMICs contributing to road-related studies (Alimo et al., 2022), Adeniji, Mabuza & Titus (2020) nonetheless highlight research gaps, particularly in spatio-temporal road-related injury analyses. This study contributes to filling this research gap.

This paper aims: (1) to explore the potential of spatio-temporal analysis methods, drawn from geomatics, to improve the spatial understanding of RTC[1] patterns; and (2) to apply these methods to pinpoint crash prone sites within the City of Cape Town (CoCT) municipality.

The paper focuses on RTCs in the years 2017, 2018, 2019 and 2021. The 2020 RTC data, though accessible, was not included due to concerns about COVID-19 lockdown effects impacting atypical spatial and temporal distributions. The CoCT RTC data are captured by city officials (usually police) using road names rather than by longitudinal and latitudinal coordinates. Numerous suburbs share similar road names, and suburb boundaries are indistinct, complicating the spatial analysis of RTC distributions and impacting the identification of areas needing road safety interventions. To enable spatial analysis, the RTCs that are analysed in this paper have been narrowed down to road intersections[2]. The results documented in this paper, therefore, may not be generalised to all RTCs in the city.

2. LITERATURE REVIEW

Spatio-temporal analyses are widely adopted data pattern interrogation methods due to their ability to unearth the underlying trends exhibited by data across facets of space and time. As space and time are fundamental to real-world processes (Atluri, Karpatne & Kumar, 2018) and given that the emergence of technology has facilitated their acquisition (Shekhar et al., 2015), spatio-temporal analyses are relevant across various fields. In road safety studies, spatio-temporal analyses have been used to offer geographical insight, which can be visualised if conducted in a Geographical Information System (Achu et al., 2019; Ma, Huang & Tang, 2021). Additionally, they also enable predictions about a phenomenon's future occurrence. Consequently, subjecting RTC data to spatio-temporal analyses locates and quantifies RTC prone sites while also revealing their temporal distributions. This section reviews several RTC studies that have involved spatial and temporal analyses. The RTC studies are organised by spatial themes based on scope and analysis similarity, and thereafter by their temporal aspects.

Several of the transportation studies that utilised spatio-temporal analyses to investigate RTCs (Osayomi & Areola, 2015; Achu et al., 2019; Cheng, Zu & Lu, 2019; Ma, Huang & Tang, 2021; Tola et al., 2021; Hazaymeh, Almagbile & Alomari, 2022) were interested in determining RTC spatial clusters. The spatio-temporal analyses used in these studies involved spatial autocorrelation, a spatial analysis method that relies on spatial statistics to detect spatial patterns (Li, 1996) either at a global or local level. The choice of whether to perform a global or local spatial autocorrelation depends firstly on the coverage of the RTC dataset under investigation (Boots, 2002) and the scale of the outcomes sought. Osayomi & Areola (2015), Achu et al. (2019), Tola et al. (2021) and Hazaymeh, Almagbile & Alomari (2022) have undertaken global spatial autocorrelation, whereas Ma, Huang & Tang (2021) implemented a local spatial autocorrelation. Cheng, Zu & Lu (2018) used both a global and local spatial autocorrelation to spatially analyse the temporal evolution of RTCs at road

[1] RTCs refer to an "incident, event, collision or crash between two or more vehicles, a vehicle and a train, a vehicle and a cyclist, a vehicle and a pedestrian, a vehicle and an animal, a vehicle and a fixed object, such as a bridge, building, tree, post, etc, or a single vehicle that overturned on or near a public road" (Arrive Alive, 2022). Road traffic crashes, in this paper, extend beyond vehicle collisions, encompassing all incidents within road environments, as not all users depend on motorised transport (Onywera & Blanchard, 2013).

[2] Road intersections, in this paper, may refer to a junction of roads, the horizontal position shared between an overhanging bridge and the road beneath and access points such as petrol station entrances and exits.

intersections in Songling and Wujiang (China) for 2016. In addition to the spatial autocorrelation analysis, these studies also used a hotspot analysis to categorise the RTC clusters into ‘hotspots’ or ‘coldspots’.

Osayomi & Areola (2015) and Tola et al. (2021) used global spatial autocorrelation to examine RTCs in terms of severity. Ma, Huang & Tang (2021) explored RTC severity through a local spatial autocorrelation approach, as indicated by their analysis of RTC neighbourhood distances. Osayomi & Areola (2015) aimed to identify RTC hotspot regions in Nigeria over six years (2002 - 2007), discerning spatial clustering of road traffic deaths in 2002 and 2004, accidents in 2003 and 2004, and injuries in 2002. They found consistency across Nigerian states and over their study period.

Tola et al. (2021) utilised a severity index developed by the Roads and Traffic Authority of New South Wales to determine RTC severity in Ethiopia and a specific region therein. This severity index assigned a value of 3,0 for fatalities; 1,8 for serious injuries; 1,3 for slight injuries and 1,0 for property damage. Tola et al. (2021) ensured that the RTCs for their study area displayed spatial clustering by calculating the maximum distance between RTCs to ensure that meaningful spatial patterns were revealed. This approach was also applied by Hazaymeh, Almagbile & Alomari (2022) who investigated RTC spatial clusters for both spatial and temporal crash attributes for the Irbid Governate in Jordan. Unlike Tola et al. (2021), Hazaymeh, Almagbile & Alomari (2022) opted to use a spatial autocorrelation distance that ensured each RTC had between four and six neighbours following Griffith’s (1996) recommendation. Hazaymeh, Almagbile & Alomari (2022) were interested in determining RTC hotspot road segments. Likewise, Ma, Huang & Tang (2021) and Achu et al. (2019) were also interested in determining RTC hotspot roads rather than RTC hotspot locations.

In the case of the study by Hazaymeh, Almagbile & Alomari (2022), RTC data obtained from the Jordan General Security Bureau required aggregation to road centreline spatial data. Ma, Huang & Tang (2021) achieved RTC hotspot road segments for their study by using a line density analysis. In addition to the line density analysis, Ma, Huang & Tang (2021) implemented two clustering analyses and a two point density spatial analysis for RTCs in Wales in the United Kingdom for 2017 over 22 counties in terms of crash counts and crash severity. While both density and clustering analyses produced similar results – the clustering analyses were found to be associated with greater locational accuracy while the density analyses were functionally simpler and more efficient.

Achu et al. (2019) studied crash attributes and RTC clusters in Thrissur District (India) (2013-2015), using Kerala State Crime Records Bureau data. Initially, they needed to geocode the non-spatial RTC data before conducting spatial analysis. Their spatial autocorrelation results unveiled RTC spatial clustering across the study years, including daytime incidents, heavy vehicle involvement, and state highways. Their hotspot analyses revealed several road segments that transitioned from coldspots to hotspots in specific study years, indicating fluctuations in RTC frequency at particular geographic locations. Significant hotspots occurred for all crash attributes except for the temporal attributes. Further analyses involved assessing the intensity of the hotspots by using kernel density estimation, which highlighted critical traffic routes.

Levine, Kim & Nitz (1995) and Alkhadour et al. (2021) assessed RTCs by using a nearest neighbour index tool. In the case of Alkhadour et al. (2021), the nearest neighbour index (NNI) ensured spatial and temporal patterns among RTCs in Amman (Jordan) (2017-2019). Levine, Kim & Nitz (1995), instead, used the nearest neighbour index tool in addition to statistical measures to investigate the spatial distribution of RTCs in Honolulu (United States). These findings were contrasted with the spatial distributions of population and employment to identify potential associations between RTC cluster attributes and population or employment patterns. Alkhadour et al. (2021), on the other hand, used the spatial patterns derived from the NNI tool to pinpoint RTC hotspots through hotspot analysis, enabling the deduction of areas with elevated RTC risk.

