Journal of Transport Geography 101 (2022) 103350

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Logistics sprawl and the change in freight transport activity: A comparison of three measurement methodologies

Nadia M. Trent[a,*], Johan W. Joubert[b]

[a] Waikato Management School, University of Waikato, New Zealand
[b] Centre for Transport Development, University of Pretoria, South Africa

ARTICLE INFO

Keywords:
Logistics sprawl
Freight transport activity
GPS data
Activity chain
Methodology comparison

ABSTRACT

Logistics sprawl has been observed in many urban areas around the world, but the assumed link between increased logistics sprawl and increased freight transport activity has little empirical evidence. Because this link could influence policy-making to either fight or facilitate logistics sprawl, this paper investigates the implications of using different methodologies to quantify the relationship. Three different methodologies are compared. The first two methodologies are from well-respected sources in literature and the third is proposed in this paper. The methodologies measure transport activity related to logistics sprawl in three urban areas in South Africa between 2010 and 2014. The results from the methodologies are contradicting with the methodology proposed in this paper questioning the link between logistics sprawl and freight transport activity altogether. The comparison of the methodologies also shows that it is essential to include empirical data of actual vehicle movement when investigating logistics sprawl's impact on transport activity.

1. Introduction

Urban logistics systems have been adapting to sprawling populations, supply chain globalisation, changing consumer behaviour, and prescriptive land use planning for the past few decades (Aljohani and Thompson, 2016; Andreoli et al., 2010; He et al., 2018; Sakai et al., 2015; Kang, 2020a). Researchers have actively studied this phenomenon by measuring the geospatial spread of (predominantly) distribution centres and warehousing facilities. With a few exceptions such as Seattle (Dablanc et al., 2014), the Noord Holland and Zuid Holland provinces in the Netherlands (Heitz et al., 2017) and São Paulo (Guerin et al., 2021), these studies present evidence from Europe (Dablanc and Rakotonarivo, 2010; Heitz et al., 2020; Kumhálová et al., 2019; Strale, 2020), North America (Cidell, 2010; Jaller et al., 2017; Kang, 2020a, 2020b, 2020c), Asia (He et al., 2019; Sakai et al., 2015, 2017), and India (Gupta and Garima, 2017) that urban logistics systems adapt by moving distribution centres and warehousing facilities further away from densely populated city centres. This trend of decentralisation has been labelled "logistics sprawl".

As the field progresses, however, so do the concepts of urban logistics systems and logistics sprawl. One significant progression is the discourse regarding the types of facilities that should be considered as part of the urban logistics system and, by extension, included in studies of logistics sprawl. Although the majority of logistics sprawl researches focus exclusively on distribution centres and warehouses, Gardrat (2021) points out that even among these researches, the objects of study can "widely vary". We have observed this variance to be the result of the authors' notion of the urban logistics system, the specific framing of the research question, or the data available. In this study, we adopt a broader view of the urban logistics system, akin to that of Gardrat (2021), that considers all the facilities in the supply chain — from manufacturing to distribution to retail — as contributing to the geospatial extent of freight flows in an urban area. Thus we agree that a focus on distribution centres and warehouses (also termed 'freight terminals') "despite their significant role in urban freight mobility, is not enough to explain the generation and structure of freight flows" (Gardrat, 2021).

Regardless of the scope of facilities included in researches of logistics sprawl, it remains true that the general impetus of this work is the concern about the potential negative impact of logistics sprawl (Aljohani and Thompson, 2016; He et al., 2018; Yuan, 2018). Without evidence about whether logistics sprawl makes net positive or net negative impacts on an economy, its people, or the environment, how can policy-makers know whether it is a trend to fight or to facilitate?

* Corresponding author.
E-mail addresses: nadia.trent@waikato.ac.nz (N.M. Trent), johan.joubert@up.ac.za (J.W. Joubert).

https://doi.org/10.1016/j.jtrangeo.2022.103350
Received 9 June 2021; Received in revised form 11 April 2022; Accepted 11 April 2022
Available online 30 April 2022
0966-6923/

In this paper, we explore one facet of the logistics sprawl impact debate: the link between logistics sprawl and freight transport activity (hereafter referred to simply as *transport activity*). Higher transport activity — requiring more kilometres to transport the same tonne — is worse than lower transport activity as more emissions are generated per tonne of freight. The location and nature of activities that generate and attract freight is one of three elements in the urban logistics system proposed to contribute to transport activity (Gardrat, 2021). Logistics sprawl quantifies how these "locations of activities" have spread further apart. Therefore, a prevailing assumption in literature is that greater logistics sprawl results in higher transport activity, which results in more emissions. However, this assumption is not proven (Aljohani and Thompson, 2016). One counter-argument is that supply chains adapt to changes in such a way that overall transport activity may actually be decreased (Kang, 2020b; Sakai et al., 2017).

Whether logistics sprawl does increase transport activity or not is consequential in the policy stance taken towards it. Unfortunately, very few studies seek to verify this relationship. We have found only two studies, one by Dablanc and Rakotonarivo (2010) and another by Sakai et al. (2017), that did so empirically. With so few studies, it is understandable that a standardised methodology has not yet emerged. In fact, methodologies are custom-made to leverage the available data. Logistics sprawl datasets generally focus on identifying facilities while empirical data regarding transport activity are seldom available. We maintain that methodologies that do not include empirical data when evaluating the relationship between logistics sprawl and transport activity could put forward the wrong policy conclusions. Therefore we focus on methodologies that were based on empirical data of transport activity as we investigate the following two research questions:

1. *How does the relationship between the change in transport activity and logistics sprawl differ depending on the methodology used?*
2. *What are the implications of using different methodologies to evaluate the relationship between logistics sprawl and a change in transport activity?*

We address these questions by measuring the change in transport activity that resulted from logistics sprawl in three urban areas in South Africa using three different methodologies. We use the same dataset of commercial vehicle Global Positioning System (GPS) traces for each of the three methodologies and then compare the findings and policy signals that result. Using commercial vehicle GPS traces to identify logistics facilities is a recent development (Viljoen and Joubert, 2019), but is novel when applied to the study of logistics sprawl. Logistics facility locations are extracted from the GPS traces of commercial vehicles by identifying clusters of logistics activities created by thousands of vehicles over a period of time (see Section 3).

In the next section, we briefly review the landscape of logistics sprawl studies before focusing on the studies that developed two of the three methodologies that we will use. In Section 3, we present the dataset and the three urban areas investigated in this study. Section 4 investigates whether logistics sprawl occurred in the three study areas. We address the first research question in Sections 5–7. In Sections 5–6, we implement two methodologies from previous studies. In Section 7, we present a new methodology tailored to leveraging the empirical detail of transport activity in the GPS dataset. The second research question is addressed in Section 8, where we discuss the implications of using different methodologies before concluding.

2. Literature review

In the literature regarding logistics sprawl, three questions are typically addressed: *'why are logistics facilities moving outward?'*, *'to what extent has logistics sprawl occurred?'*, and *'what is the impact of this trend?'*

2.1. Why are logistics facilities moving outward?

Many studies have investigated the factors that led to logistics sprawl in specific urban areas. Some studies use rich contextual narratives (Dablanc and Rakotonarivo, 2010; He et al., 2019; Strale, 2020) or comparisons of data trends (Sakai et al., 2016) to describe influencing factors. Other studies use quantitative approaches to pinpoint the most significant decision factors that influence facility location (Bowen, 2008; Cidell, 2010; Kang, 2020b, 2020c). Although the possibility of context-specific exceptions exist (Cidell, 2010; Kang, 2020c), a few general narratives are supported by the literature. Supply chains have restructured in response to the global economy, changing consumer behaviour, and rapid advances in technology, communication, and logistics management (Kang, 2020a; Sakai et al., 2015, 2017). Thus larger land parcels at cheaper prices with quicker access to regional transport infrastructure are desired. Facilities have also been moved outward to avoid public opposition in densely populated or affluent areas (He et al., 2019; Strale, 2020; Yuan, 2018), and to capitalise on local government incentives for economic development (Strale, 2020). Although the "why" question is context-specific and complex to investigate, a firm foundation of inquiry exists.