The temporal analyses by Levine, Kim & Nitz (1995) and Ma, Huang & Tang (2021) were approached from a statistical perspective. Levine, Kim & Nitz (1995) observed a higher concentration of RTCs on weekdays compared to weekends, with fatal incidents more likely to occur at night. Ma, Huang & Tang (2021) noted RTC occurrences between 7am and 9am, 5pm and 6pm, and on Saturdays throughout the year, with heightened activity in July and August. The temporal distribution of the RTCs involved in the studies by Osayomi & Areola (2015); Cheng, Zu & Lu (2018); Tola et al. (2021) and Hazaymeh, Almagbile & Alomari (2022) were assessed spatially. Hazaymeh, Almagbile & Alomari (2022) found that RTCs in Jordan were most observed during the months of July and August. Additionally, Hazaymeh, Almagbile & Alomari (2022) and Tola et al. (2021) noted higher RTC frequencies on Sundays and Thursdays. Cheng, Zu & Lu (2018) identified January and February as periods with weaker consecutive RTC hotspots.

In summary, the examination of various methods for analysing RTCs reveals that the findings from studies employing spatial autocorrelation and hotspot analyses align closely with the objectives of this research. The utilisation of spatial autocorrelation and hotspot analyses is justified as they establish the credibility of spatial clusters derived from the spatial data based on statistical measures rather than mere spatial occurrences. This alignment with previous research is crucial for this study, which faces challenges related to the lack of spatial definition in the RTC data. By leveraging spatial autocorrelation and hotspot analyses, the spatial inconsistencies inherent in the RTC data do not affect the spatial partitioning of RTC data into meaningful clusters. As a result, the accuracy and reliability of the spatial analyses conducted are enhanced. Furthermore, the reviewed studies showcased the flexibility of spatio-temporal analyses in scrutinising RTCs, with several investigations revealing the spatial and temporal patterns of specific RTCs.

3. METHOD

The RTC dataset acquired from the CoCT included crash data over 2017, 2018, 2019 and 2021, organized into separate excel worksheets, one for each study year. These study years were chosen as they were the most recently captured and available for dissemination at the time the data was requested. Data fields included: road name, intersection description (where applicable), kilometre value, police station, crash date, day and time, road type, sign visibility, crash type, specified cause, vehicle reference number, vehicle type, vehicle number plate, manoeuvre, travel direction, person type, injuries, population group, gender, age, liquor suspected and liquor tested.

Before the data was able to be assessed spatially, it first needed to be prepared, deduplicated and geocoded. This process is summarised below.

Figure 1: Workflow prior to spatio-temporal assessment

The separate excel worksheets containing the RTCs for each study year were firstly combined into a single excel worksheet. Subsequent data preparations included expanding road name abbreviations, assigning person codes to everyone in a crash vehicle, merging crash date and time, and splitting crash addresses into roads and suburbs. Additionally, it also involved extracting road intersection crashes marked with an “X” between road names so that these would be geocoded as they were the only spatially definable crashes. All the data preparation steps were conducted within excel and python software using several scripts. Subsequently, once the road intersection crashes were extracted, these were deduplicated before they were geocoded. The crash data required deduplication as the data collected was captured by crash victim and not for each crash. Failure to deduplicate records would have resulted in crash number misrepresentations at each site, as multiple records were linked to the same crash location. The crash deduplication was undertaken in the Microsoft SQL Server Management Studio (SSMS). Figure 2 below depicts the physical entity relationship diagram used as the foundation for the deduplication process.

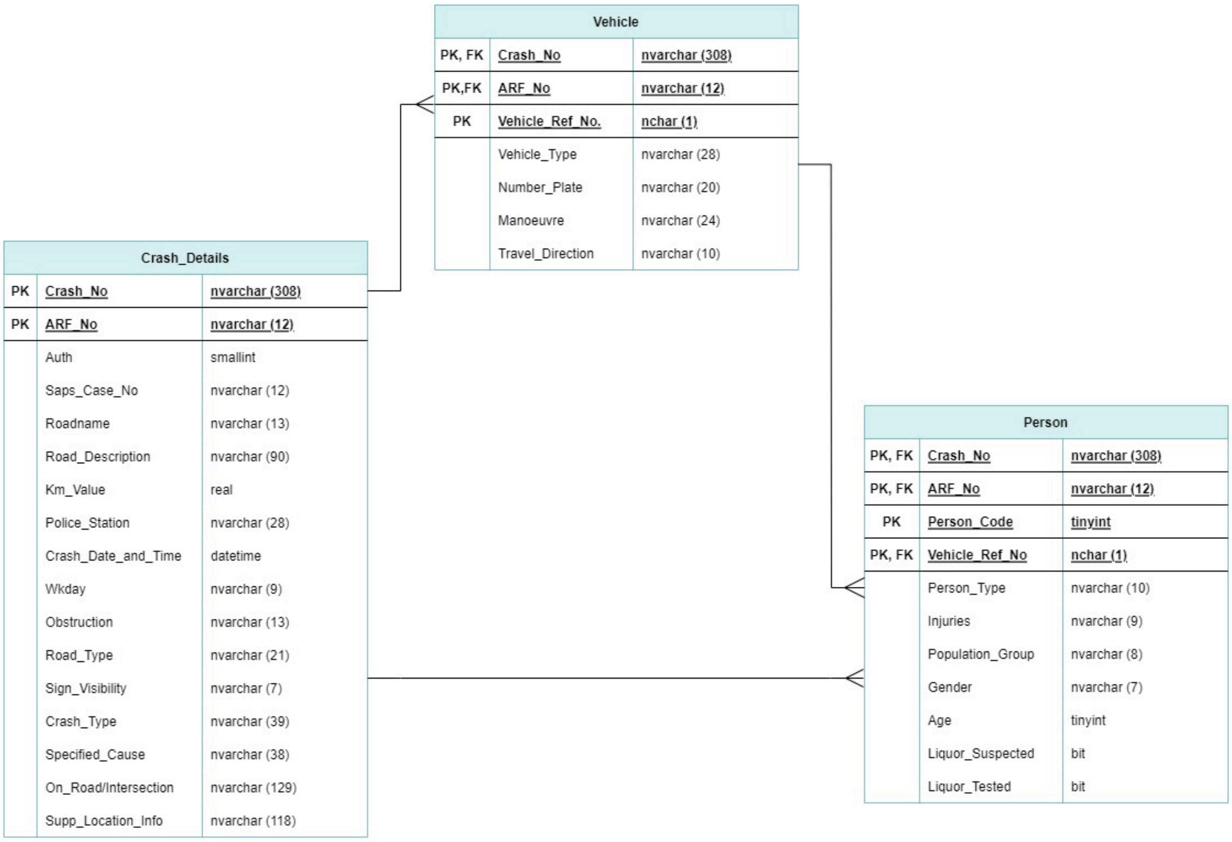

Figure 2: Physical ER diagram for RTC data deduplication

The three entities used to categorise the crash attributes were “Crash_Details”, “Vehicle” and “Person”. From the “Crash_Details” table, road intersection crash records were derived, while the “Vehicle” and “Person” tables contained the relevant crash data for the vehicles and the individuals involved, respectively. These three entities were related to each other by primary keys (PK) and foreign keys (FK) as indicated in Figure 2. As such, crash data was not lost through the deduplication process and the data integrity was upheld.

Once the records were deduplicated and thus associated to each crash (as opposed to each crash victim), crash records were passed through OpenRefine software to extract unique crash road intersection sites. While this is not a necessary step before geocoding, it was employed so that fewer geocoding resources (i.e. Esri credits) were used and so that the processing time was reduced.

The unique crash road intersection sites were thereafter geocoded within the ArcGIS Pro software. The first geocoding attempt was undertaken by using the CoCT geolocator called "Here/GC_CoCT" which was accessed by connecting to the CoCT server. From the CoCT locator, 11 741 RTC intersection sites out of 18 843 unique road intersection crashes were matched to a geographic location. Of the remaining 7 102 unique road intersection crashes, 5 638 were not matched to any specific location while 1 464 had several possible match locations and hence were labelled "tied". For the "tied" RTCs, supplementary crash location data from the original dataset was consulted to determine the most appropriate crash location. To resolve the "tied" RTCs that had no additional location information, the first locational crash site among the other matched crash sites was chosen. Two examples of how the "tied" crashes were geocoded are shown in Figure 3.

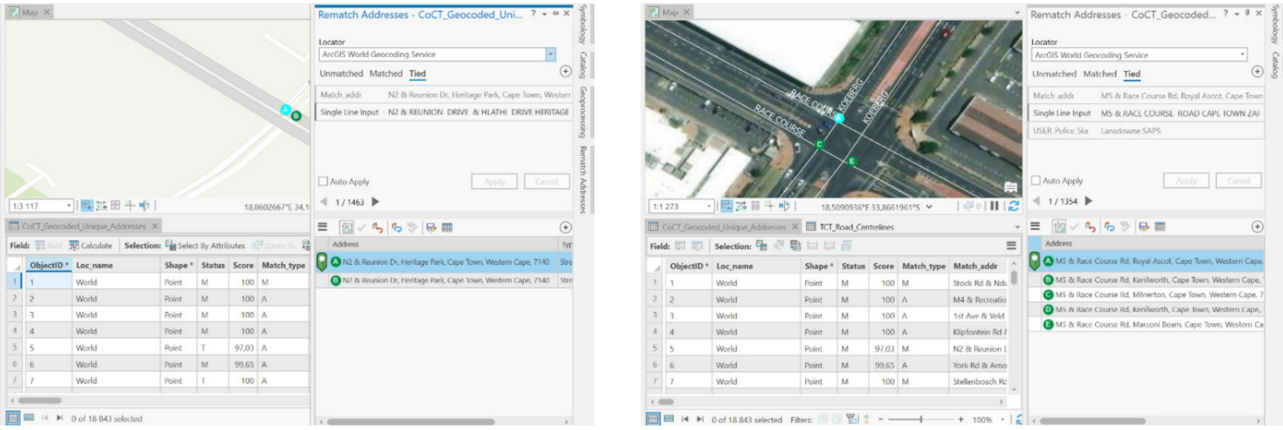

Figure 3: Two examples of how the "tied" RTCs were resolved

The "tied" geocoded RTC results that proved to be uncertain ended up being "unmatched" and hence added more crash records to the existing 5 638 "unmatched" records. To geocode the "unmatched" RTCs, the ArcGIS World Geocoding Service geolocator was used. To employ the ArcGIS World Geocoding Service, users must purchase credits. According to Esri, geocoding tools require 40 credits to geocode 1000 addresses. Although this locator could have been utilised initially in the geocoding process, it was chosen as a final option due to its associated costs. Of the 5 682 "unmatched" RTCs, the ArcGIS World Geocoding Service matched 3 176 crashes with a crash location, however 697 records were "tied" and 1 809 RTCs remained unmatched. To resolve the "tied" and "unmatched" crashes, google maps, open street map road centreline spatial data and the CoCT road centreline spatial data were consulted. A total of 280 unique crash records remained "unmatched" either because the crash address had non-intersecting roads or were erroneously recorded as road intersections (i.e. identical road names).

After reviewing the geocoding process it became apparent that geocoding non-intersection crashes would be unfeasible. This was largely due to the multitude of potential crash locations along a road section, presenting an infinite number of possibilities compared to road intersections. As a result, only RTCs that transpired at road intersections were geocoded.

The 18 563 geocoded unique RTC intersection locations were brought thereafter into SSMS to assign coordinates to the crash records with the same crash address as the unique RTC intersection crashes. The coordinates that were used to geocode the RTC intersections were based on the World Geodetic System of 1984 (WGS84) datum and Universal Transverse Mercator (UTM) projection.

Since the RTC intersections were spatially defined, they were able to be spatially assessed. The spatio-temporal analyses used to assess the geocoded RTC intersection locations was a global spatial autocorrelation, an emerging hotspot analysis and a local outlier analysis. The figure below illustrates the spatio-temporal workflow employed.

Figure 4: Spatio-temporal workflow summary

The first spatio-temporal analysis step involved performing a global spatial autocorrelation to determine whether the RTC counts at road intersections demonstrated spatial clustering. Two global spatial autocorrelations were performed, one for the road intersection RTCs between 2017 and 2019 and the other for RTCs at road intersections for 2021. The RTC data was assessed in this manner, as opposed to all the RTCs collectively, as the data had an Inherent gap, i.e. the 2020 data that was omitted. Ignoring this data gap and assessing the RTCs collectively could have disrupted subsequent spatio-temporal analyses and have introduced potential inconsistencies. As such, to maintain data integrity, the RTC data underwent cumulative assessment across consistent study years. Depending on the outcomes of the two global spatial autocorrelation analyses, RTCs were subjected to a hotspot analysis.

The unique road intersection locations from 2017 to 2019 were utilised to compute one global spatial autocorrelation, while the other involved the unique RTC intersection locations from 2021. The crash attribute on which the global spatial autcorrelation was performed was the crash frequency at each unique crash site as their spatial patterns were of interest for this study. The spatial relationships between RTCs were conceptualised with a fixed distance band so that each RTC neighbourhood distance delineation was consistent throughout all RTCs. The neighbourhood distance for the RTCs was not computed to ensure unrestricted delineation of RTC neighbourhoods. Manhattan distances were used to discern RTC neighbourhoods as opposed to euclidean distances since RTCs were anticipated to align to the road pattern.

A hotspot analysis was utilised to determine RTC hotspot and coldspot road intersections. The generation of a space time cube enhanced the spatial and temporal characteristics of the RTC data and ensured that the results of the hotspot analyses were defined spatially and temporally. Both hotspot analyses and space time cubes were generated for the two spatial analyses (i.e. one for RTCs between 2017 and 2019 and the other for RTCs in 2021). The space time cubes were created by point aggregation since the geocoded RTC data was point spatial data. The time step interval used to create the space time cube and subsequently the hotspot analysis results was one month. Had the space time cube been subjected to 3D visualisation, the spatial distribution of RTCs would have been able to be displayed for each month in the study period assessed. This visualisation was omitted from the study as it was beyond the scope defined in the initial aim.

To avoid limiting spatial analysis, RTC neighborhood distances were not specified. However, the number of spatial neighbours for each RTC was set to four so as to assess high or low crash occurrences, following Griffith's (1996) recommendation.The temporal neighbourhood used to discern RTC hotspots and coldspots was set to one month. This meant that hotspots and coldspots were determined by comparing the RTCs to four spatial RTCs and one temporal RTC.