2.2. To what extent has logistics sprawl occurred in an area?

Quantifying the extent of logistics sprawl has received the most attention in this field over the past 20 years, yet the literature showcases a diversity of data sources and methodologies used (He et al., 2018). Kang (2020a) defines two dimensions of measurement. In the first dimension, studies define the centrality and/or concentration of logistics facilities. Spatial centrographic techniques are typically used to pinpoint the centroid of logistics facilities (centrality). The average distance to this centroid then defines the concentration. In some cases, a fixed point like a central rail terminal or port is used instead of the centroid. The second dimension defines whether these measures (centrality and concentration) are expressed absolutely, considering only the logistics facilities, or relative to other phenomena like population sprawl. Relative measures are becoming more popular due to their greater explanatory power (for example Kang (2020a); Sakai et al. (2017); Strale (2020)).

The wealth of empirical studies in this area provide a well-trodden path to new researchers. However, the final question, that of the impact of logistics sprawl, has yet to receive the same attention (Aljohani and Thompson, 2016; Sakai et al., 2020).

2.3. The impact of logistics sprawl

The impact of logistics sprawl is felt by both public and private urban freight stakeholders. The private stakeholders are the logistics companies, who own or rent the logistics facilities, and transport operators. The impact on both these stakeholders is primarily economic. Logistics facilities trade off costs within the context of land use regulation, pricing, and opposition from communities (Lindsey et al., 2014). Meanwhile transport operators must respond to the needs of the changing freight landscape while maintaining their cost efficiencies and asset utilisation.

The public stakeholders are the communities who are serviced by these sprawling freight landscapes. They benefit from the economic activity and suffer the externality costs of noise, road wear and more. Concerns about justice arise because these benefits and costs are disproportionately distributed (Cidell, 2015; Yuan, 2018).

The concern about negative externalities in general, and the injustice of the distribution of the costs and benefits specifically, call for policy intervention. However, one glaring caveat to this discussion is a lack of empirical evidence. Of importance to this study is the lack of evidence regarding logistics sprawl and its impact on transport activity. A number of authors have *modelled* changes in transport activity for either hypothetical cases (for example Wagner (2010) and Wygonik and Goodchild

(2018)) or empirical cases (for example Gardrat (2021)). But we have found only two researches that measured the change in transport activity based on empirical freight trip data.

2.4. Empirical studies of logistics sprawl and transport activity

Dablanc and Rakotonarivo (2010) were the first to quantify the change in emissions caused by logistics sprawl. Their study focussed on large parcel and express transport companies in Paris, France, consequently defining “logistics facilities” as the terminals (cross-docking facilities) used by these companies. From their comprehensive dataset of facility locations, they could calculate the centroids of these facilities in 1974 and 2008. They determined that the average distance from a facility to the centroid had increased by 10km during this period. This indicated logistics sprawl. Unfortunately, the emissions analysis based on this finding was limited by a lack of data. Without extensive data about shipments from these facilities, they had to formulate assumptions about the freight trips generated by each facility, the percentage of freight trips heading into Paris, and the delivery fleet composition. These assumptions were based on contextual knowledge and industry averages. Their results showed an increase of 14,700 tonnes of CO_2 per year. Limited by their data, they were unable to account for any economic or transport behaviour changes over the three decades. Arguably, this is a simplification that could skew results.

Powered by data from the detailed Tokyo Metropolitan Freight Survey (TMFS) conducted in 2003 and 2013, Sakai and his collaborators could establish a more rigorous methodology (Sakai et al., 2015, 2017). The TMFS provides facility-level data about the number of shipments (truck trips), loads per shipment, and the origins and destinations of the shipments. Sakai and collaborators included distribution centres, truck terminals, warehouses, intermodal facilities and oil terminals in their definition of logistics facilities and measured the absolute sprawl of these facilities by the change in distance to a fixed point: the Tokyo Railway Station (Sakai et al., 2015). They also included relative measures by contrasting logistics sprawl to population sprawl and shipment origins and destinations (Sakai et al., 2015, 2017).

In Sakai et al. (2017), they measured the change in transport activity between 2003 and 2013 by comparing average shipment distances, vehicle-kilometres per tonne, and the facilities’ distance optimality gaps (DOGs). They found that transport had become more efficient over 10 years. By encouraging load consolidation and more optimal facility location, the logistics sprawl of the set of facilities considered in their study served to reduce the negative impact of the sprawling customers and suppliers in the Tokyo Metropolitan Area (TMA).

But this analysis was not without its simplifications. The TMFS provides shipment data aggregated to the municipality level (inside study area) and prefecture level (outside study area), necessitating a few assumptions. Firstly, if a truck made multiple deliveries/pick-ups in a municipal area, the shipment distance was the distance from the facility to *one* point in the municipal area. Thus any back-and-forth driving inside the municipality is ignored. Secondly, if a truck tour visited multiple municipalities in one trip, it was treated as multiple shipments — one to each municipality. This simplification distorts the impact of routing on actual vehicle-kilometres per tonne. Finally, all shipments to or from municipalities outside of the study area were “cut-off” at the nearest cordon point on the study area’s perimeter. Effectively, only the distance to the perimeter was considered. While this helped to bound the results to transport activity changes *in* the TMA, it ignored potentially meaningful results about the overall change in transport activity — especially considering the trend of larger logistics facilities covering broader catchment areas.

The methodologies used in these researches were dictated by shipment data availability. While there are a number of ways in which urban planners determine the location of logistics facilities, shipment data is notoriously difficult to obtain on a disaggregate level (Sakai et al., 2020; Trent et al., 2020). The advent of vehicle telematics (GPS tracking) in commercial vehicles opens up new possibilities in studying the link between logistics sprawl and transport activity.

3. The GPS dataset for three urban areas in South Africa

To our knowledge, no other study has determined the spatial distribution of logistics facilities based on the GPS traces of commercial vehicles. Studies use data sources that identify the locations (and characteristics) of facilities (Dablanc and Rakotonarivo, 2010; He et al., 2019; Sakai et al., 2015), or use counts of facilities or firms per county, ZIP code, or municipal area (Cidell, 2010; Kang, 2020a, 2020c; Jaller et al., 2017). Some studies have also used various indirect data sources like employment data (Strale, 2020), and cadastral data or satellite imagery (Krzysztofik et al., 2019; Strale, 2020). Comparing the strengths and weaknesses of the wide variety of data sources used to identify facilities is beyond the scope of this paper.

Despite the many ways in which facility locations are determined, it is seldom that researchers have access to data about the truck traffic generated by logistics facilities. The one notable exception is the TMFS survey (Sakai et al., 2015, 2017) which has shipment data, but even that survey lacks detail on how vehicles travelled to deliver the shipments. GPS traces offer promising opportunities as a standardised, ubiquitous, and more detailed data source of transport activity.

There are limitations to GPS datasets. Firstly, extracting facility locations from the GPS traces of vehicles is not trivial, requiring algorithmic techniques and experimentally-based assumptions. Fortunately, a growing body of work on this topic is constantly increasing the reliability of these methods (De Beer and Joubert, n.d.; Joubert and Axhausen, 2013; Joubert and Meintjes, 2015a, 2015b; Trent et al., 2020). Secondly, GPS traces do not report on vehicle type or owner, load factors, commodities, or trip purpose. The traces merely report where the vehicle was at what time and whether the engine was running or not. However, a number of studies have shown that this limitation can be overcome (Joubert and Axhausen, 2011; Ma et al., 2016; Viljoen and Joubert, 2019; Yang et al., 2014). In the next section, we describe the dataset used in this paper and how these limitations are addressed using algorithmic techniques.

3.1. This study’s dataset

The dataset underlying this paper was obtained from *Digicore Technologies* who provided the GPS traces of tens of thousands of commercial vehicles subscribed to their *Ctrack* telemetry service from 2010/01 to 2014/05 (53 months). Nearly 16,000 of these vehicles operated in and around the three study areas of this paper. The GPS traces of these vehicles over the 53-month period constitutes the dataset for this study. Based on national vehicle registration statistics (Electronic national administration traffic information system (eNaTIS), 2014), a conservative estimate pegs the sample size between 1%–2% of the total commercial fleet in the study areas during the four-year period.

Selection bias in the data is acknowledged. Commercial vehicle telematics is a booming industry in South Africa. In 2014, it was estimated that more than 20% of the commercial vehicle fleet were being tracked (Automotive Fleet, 2015), with market penetration expected to rise to 32.5% in 2020 (BergInsight, 2015). Because the market drivers behind commercial vehicle telematics (security and fleet efficiency) are broadly relevant, we assume that third-party telemetry services are prevalent across all industries. However, the confidentiality agreements with the data provider prevent us from verifying whether all industries are represented in our dataset. While we cannot verify the extent of the selection bias, we note that *Ctrack* was one of the primary market players at the time when the data was collected. It is evident from the GPS traces that the sample covers a broad geographic range within Southern Africa.

For this study, we use five monthly snapshots of GPS traces: March 2010, March 2011, March 2012, March 2013, and March 2014. While we could have used a number of months other than March, there is plausibility in our choice. The main summer break in South Africa is over December and January, both months having irregular, abnormal and commodity-specific logistics patterns. February, being not only a short month, often still exhibits some recovery patterns following the summer break. Consequently, March is the first month in the calendar year that one can argue has somewhat *normal* delivery and logistics patterns. One other reason for choosing March is the extent of the longitudinal data set, which runs from January 2010 until early May 2014. Since we cannot be sure of the completeness of April, and definitely not May, we use March to allow us another temporal point in 2014.

Using the methodology of Joubert and Axhausen (2011), the GPS traces were converted into chains of what those authors referred to as *minor* and *major* activities for each vehicle, using the same threshold duration of 300 minutes (5 hours) to distinguish between the two activity types. Minor activities are identified when the vehicle is switched off for shorter periods (<300 minutes) — presumably to load or unload freight at logistics facilities. Major activities are defined when the vehicle is stationary for extended periods (>300 minutes), such as overnight parking at a depot, for example. Commercial vehicle activity chains that executed *one or more* of their activities within one of the urban areas were retained.

Associated with each activity is its GPS coordinates that may be subject to signal noise. As a result, two vehicles performing an activity at the same facility will likely have different GPS coordinates associated with the two activities, even if only different by a few metres. Therefore, to associate activities with actual facilities, we use the density-based clustering technique of Joubert and Meintjes (2015b). Geographic clusters of minor activities are defined as logistics facilities. Importantly, there is no distinction between the types of facilities in the data whether retail, warehouses, transport terminals, or factories. All these facilities are implicitly included. This is in line with Gardrat (2021) who argues that the location of all facilities that generate and receive freight contribute to freight flows and, therefore, transport activity in an urban area.

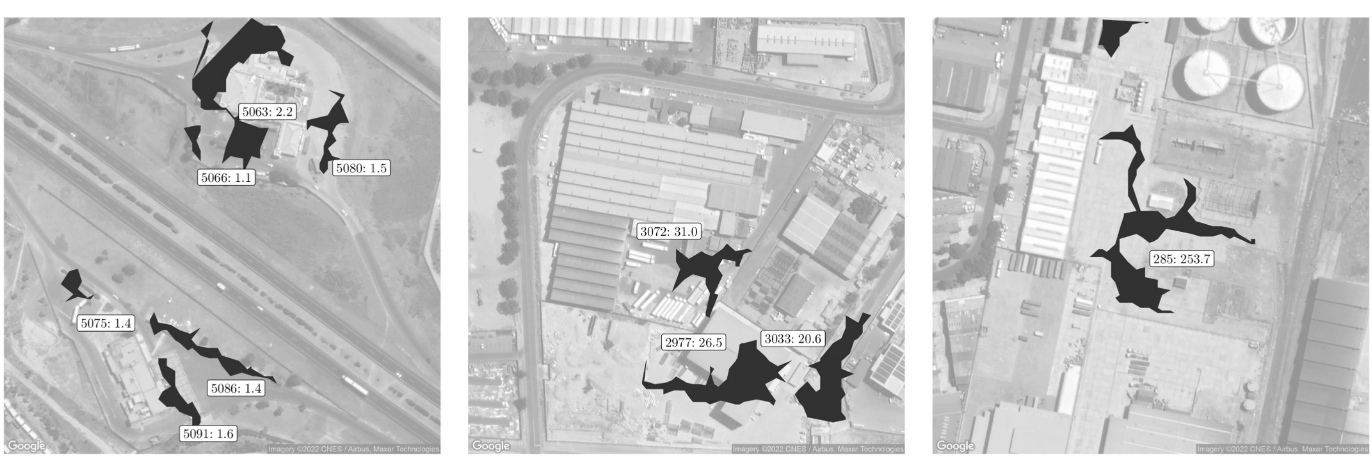
We do acknowledge that this approach, like other attempts to explain logistics sprawl, has limitations. The first limitation is that major activities are not included in the density-based clustering. The implication is that the first *loading* activity of a vehicle at its origin may be lost. However, if the location of that first *loading* activity is indeed an active logistics facility, it is highly likely that many other minor activities would also be performed there throughout the month. Therefore, the probability that a facility would be completely left out because some vehicles start or end some of their activity chains there with a major activity, is negligible.

The second limitation, as illustrated in Fig. 1, is that it remains problematic to distinguish between activity types. Fig. 1a shows the refuelling stations and truck stops on either side of a freeway. The three different facilities, in each case, could possibly represent different functional areas. One area might be the parking (facilities 5080 and 5075), another the forecourt (5063 and 5091), while another might be the dedicated truck refuelling points (5066 and 5086). One cannot simply remove any one of them and classify them as non-logistics facilities because, in the absence of more data, it is not clear whether a visit to the forecourt store is for buying a packet of sweets (non-logistics) or for replenishing the milk (a logistics delivery). Similarly, vehicles may visit the refuelling points to either fill up their own tanks, or deliver fuel to the underground bunkers, both located in close proximity.

There is, however, a counter-argument that the geospatial spread of all activities (non-logistics and logistics) performed by a commercial vehicle are dictated by the sprawl of the facilities that make up its route. Therefore, when examining the impact of logistics sprawl on transport activity, it is not unreasonable to include all activities conducted by a commercial vehicle. In addition, the density-based clustering uses parameters that would exclude outliers, therefore clusters are only created where many activities are concentrated, excluding isolated activities that are not part of business-as-usual.

One possibility to refine the identification of facilities, which we leave for future work, is to consider the number of unique vehicles performing activities at a facility and the total number of activities at a facility. The ratio of the two could indicate the average number of activities per unique vehicle. These ratios are expressed in the labels of the facilities of the different subfigures. The refuelling facilities (Fig. 1a) have very low ratios varying between 1.1 and 2.2. This suggests that a fuel stop is visited quite infrequently by a specific vehicle. Compare this to the distribution centre, facility 3072 in Fig. 1b, where the ratio is significantly higher, 31.0. While some vehicles may drop goods infrequently, the same distribution vehicles revisit the centre repeatedly, driving up the ratio. And finally, we see in Fig. 1c a large fuel franchise's distribution centre, where the same small fleet of tanker vehicles perform very many activities during a month, driving the ratio up to a high 253.7.

In the absence of a central business register in South Africa, there is



(a) Large refuelling stations and truck stops highlighting different functional areas.

(b) Freight carrier distribution.

(c) Fuel depot.