Subsequent to the hotspot analyses, local outlier analysis was conducted. This analysis was referred to in section 2 as a local spatial autocorrelation. Since the interest of this study was to determine RTC hotspots at the road intersection scale, local spatial autocorrelations were used to reveal RTC intersection spatial patterns. The local outlier analysis, like the hotspot analysis was performed using the space time cubes created earlier to derive spatially and temporally influenced RTC clusters and outliers.

4. RESULTS

Before using the RTC locations that had been geocoded with the CoCT geolocator, several of these RTCs were cross-referenced with the ArcGIS World Geocoding Service for consistency assessment. The RTCs were arbitrarily chosen for the consistency assessment and spatially distributed throughout the entire study area. Five examples of these geolocator cross-reference geocoding tests are shown below.

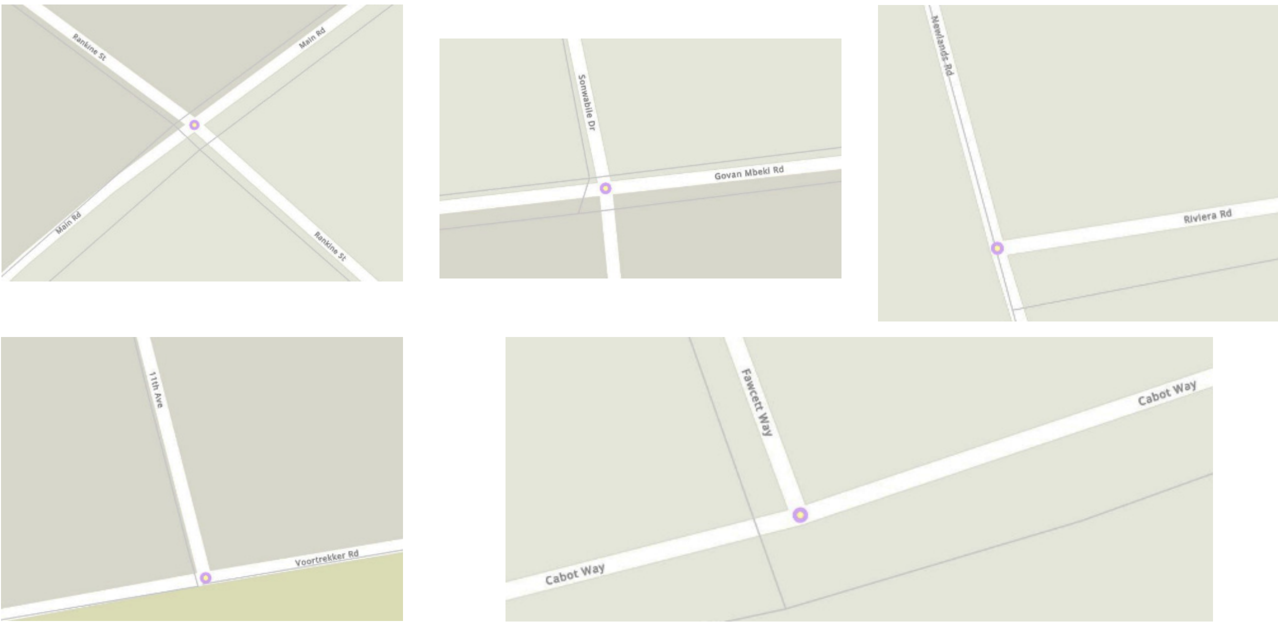

Figure 5: Five examples of the cross-reference geocoding tests conducted between the CoCT (purple points) and ArcGIS World Geocoding Service (yellow points) geolocators

In the illustrations above, the purple points indicate the spatial locations of the RTCs as geocoded by the CoCT geolocator while the yellow points indicate the ArcGIS World Geocoding Service equivalent. The grey lines represent the CoCT road centreline spatial data, and the backdrop is an ArcGIS street basemap. For the cross-reference tests conducted, all the RTCs geocoded from the CoCT geolocator and the ArcGIS World Geocoding Service coincided. As such, the RTCs from both geolocators were found to be consistent and hence were used to obtain coordinates for all the RTCs.

The geocoding and crash deduplication resulted in a total of 121 188 intersection crashes for the CoCT in 2017, 2018, 2019 and 2021. The following graphic illustrates the spatial distribution of these crashes.

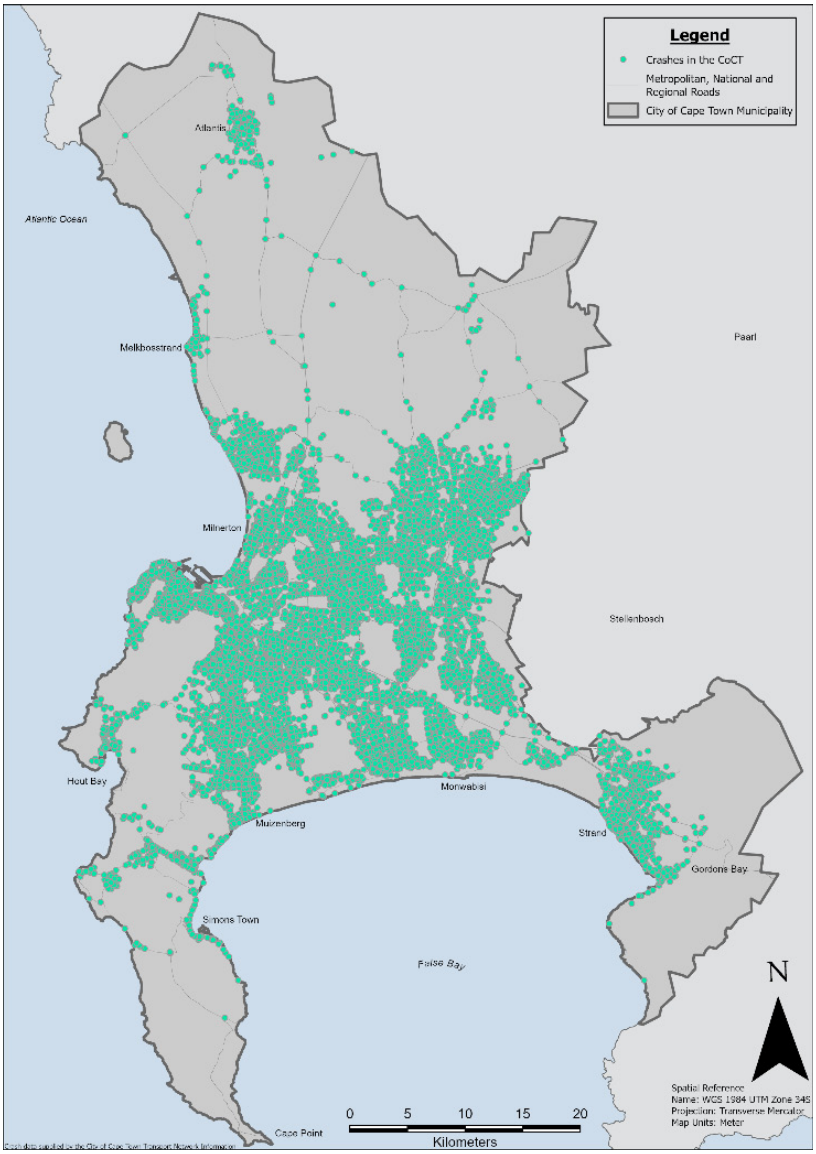

Figure 6: Spatial distribution of intersection crashes in the CoCT in 2017, 2018, 2019 and 2021

While the figure above illustrates that RTCs at road intersections appeared to mostly occur near the central regions of the CoCT, this result was insufficient to determine whether the road intersection RTCs displayed any spatial clustering. As such the global spatial autocorrelation analysis was performed and the results shown below.

Both global spatial autocorrelations shown below indicated that the RTCs at road intersections were spatially clustered. These spatial clusters were found to be associated with less than one percent being random, as indicated by the p-value. The results of the emerging hotspot analysis below (Figure 8 and Figure 9) provide a visual of these spatial clusters according to whether they were associated with high crash counts (i.e. hotspots) or low crash counts (i.e. coldspots). The first hotspot analysis demonstrates the emerging hotspot and coldspot RTC intersections for RTCs that occurred in 2017, 2018 and 2019.

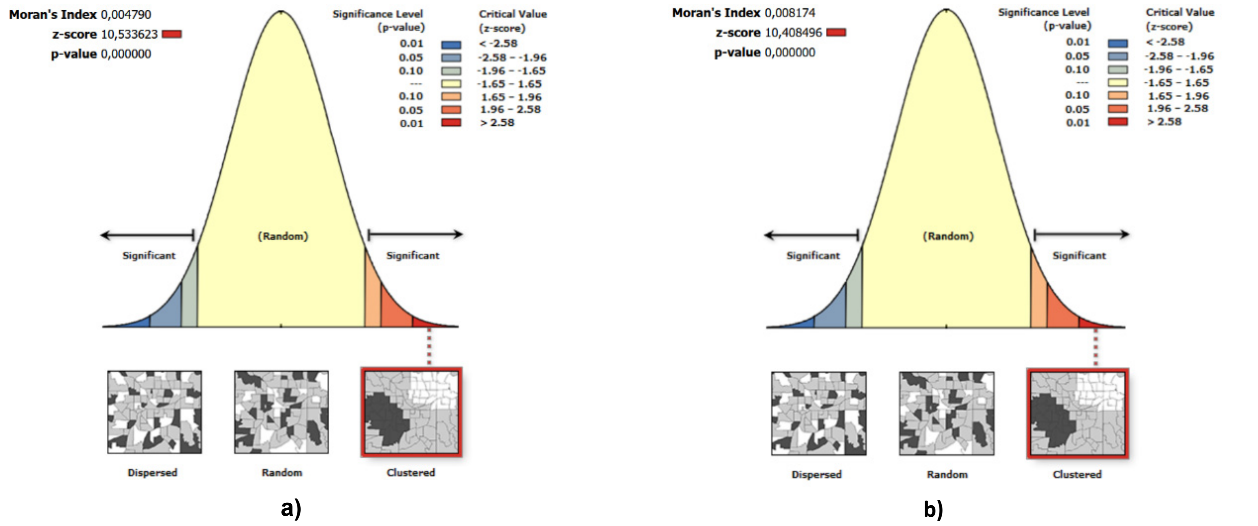

Figure 7: Global spatial autocorrelation results for RTCs in a) 2017, 2018, 2019 and b) 2021

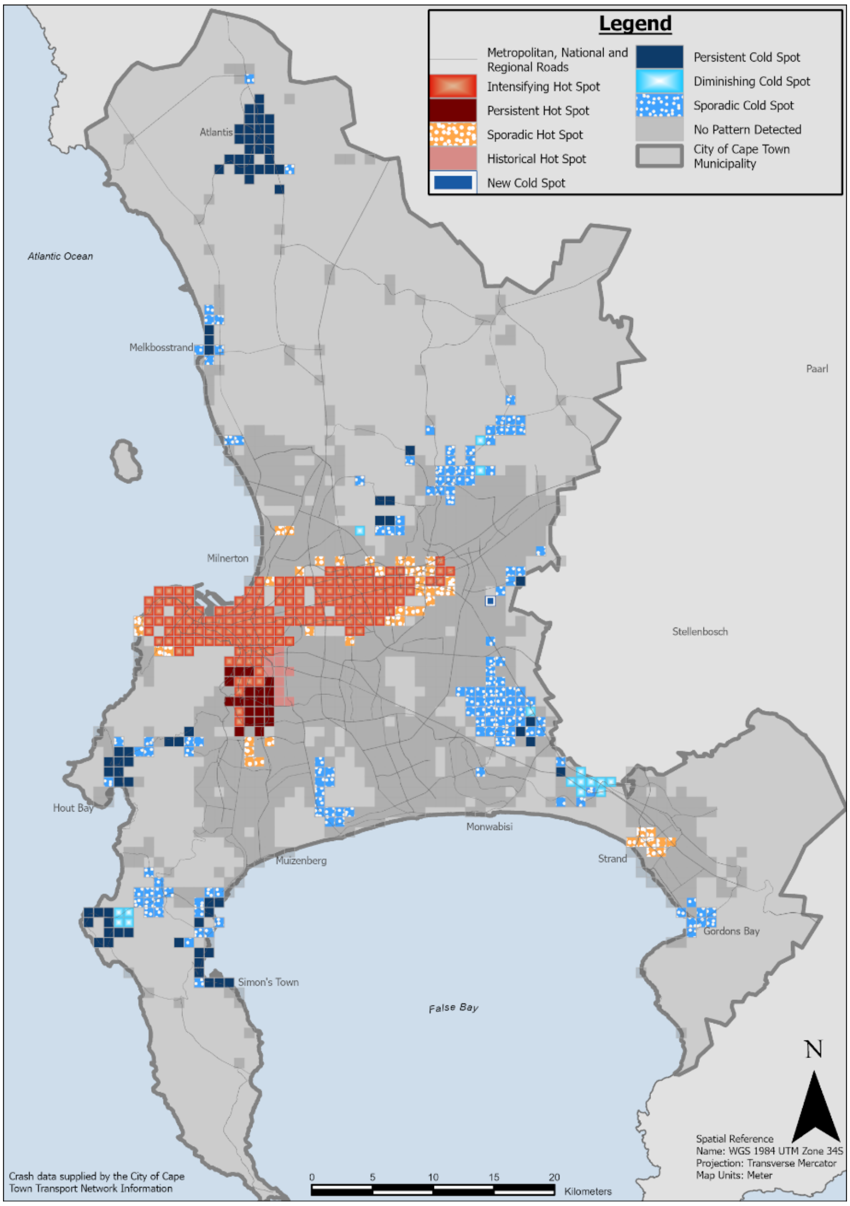

Figure 8: Spatio-temporal distribution of hotspot and coldspot RTC intersections in the CoCT in 2017, 2018 and 2019