Fig. 1. Three examples showing the number of activities per unique vehicle for different facility types. Each facility is represented by a polygon that is the density-cluster hull. For example, the label 5063: 2.2 represents the facility ID (5063) and the number of activities per unique vehicle (2.2). Graphics produced using Kahle and Wickham (2013).

no way to classify or categorise facilities more accurately. And, supported by Gardrat (2021), any a priori assumption about facility types to in- or exclude may inevitably lead to unintended consequences.

3.2. Three urban areas

This study focusses on the three most prominent urban areas in South Africa namely the City of Cape Town Metropolitan Municipality (Cape Town) in the Western Cape province, the eThekwini Metropolitan Municipality (eThekwini) in the KwaZulu-Natal province and the entire Gauteng Province (Gauteng). Fig. 2 shows the areas in context of South Africa.

With its roots in the historic discovery of the Witwatersrand gold reef, Gauteng developed into a polycentric megacity consisting of five municipalities — three of which are metropolitan. It is the most densely populated urban area in South Africa with an area of $18{,}178\,\text{km}^2$ and a population exceeding 13 million in 2016 (Statistics South Africa, 2016).

Cape Town and eThekwini are two metropolitan municipalities that grew out of port cities. Both municipalities cover a similar area of $2{,}400\,\text{km}^2$–$2{,}600\,\text{km}^2$. Despite these similarities, Viljoen and Joubert (2019) discuss two distinct urban development trajectories that resulted in very different road transport networks and freight traffic trends. For Cape Town, we extend the metropolitan boundary to include the *functional* area that also includes the towns of Stellenbosch (east), Paarl and Wellington (northeast) and Malmesbury (north).

Fig. 3 reports the number of unique commercial vehicles tracked in the dataset as well as the number of activity chains extracted. We acknowledge the drop in 2014 but in the absence of additional data, we can only speculate about the cause: varying from substantial labour strikes in the platinum mining industry (and its effect on international commodity prices) to the troubled political tenure and instability of the incumbent presidency at the time. In the next section, we determine the extent of logistics sprawl in these areas between 2010 and 2014.

4. Logistics sprawl in the study areas

The starting point for all three methodologies is quantifying whether logistics sprawl or concentration occurred in an urban area. Similar to Dablanc and Rakotonarivo (2010), we first located the centroid of logistics activity at different points in time and then measured whether the mean distance of the logistics facilities to the centroid have changed over time. This spatial centrographic technique is more intuitive in the polycentric study areas than the approach taken by Sakai et al. (2015) that uses the Tokyo Railway Station as a fixed urban centrepoint.

Most studies of logistics sprawl calculate an unweighted centroid. When an unweighted centroid is used, the relative volumes of logistics activities produced by facilities are not incorporated. A facility that generates only five outgoing trips per month has the same influence on the measurement as a large distribution centre that produces five trips per day. This simplification is usually necessary because shipment data are notoriously scarce. From the GPS data in this study, we can calculate the number of outgoing trips from each facility. We leverage this to

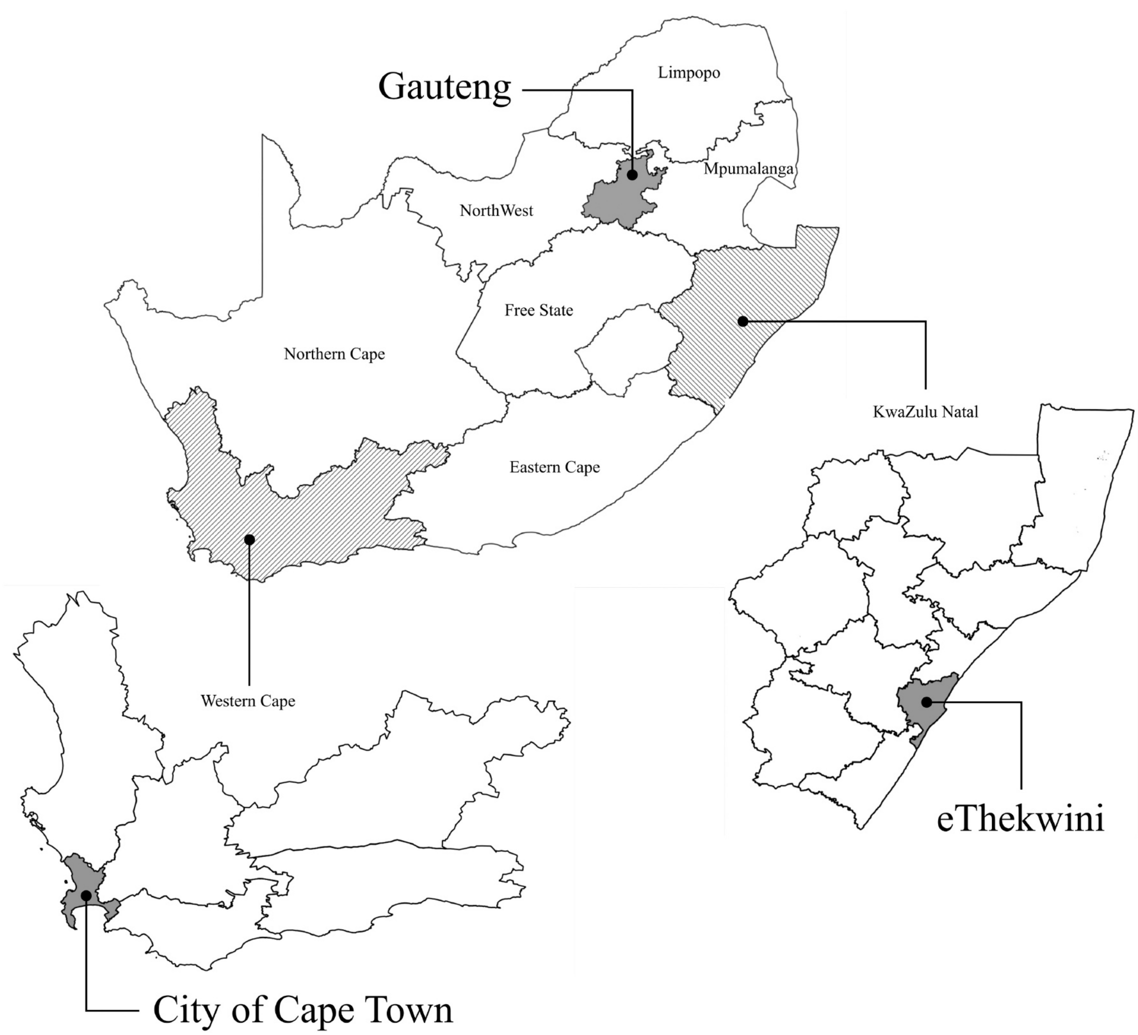


Fig. 2. Gauteng, Cape Town, and eThekwini in context of the rest of South Africa. Source: Viljoen and Joubert (2019).