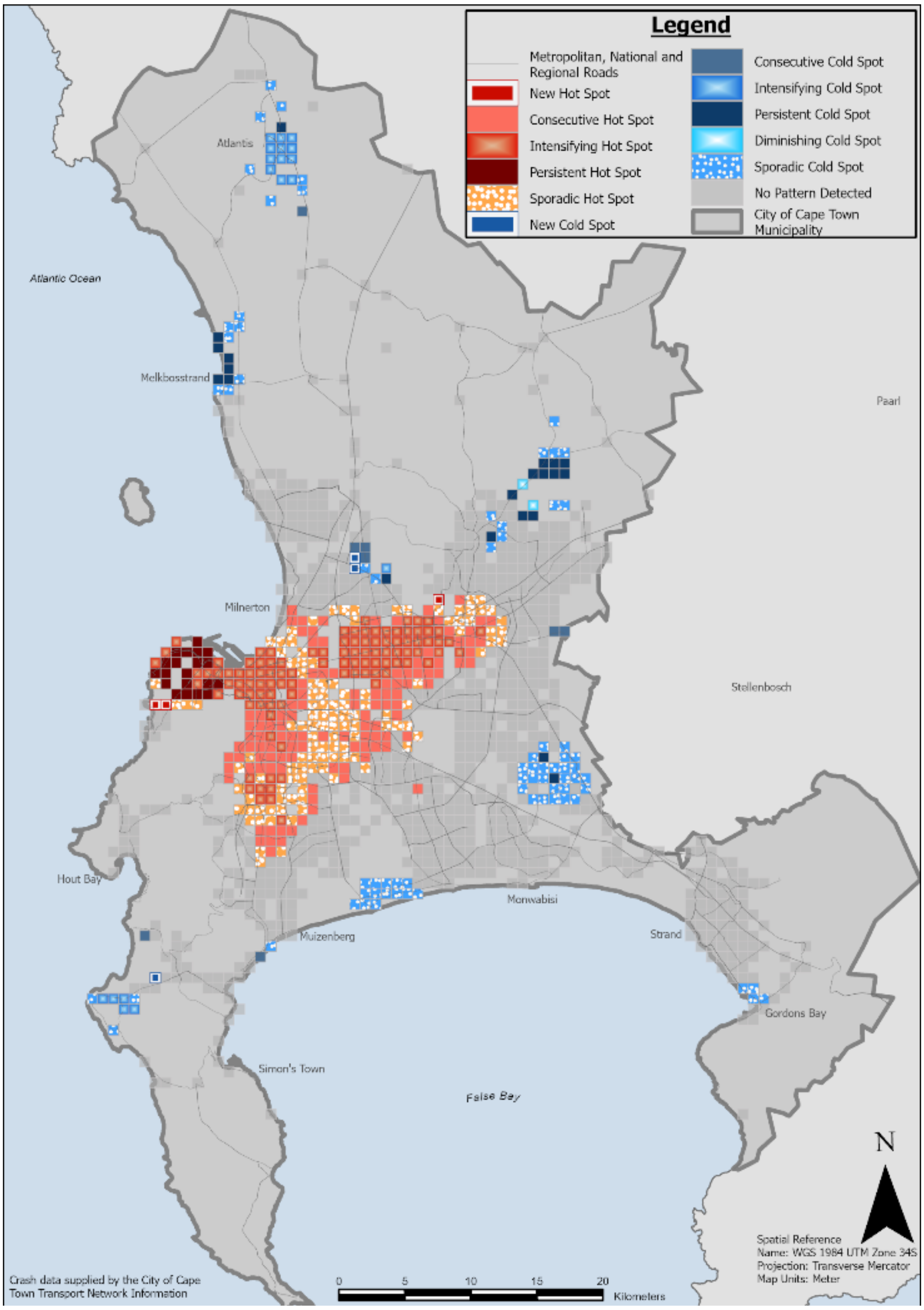

Figure 9: Spatio-temporal distribution of hotspot and coldspot RTC intersections in the CoCT in 2021

In 2017, 2018 and 2019, RTC intersection hotspots were comprised of intensifying, persistent, sporadic and historical hotspots. Intensifying hotspots indicated crash sites surrounded by higher crash counts than the CoCT's global average for at least 32 months (2017-2019) that intensified with time. These appeared to be the most predominant at intersections that occurred along the central and western regions of the CoCT. Most of the persistent hotspot crashes were clustered in the southwest regions and indicated intersections surrounded by high crash counts for 90 percent of the time interval compared to the global average. These crash counts were not, however, significantly intensifying like at the intensifying hotspot intersections. There were several sporadic hotspot intersections found around the central regions of the CoCT in 2017, 2018 and 2019. These hotspots indicated intersections where RTC neighbour counts were significantly high during some months between 2017 and 2019 and statistically insignificant during others compared to the average crash count for the CoCT. At these intersections, surrounding crash counts were statistically significantly high especially in the final time step i.e. December 2019. The historical hotspot intersections appeared to occur between the intensifying hotspots, persistent hotspots and insignificant crash count intersections. These hotspots indicated intersections where surrounding crash counts were statistically significantly high for at least 90 percent of the time interval (i.e. 32 months) but not in December of 2019. Unlike the hotspots, the coldspots, on the other hand, indicated intersections where surrounding crash counts were statistically significantly lower than the global average crash count for the CoCT between 2017 and 2019. The different coldspot categories appear to have mostly occurred along the periphery of the CoCT with a few sporadic and persistent hotspots occurring in the northeast regions of the CoCT.

Figure 9 illustrates the spatio-temporal distribution of hotspot and coldspot RTC intersections in the CoCT in 2021.

Like with the hotspot and coldspot RTC intersections observed in Figure 9, coldspots appeared along the peripheries of the study area while the hotspots were more centralised. In 2021, RTC intersection hotspots were labelled as “new”, “consecutive”, “intensifying”, “persistent” and “sporadic”. The new hotspots indicated intersection sites which observed statistically significantly high neighbour crash counts in the final time step i.e. December 2021. There were three such locations, each located near central business district regions. The consecutive hotspots, which appeared to occur in and around the central regions of the CoCT, signalled higher than average neighbour crash counts for less than 90 percent of 2021. At consecutive hotspots, neighbouring crash counts at road intersections were statistically significantly higher than the average crash count for the CoCT, at least for November and December of 2021. The intensifying hotspots were once again found near the central and western regions of the CoCT. The sporadic hotspots, however, appeared to have expanded from the central regions in 2017, 2018 and 2019 into western regions in 2021.

While the emerging hotspot analyses identified road intersections with high crash count occurrences, these were determined in terms of the RTC neighbours. In other words, the hotspots that were identified indicated road intersections with high crash count RTC neighbours. These road intersections were therefore deemed hotspots based on how their RTC neighbours compared to the average crash count for the CoCT. The results of the local outlier analysis, however, examined the crash counts at each intersection at the road intersection scale to determine road intersections with high crash counts. The local outlier results are shown in Figure 10.

In Figure 10, there are several cluster and outlier categories illustrated for crash counts in 2017, 2018 and 2019. These include high-high clusters symbolized in pink, high-low outliers symbolized in red, low-high outliers symbolized in dark blue and low-low clusters symbolized in light blue. The high-high clusters indicate intersections which were found to have observed high RTC counts and have been surrounded by high RTC crash counts compared to the global average crash count. These high-high clusters therefore indicate intersections that are often prone to RTCs and where road safety interventions are required. High-high clusters appear to have occurred centrally in the CoCT across 2017, 2018 and 2019. The low-low clusters indicate intersections where neighbouring crash counts were low and where low crash counts were observed. These appeared to occur along the periphery of the central regions in the CoCT. The high-low and low-high outliers indicate interesting crash sites. The high-low outliers indicate intersections where crash counts were high but were surrounded by RTC neighbours with low crash counts. These intersections are therefore also intersections of road safety concern. The low-high outliers, on the other hand, indicate intersections that observed low crash counts but that were surrounded by high crash counts compared to the CoCT crash count average. These intersections could indicate intersections where road safety interventions have been introduced and hence have resulted in fewer crashes being observed. The high-low outliers appear, unsurprisingly, to have occurred near low-low clusters whereas the low-high outliers appear to have occurred near high-high clusters and multiple outlier and cluster types. Additionally, non-significant RTC intersections are symbolized by white tiles while the purple tiles demonstrate intersections where there have been various cluster and outlier types through the study period.

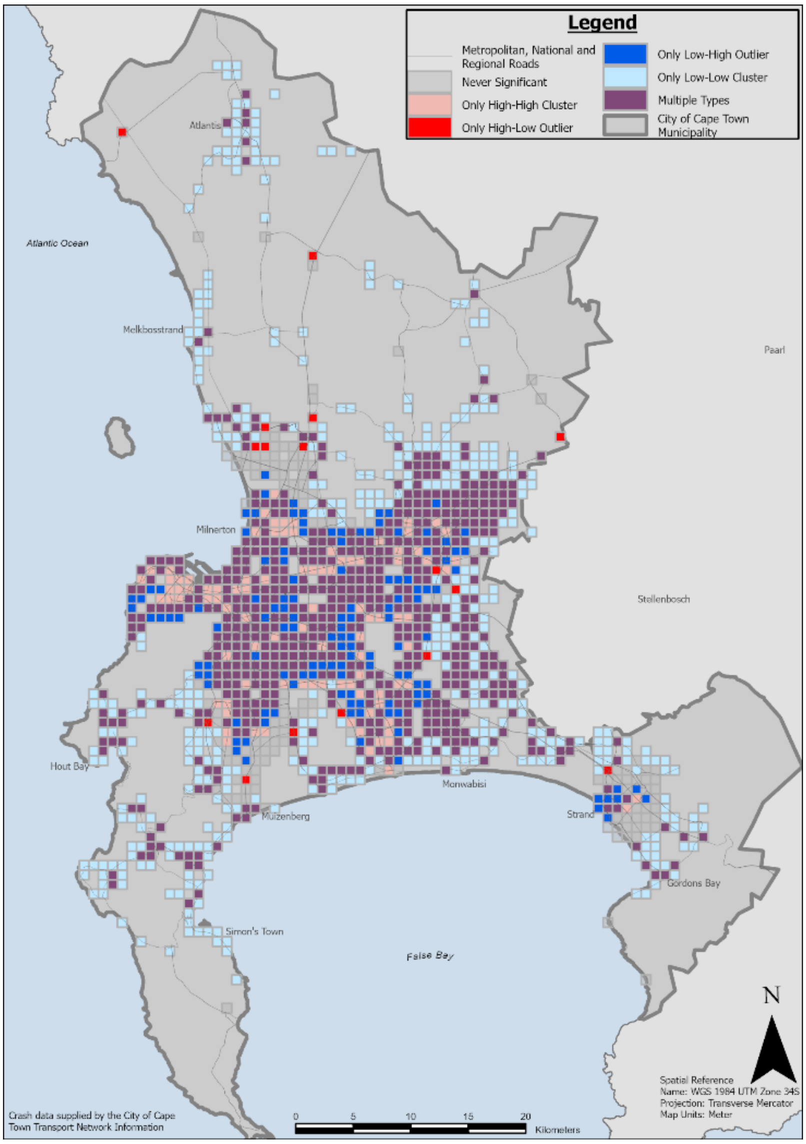

Figure 10: Spatio-temporal distribution of local outlier RTC intersections in the CoCT in 2017, 2018 and 2019

Figure 11 illustrates the local outlier results for the RTCs in 2021. It reveals a spatial distribution of RTC clusters and outliers akin to those observed in 2017, 2018, and 2019. High-high clusters, multiple cluster and outlier types as well as low-high outliers appear to cluster in and around the central regions of the CoCT and expanding to the west. There are however some differences between the clusters and outliers in the previous figure and those presented in the figure above. Several high-high clusters in 2017, 2018 and 2019 transitioned into multiple cluster and outlier types whilst others transitioned into low-high outliers. Likewise, several multiple cluster and outlier types in 2017, 2018 and 2019 changed into no patterns in 2021. While transitions could potentially suggest road safety interventions, intersections that have remained high-high RTC clusters or high-low outliers from 2017-2019 and into 2021 signal ongoing traffic issues requiring intervention.

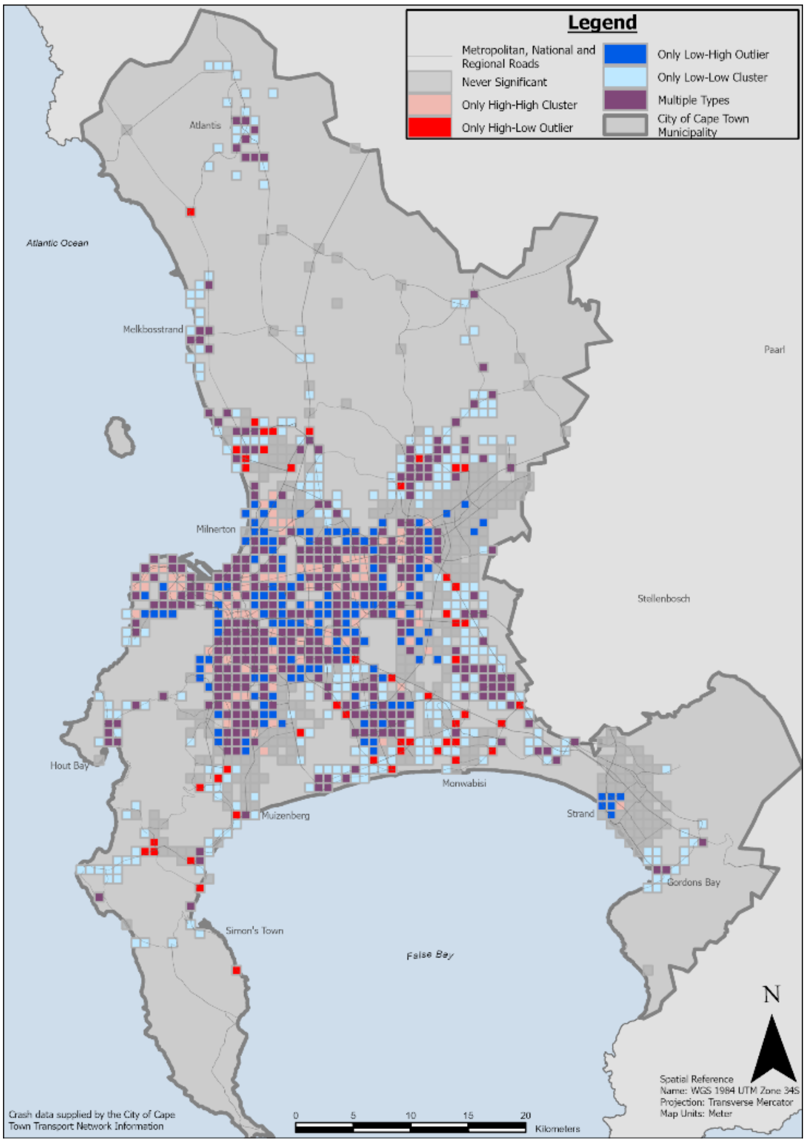

Figure 11: Spatio-temporal distribution of local outlier RTC intersections in the CoCT in 2021

5. CONCLUSIONS

This paper set out to explore the potential of methods drawn from geomatics to improve the spatial understanding of RTC patterns, and to apply these methods in the spatial definition of RTCs in Cape Town.