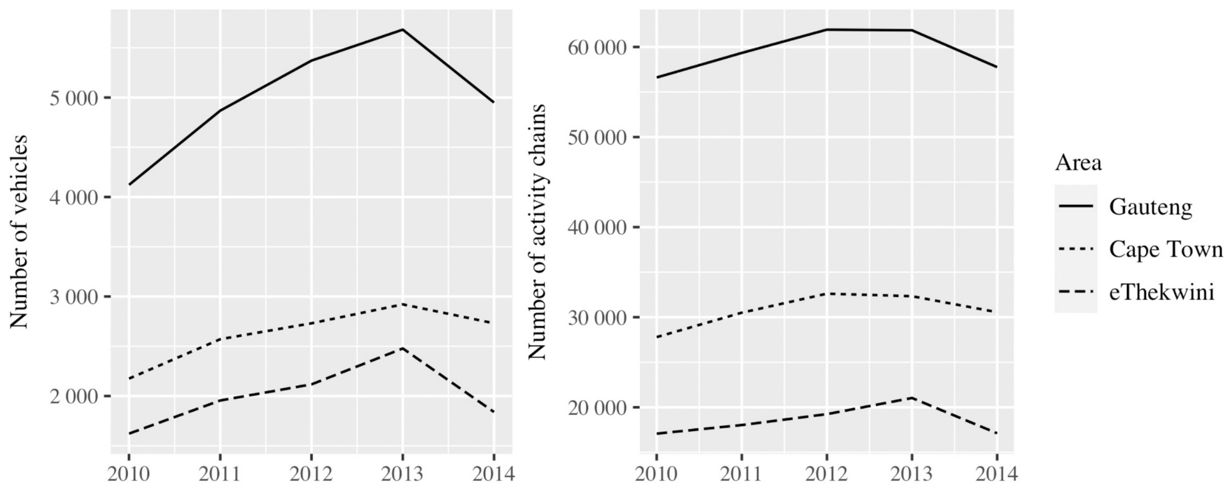


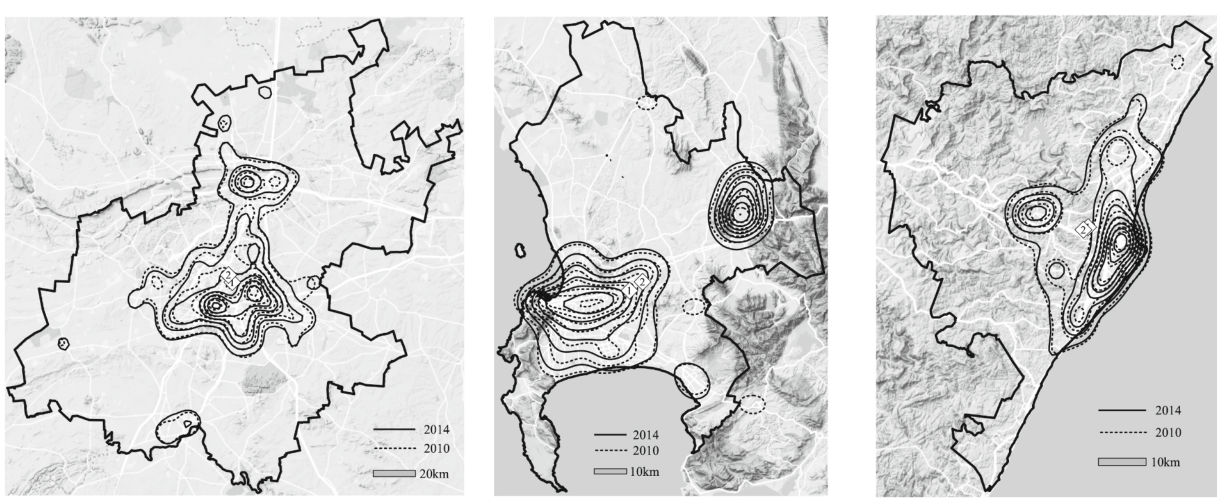
Fig. 3. Number of unique commercial vehicles and activity chains extracted from the dataset for each study area during each time period.

calculate logistics sprawl based on a weighted centroid and compare this to the results based on an unweighted centroid.

4.1. Locating the weighted and unweighted centroids

The table in Fig. 4 compares the movement of the weighted centroids between 2010 and 2014. The image plots the weighted centroids on contour maps of logistics facility density. In Gauteng, the weighted centroid shows a more pronounced shift west and, instead of a shift south as indicated by the unweighted centroid, it shows a shift north. These differences highlight the volume of logistics activity in the northwest of Gauteng. In Cape Town and eThekwini the differences in the shifts of the weighted and unweighted centroids are less pronounced. Both approaches show that Cape Town's centre of logistics activity moved northeast, away from the dense development of the urban centre and geographic constraints. In eThekwini, both approaches show a shift southwest, although the weighted approach shows that more logistics activity is moving inland along the N3 (national highway) corridor as the western shift is greater and the southern shift smaller when weighting is considered.

The shifting centroids show how the clusters of logistics facilities in each area are moving geographically. Next, the mean distances to these centroids measure how tightly these facilities are clustered around their respective centroids.



(a) Gauteng. (b) Cape Town. (c) eThekwini.

Area	Centroid	East / West	North / South
Gauteng	unweighted	0.04 km W	0.94 km S
	weighted	1.60 km W	2.93 km N
Cape Town	unweighted	2.17 km E	1.80 km N
	weighted	2.37 km E	1.73 km N
eThekwini	unweighted	0.24 km W	0.62 km S
	weighted	0.94 km W	0.42 km S

Fig. 4. The shift in the weighted centroid of logistics facilities from 2010 (marked '1') and 2014 (marked '2') mapped on the contours of the density of logistics facilities. The table compares the shift in the weighted and unweighted centroids.

4.2. Mean distance to the centroid

The mean euclidean distance of the facilities to the weighted and unweighted centroids ($\overline{D}$) for each period are shown in Table 1. We define $\Delta\overline{D} = \overline{D}_{2014} - \overline{D}_{2010}$. An increase in $\Delta\overline{D}$ would suggest that logistics facilities have moved further away from each other, while a decrease would indicate the opposite.

The WMW test rejected the null hypothesis that the 2010 and 2014 distributions of D in Gauteng or Cape Town were drawn from the same population. In the same vein, the two-sample t-test (hereafter 't-test') rejected the null hypothesis that the means of the underlying distributions were equal in either Gauteng or Cape Town. (All statistical tests in this study were performed at a 95% confidence level.) Therefore, the increase in $\Delta\overline{D}$ — whether measured to the weighted or unweighted centroid — implies that logistics facilities spread outward.

In the case of eThekwini, both the tests failed to reject the null hypotheses, regardless of which centroid was used in the calculation of D. So even though $\Delta\overline{D}$ decreased, suggesting that logistics facilities moved closer to each other, it cannot be said with confidence. Although there is no conclusive evidence of logistics sprawl or concentration in eThekwini between 2010 and 2014, we do not exclude the study area from the remainder of the analysis. Instead, we view it as a control area where the null hypothesis that $\Delta\overline{D} = 0$ cannot be rejected.

In all three study areas, there is a difference in $\Delta\overline{D}$ depending on whether the weighted or unweighted centroid was used and this offers insights about the sprawl of facilities that produce more logistics activity. In Gauteng, $\Delta\overline{D}$ increases from 1.52km to 2.04km when the centroid is weighted. This implies that facilities that produce more outgoing trips, tend towards the outskirts. The same trend is observed in Cape Town with $\Delta\overline{D}$ increasing from 1.73km to 1.82km. In eThekwini, the weighted centroid approach shows that $\Delta\overline{D} = -0.15\,\text{km}$ compared to -0.25km. Again, this suggests that facilities with more outgoing trips tend towards the outskirts as these did not concentrate as much as other facilities. However, the results for eThekwini remain statistically insignificant.

The observations in all three areas align with a number of other studies that found that larger, more active facilities moved outward to where there is cheaper and larger land tracts available and quicker access to regional transportation (Andreoli et al., 2010; Bowen, 2008; Cidell, 2010; Sakai et al., 2017). Being mindful of the wide variety of data sources and methodologies used across studies of logistics sprawl (He et al., 2018; Gardrat, 2021), we venture to compare the extent of logistics sprawl over time in Table 2. If we assume, for the sake of comparison, that sprawl occurs linearly over time, then the extent of sprawl in Gauteng and Cape Town is comparable to other urban centres. The lack of logistics sprawl in eThekwini is similar to the results for Seattle as reported by Dablanc et al. (2014).

This result, that logistics sprawl occurred in Gauteng and Cape Town and that there is no conclusive evidence of sprawl or concentration in eThekwini, is relevant to all three the methodologies presented in the remainder of this paper.

Table 2
Comparing the extent of sprawl observed in the study areas to sprawl observed in other urban areas.

Urban area	Period	Overall sprawl (km)	Annual sprawl (m)	In relation to	Reference
eThekwini	2010–2014	0[a]	0[a]	Centroid	Section 4
Cape Town	2010–2014	+1.82	+455	Centroid	Section 4
Gauteng	2010–2014	+2.04	+510	Centroid	Section 4
Seattle	1998–2009	−1.29	−117	Centroid	Dablanc et al. (2014)
Paris	1974–2008	+10.00	+294	Centroid	Dablanc and Rakotonarivo (2010)
Gothenburg	2000–2014	+4.20	+300	Centroid	Heitz et al. (2020)
64 US Metros	2003–2016	+4.70	+361	Centroid	Kang (2020a)
Yangtze River Delta	2005–2015	+4.02	+402	Centroid	Heitz et al. (2019)
Southern California	1998–2014	+6.76	+423	Centroid	Jaller et al. (2017)
Los Angeles	1998–2009	+9.74	+885	Centroid	Dablanc et al. (2014)

[a] No statistically significant evidence of logistics sprawl or concentration.

5. Methodology A: Mean distance to centroid

To calculate the change in emissions caused by logistics sprawl, Dablanc and Rakotonarivo (2010) multiplied the change in the mean distance to the unweighted centroid ($\overline{D}$) between 1974 and 2008 with an estimation of outgoing freight volumes from each facility. In their study, $\overline{D} = 10\,\text{km}$ when using an unweighted centroid. Based on government reports and industry interviews, it was assumed that each facility generated 193 tonnes of freight per day in 2008 and that 30% of this freight was destined for the Paris area. Light delivery vehicles reportedly delivered 82% of Paris parcels while trucks delivered the remaining 18%. Using these parameters and a unitary measure of CO_2 emissions, they calculated that annual emissions in this sector had increased by 14,700 tonnes over three decades. The authors are thorough in qualifying the limits of their simplifying assumptions. However, for at least seven years, this was the only empirical evidence of the net impact of logistics sprawl on the environment.

Although Dablanc and Rakotonarivo (2010) reported the overall increase in emissions, we can use the same approach with our dataset to calculate the change in transport activity.

5.1. Calculating changes in transport activity

To investigate the change in transport activity, we calculate the change in the total kilometres travelled, which we define as Δkm. We do not have industry averages and interview data like Dablanc and Rakotonarivo (2010), but we can determine how many vehicles left each

Table 1
The change between 2010 and 2014 in the mean euclidean distance of logistics facilities to the respective weighted and unweighted centroids ($\Delta\overline{D}$), calculated in each area.

Centroid	Year	Gauteng		Cape Town		eThekwini	
		$\overline{D}$ (km)	# facilities	$\overline{D}$ (km)	# facilities	$\overline{D}$ (km)	# facilities
Unweighted	**2010**	28.64	8,401	21.42	3,899	13.09	2,673
	2014	30.16	8,766	23.15	4,349	12.84	2,733
	$\Delta\overline{D}$	**1.52**		**1.73**		**-0.25**[a]	
Weighted	**2010**	28.58	8,401	21.00	3,899	13.04	2,673
	2014	30.62	8,766	22.82	4,349	12.89	2,733
	$\Delta\overline{D}$	**2.04**		**1.82**		**-0.15**[a]	

[a] Result is not significant based on Wilcoxon-Mann-Whitney (WMW) and two-sample t-test hypothesis tests.

facility (outbound trips) in 2014. We base our calculation on 2014's outbound trips only to emulate the original methodology. Using only outbound trips also prevents double-counting in our dataset. Table 3 tabulates the results when either the weighted or unweighted centroid is considered.

Using Methodology A and an unweighted centroid, we conclude that logistics sprawl produced an additional 676,680km in Gauteng, and an additional 376,481km in Cape Town for the month of March. If we consider the weighted centroid approach, the additional kilometres in Gauteng increase by 34.21% (676,680 → 908,175) and in Cape Town by 5.20% (376,481 → 396,067). In the case of eThekwini, the number of outbound trips in March 2014 were 143,232 but because we have assumed that $\Delta\overline{D} = 0$ for both the weighted and unweighted centroid, this methodology reports that there was no change in the kilometres in the area on the basis of logistics sprawl. This initial answer has, however, two notable caveats.

5.2. Caveats to Methodology A

The first caveat of this methodology is that the initial answer ignores the gradual nature of logistics sprawl by basing the calculation on 2014's activity alone. Table 4 shows the gradual increase of freight activity in the areas. However, to be confident in the calculation of a year-on-year Δkm, we need to be sure that the year-on-year distributions of D are significantly different. The WMW and t-test failed to prove this. While a four-year timespan may ignore gradual sprawl, it seems that shorter timespans cannot be confidently used with the current dataset.

The second caveat to this methodology regards the use of the change in the mean distance to centroid ($\Delta\overline{D}$) as the multiplier. Fig. 5 shows the cumulative distribution of the number of outgoing trips from facilities as these facilities' distances from the weighted centroids increase. It is clear in Gauteng and Cape Town that more outgoing trips are originating further away from the centroid in 2014 compared to 2010. In eThekwini, where the metrics show no evidence of logistics sprawl or concentration, the difference between the cumulative distributions is, as expected, nearly imperceptible.

The quartiles indicated on the graphs express bands around the weighted centroid. In Gauteng, for example, the first quartile (25%) of trips originating closest to the centroid sprawled approximately 15.91 km − 13.04 km = 2.87km while, if we consider the closest 50% of trips, sprawl was approximately 25.81 km − 21.65 km = 4.16km. In the third quartile, the difference was only 1.45km. Meanwhile, for those 25% of trips originating furthest from the centroid, the distance to the centroid actually decreased from 2010 to 2014. These quartile values are a far cry from the multipliers shown in Table 3. Using the $\Delta\overline{D}$ as the multiplier assumes that facilities and, importantly, the volume of logistics activity, are spread out equally. Fig. 5 and the differences in the weighted and unweighted centroids, show that this is not accurate.

Another question raised about using $\Delta\overline{D}$ is whether the trips originating farthest from the centroids are actually headed for the centroid? In other words, is it the distance from the centroid that dictates the kilometres travelled, or are these trips heading to destinations away from the centroid? When Dablanc and Rakotonarivo (2010) used this approach in the case of parcel deliveries into the Paris urban area, this question was probably irrelevant as most traffic was heading towards the centroid. However, when dealing with logistics sprawl in general and in polycentric urban areas, like Gauteng and Cape Town, this is pertinent.

The final caveat is that using the total of outgoing trips as a multiplier does not account for changes in the underlying economic activity. Methodology B improves on the caveats by accommodating changes in economic activity and considering *where* commercial vehicles travel to.

Table 3
Additional kilometres (Δkm) resulting from logistics sprawl according to Methodology A. Results differ based on whether the weighted or unweighted centroid was used.

Area	# outgoing trips	Centroid	$\Delta\overline{D}$ (km)	Δkm (km)
Gauteng	445,184	Unweighted	1.52	676,680
		Weighted	2.04	908,175
Cape Town	217,619	Unweighted	1.73	376,481
		Weighted	1.82	396,067
eThekwini	143,232	Unweighted	0[a]	0
		Weighted	0[a]	0

[a] No statistically significant evidence of logistics sprawl or concentration.

6. Methodology B: Survey-based shipment calculations

Sakai et al. (2017) used three metrics to investigate the change in transport activity (they refer to *shipping efficiency*) between 2003 and 2013. Firstly, a change in vehicle-kilometres per tonne measured the increase/decrease in transport activity required per tonne circulating in the economy. Secondly, they tracked the change in the average shipping distance (ASD) for all trips to and from customers and suppliers. Finally, they calculated the change in the DOG, which is the gap between the actual ASD and the minimum ASD that would've resulted if a facility was located optimally with respect to its customers and suppliers. All three metrics account for changes in the size of the economy and the impact of the sample size.