With regard to the potential of geomatics methods in improving spatial understanding, by focusing on road intersection RTCs with geographical coordinates, the study was able to employ spatio-temporal analyses effectively. Through spatio-temporal analyses, particularly hotspot and local outlier analyses, sites prone to high RTC counts can be pinpointed and used to target road safety interventions. The RTC sites that require road safety remedies are revealed by spatio-temporal analyses as hotspots in hotspot analyses and as high-high clusters and high-low outliers in local outlier analyses. Additionally, by employing 3D visualisation, it would become feasible to discern the individual temporal contributions to the overall findings and hence determine at which times of the year, road safety interventions would be required.

With regard to the spatial definition of RTC patterns in Cape Town, the hotspots pinpointed road intersection neighbourhoods with high crash counts, while the high-high clusters and high-low outliers highlighted specific areas and subsequently, road intersections of concern. Throughout the study period, several hotspots were identified from new, sporadic, consecutive, persistent and intensifying; all of which predominantly occurred in the central regions of the CoCT. Consequently, road intersections within the central regions of the CoCT are areas that require road safety attention. Additionally, the spread of high-high clusters across the suburbs, spanning from the City Bowl to townships in the south and northern suburbs, highlights the widespread nature of the road safety problem. Furthermore, the occurrence of high-low outliers along the CoCTs peripheries suggests that even seemingly "safer" neighbourhoods are not immune to road safety challenges, emphasising the importance of comprehensive safety measures city-wide. Addressing these findings could lead to targeted interventions aimed at reducing traffic crashes and improving overall safety for residents throughout the CoCT.

Having demonstrated the effectiveness of the methods employed, further analysis of the Cape Town dataset is possible in which vehicle, victim and crash type, amongst other variables, can be disaggregated. This is the focus of subsequent phases on research.

6. ACKNOWLEDGEMENTS

The lead author acknowledges the scholarship funding support received from the Department of Transport. The authors are grateful to David Sampson from the City of Cape Town Transport Network Information for providing the crash data essential for this study. They are also grateful to Mignon Wells and Nicholas Lindenberg from the University of Cape Town for their assistance in data preparation and data geocoding.

7. REFERENCES

Achu, AL., Aju, CD, Suresh, V, Manoharan, TP & Reghunath, R. 2019. Spatio-Temporal Analysis of Road Accident Incidents and Delineation of Hotspots Using Geospatial Tools in Thrissur District, Kerala, India. *KN - Journal of Cartography and Geographic Information*, 69(4):255-265.

Adeniji, AA, Mabuza, LH & Titus, E. 2020. Magnitude, trends and prevention of road traffic accidents in the Republic of South Africa. *South African Family Practice,* 62(1):1-4. Doi: 10.4102/safp.v62i1.5032.

Alimo, PK, Agyeman, S, Sumo, PD, Bouraima, MB & Lartey-Young, G. 2022. A call for action on alarming road traffic injuries in Africa. *International Journal of Injury Control and Safety Promotion,* 29(4):423-428. Doi: 10.1080/17457300.2022.2067184.

Arrive Alive. 2022. *Definitions And Explanatory Notes*. Available at: https://www.arrivealive.mobi/definitions-and-explanatory-notes. 2 September 2022.

Atluri, G, Karpatne, A & Kumar, V. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys,* 51(4):41. Doi: 10.1145/3161602.

Boots, B. 2002. Local measures of spatial association. *Ecoscience,* 9(2):168-176. Doi: 10.1080/11956860.2002.11682703.

Cheng, Z, Zu, Z & Lu, J. 2019. Traffic Crash Evolution Characteristic Analysis and Spatiotemporal Hotspot Identification of Urban Road Intersections. *Sustainability,* 11(1). Doi: 10.3390/su11010160.

Elvik, R. 2010. Why some road safety problems are more difficult to solve than others. *Accident Analysis and Prevention,* 42(4):1089-1096. Doi: 10.1016/j.aap.2009.12.020.

Esri. n.d. *Geoprocessing tools that use credits*. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/appendices/geoprocessing-tools-that-use-credits.htm. Accessed 10 February 2024.

Griffith, DA. 1996. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In *Practical Handbook of Spatial Statistics*. S.L. Arlinghaus, D.A. Griffith, W.C. Arlinghaus, W.D. Drake, & J.D. Nystuen, Eds. Boca Raton, Florida, United States of America: CRC Press, 65-82.

Hazaymeh, K, Almagbile, A & Alomari, AH. 2022. Spatiotemporal Analysis of Traffic Accidents Hotspots Based on Geospatial Techniques. *ISPRS International Journal of Geo-Information,* 11(4):1-18. Doi: 10.3390/ijgi11040260.

Hazen, A & Ehiri, JE. 2006. Road traffic injuries: Hidden epidemic in less developed countries. *Journal of the National Medical Association,* 98(1):73-82.

Jadaan, K, Al-Braizat, E, Al-Rafayah, S, Gammoh, H & Abukahlil, Y. 2018. Traffic Safety in Developed and Developing Countries: A Comparative Analysis. *Journal of Traffic and Logistics Engineering,* 6(1):1-5. Doi: 10.18178/jtle.6.1.1-5.

Li, B. 1996. Implementing spatial statistics on parallel computers. In *Practical Handbook of Spatial Statistics*. S.L. Arlinghaus, D.A. Griffith, W.C. Arlinghaus, W.D. Drake, & J.D. Nystuen, Eds. Boca Raton, Florida, United States of America: CRC Press. 107-148.

Ma, Q, Huang, G & Tang, X. 2021. GIS-based analysis of spatial-temporal correlations of urban traffic accidents. *European Transport Research Review,* 13(1):1-11. Doi: 10.1186/s12544-021-00509-y.

Nantulya, VM & Reich, MR. 2002. The neglected epidemic: Road traffic injuries in developing countries. *British Medical Journal,* 324(7346):1139-1141.

Onywera, VO. & Blanchard, C. 2013. Road accidents: a third burden of “disease” in sub-Saharan Africa. *Global Health Promotion*. 20(4):52–55. Doi: 10.1177/1757975913502688.

Osayomi, T & Areola, AA. 2015. Geospatial analysis of road traffic accidents, injuries and deaths in Nigeria. *Indonesian Journal of Geography,* 47(1):88-98.

Sokourenko, K, Wakim, M & Bassil, M. 2019.

Tola, AM, Demissie, TA., Saathoff, F & Gebissa, A. 2021. Severity, Spatial Pattern and Statistical Analysis of Road Traffic Crash Hot Spots in Ethiopia. *Applied Sciences,* 11(19):1-18. Doi: 10.3390/app11198828.

WHO. 2004. *World report on road traffic injury prevention*. M. Peden, R. Scurfield, D. Sleet, D. Mohan, A.. Hyder, E. Jarawan, & C. Mathers, Eds. Geneva: WHO.

Younis, MW, Batool, Z, Bukhari, M, Rehman, Z ur, Shahzad, S, Rehman, A ur, Khan, AH, Yasin, M. et al. 2019. Pattern of underreporting of Road Traffic Injuries (RTIs): An investigation of missing burden of RTIs in Pakistan. *Journal of Transport and Health,* 14. Doi: 10.1016/j.jth.2019.100575.