With our dataset, we cannot reproduce the first metric as we lack load factor data. But we can calculate the ASD and DOG. But first, we must identify each facility's customers and suppliers — collectively referred to as *stakeholders* — from the activity chain data.

6.1. Identifying a facility's stakeholders

In the TMFS, the main stakeholders of a specific facility were self-reported (Sakai et al., 2017). Because the GPS data are anonymous, identifying stakeholder relationships is tricky. But complex network theory has proven useful in this regard (Viljoen, 2018). The only study we are aware of that extracts supply chain relationships from GPS traces is that of Viljoen and Joubert (2019). The facilities are the nodes of the network while commercial vehicle activity constitutes the edges. Facilities are connected if they are consecutive activities on the same activity chain. How we build on this approach to identify stakeholders is described in the remainder of this section by means of an illustrative example.

Consider the example, illustrated in Fig. 6, of identifying the stakeholders of some facility, B. We start by extracting all activity chains that visit facility B. Consider an activity chain where a vehicle starts at facility A and performs logistics activities at facilities B, C, and D before ending its trip back at facility A. The activity chain is represented as the sequence A-B-C-D-A. The first-order neighbourhood (FON) of B is all the facilities with whom B shares a *direct* edge. Fig. 6a shows how the direct edges (A-B and B-C) are mapped while the indirect edges (C-D and D-A) are ignored. The weights of the direct edge are set to one, i.e. $w_{AB} = w_{BC} = 1$.

Next, consider Fig. 6b as the activity chain A-B-D-E is added. Since A-B already exists in B's FON, the edge's weight is increased to $w_{AB} = 2$. The second edge, B-D, is added with weight $w_{BD} = 1$ and the final edge, D-E, is ignored as it is not directly connected to B. The existing edge B-C remains unchanged with the addition of this activity chain. In a similar manner, Fig. 6c and 6d add two more activity chains.

Once the FON has been constructed for facility B, we use the edge weights to differentiate its stakeholders from facilities that just happened to be visited by the same vehicle on the same day. The edge weight is an indication of how often two facilities were connected by activity chains. The higher the weight, the greater the likelihood that these facilities do business. We dealt with incoming edges (from suppliers) and outgoing edges (to customers) separately.

Table 4
Methodology A: The gradual year-on-year changes between $\overline{D}$ and commercial vehicle activity.

	Gauteng			Cape Town			eThekwini		
	$\overline{D}$(km)	# facilities	Outgoing trips	$\overline{D}$(km)	# facilities	Outgoing trips	$\overline{D}$(km)	# facilities	Outgoing trips
2010	28.64	8,401	400,372	21.42	3,899	183,197	13.09	2,673	123,396
2011	28.50	10,068	463,161	21.98	4,641	207,650	12.77	2,992	142,906
2012	29.25	10,176	483,143	22.15	4,920	221,279	13.14	3,007	154,671
2013	29.67	10,479	498,647	22.47	4,939	225,951	12.84	3,378	175,595
2014	30.16	8,766	445,184	23.15	4,349	217,619	12.84	2,733	143,232

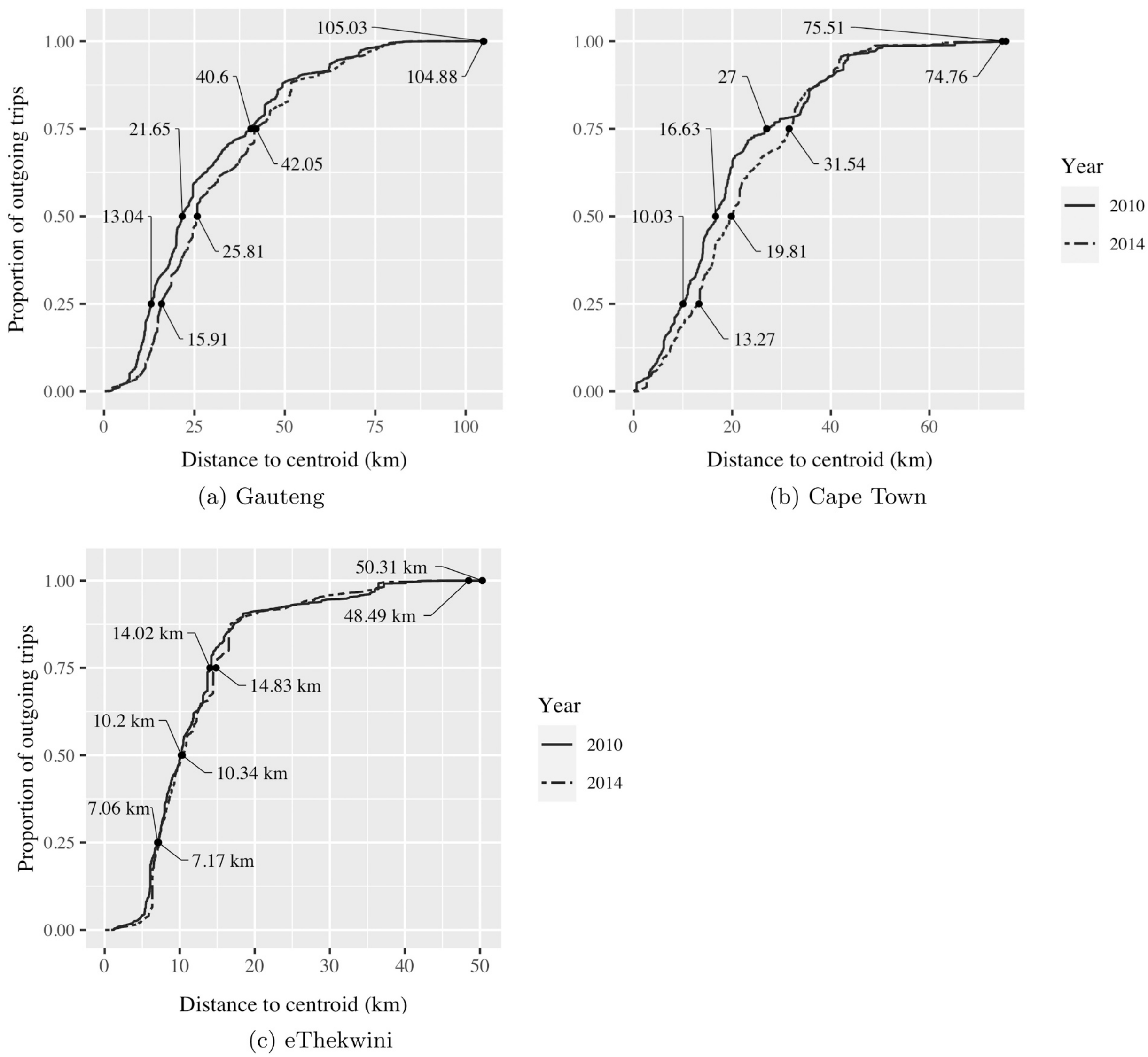
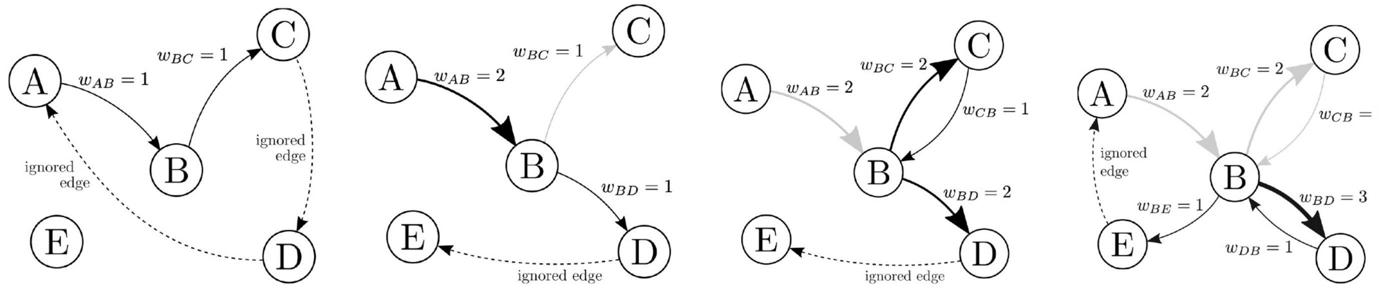


Fig. 5. Cumulative distribution of the number of outgoing trips from facilities as facility distance from the weighted centroid increases.

Consider the outgoing edges of facility B's FON as shown in Fig. 6d. B's out-degree is the sum of the outgoing weights: $w_{BD}+w_{BC}+w_{BE}=6$. Let the number of facilities in the FON be represented by n. Then, if the edges were equally weighted, each edge would carry $\frac{1}{n}=\frac{1}{3}=0.33$ of facility B's out-degree. This is the threshold we use to determine whether a facility is *important enough* to be regarded a customer. In the example $\frac{w_{BE}}{6}=0.167<0.33$, therefore facility E is not considered a customer of B. Facilities C and D are regarded customers because $\frac{w_{BC}}{6}=0.33$ and $\frac{w_{BD}}{6}=0.5$.

The filtering mechanism reduced the sizes of facilities' FONs significantly. Fig. 7 plots the percentage reductions in the size of the FON against the original size of the FON.

Only facilities that had completely homogenous FONs (i.e. equal edge weights) showed a 0% change. But it is clear that most of the FONs experienced a great reduction in size. This implies that most of the connections facilities have with other facilities are merely incidental. Only a few of a facility's connections occur frequently enough for us to assume that there exists a more recurrent business relationship. Having identified the stakeholders of each facility, we could calculate the changes in transport activity.



(a) Adding *A*-*B*-*C*-*D*-*A*. (b) Adding *A*-*B*-*D*-*E*. (c) Adding *B*-*C*-*B*-*D*-*E*. (d) Adding *B*-*D*-*B*-*E*-*A*.

Fig. 6. Illustrative example of how a weighted first-order neighbourhood (FON) is constructed for facility *B* from activity chain data. Each subfigure indicates the cumulative FON for node *B* after adding one more activity chain. Dotted lines represent those edges that are ignored because they are not in node *B*'s FON, while grey edges indicate FON edges that are not part of the current activity chain added.

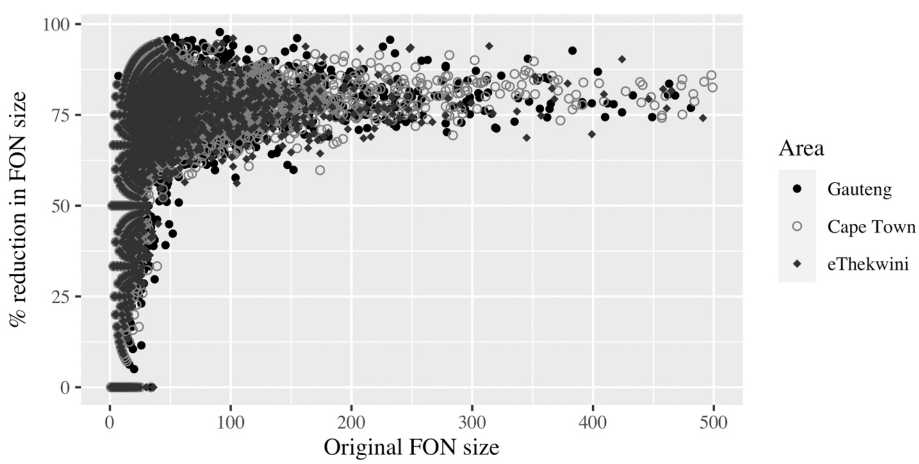


Fig. 7. Percentage reduction in the FON size when only stakeholders are retained plotted against the original (unfiltered) FON size. Thirty-two FONs with an original size >500 are omitted for better illustration.

6.2. Calculating changes in transport activity

A key difference between our dataset and that of Sakai et al. (2017) is the level of aggregation. Their dataset is aggregated to municipal (inside study area) and prefecture (outside study area) level. In our dataset, we have unique coordinates for all facilities and we use these instead of aggregating to municipality level because two of the three study areas are much smaller than the TMA. Aggregating to municipality level in these two areas would be far too crude for calculating ASD and DOG.

Another difference is that Sakai et al. (2017) use the TMA road network to calculate distances. Our GPS traces are not matched to the underlying road network. As a result, we need to estimate the actual distance travelled by a vehicle. To account for the circuitity (windiness) of the actual vehicle movement, we multiply the euclidean distance by 1.25, a multiplier empirically determined from South Africa's urban road networks.

Fig. 8 plots the distributions of the ASD and DOG metrics in each area over the four-year span. The distributions look remarkably similar over time. In fact, only in Gauteng were the differences in ASD statistically significant. Meanwhile, for the DOG of Gauteng the tests concurred that while the shape of the distributions are not similar, it cannot be said that their means are significantly different. The following interpretations bear the results of the hypothesis tests in mind.

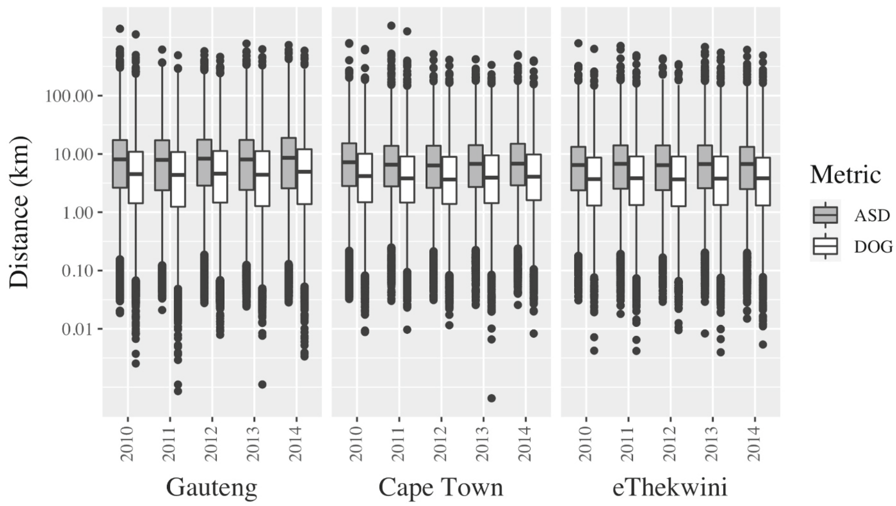
In Gauteng, the increase in the mean and median ASD supports the notion that logistics sprawl also increased transport activity over four years. Vehicles had to travel further to connect facilities with their stakeholders. We are careful to read too much into the changes in the mean and median DOG based on the hypothesis tests. Instead, we note that while there was a definite, albeit small, increase in distance between facilities and their stakeholders, it does not necessarily mean that they were less optimally located.

In eThekwini and Cape Town, the statistical tests failed to show that the differences in ASD and DOG over four years were significant. For eThekwini, this result is intuitive as there is no evidence of logistics sprawl. In Cape Town, however, this contradicts the assumption that logistics sprawl implies greater transport activity. Even though logistics facilities spread out, it would seem that industry reorganised itself so that this spreading did not impact transport activity.

These insights regarding the relationship between logistics sprawl and increased transport activity differ from what Methodology A offered. Sakai et al. (2017) also found that, contrary to the general assumption in logistics sprawl literature, logistics sprawl does not necessarily lead to greater transport activity.

6.3. Caveat to Methodology B

Methodology B improves on Methodology A by using metrics that control for the changes in the absolute volume of tonnes circulating in the economy as well as sample size. But one key caveat (discussed at length by Sakai et al. (2017)) is that the dataset cannot account for the impact of activity-chaining behaviour. In other words, only measuring the direct distances between facilities and their stakeholders does not account for how commercial vehicle activity chains had to change in response to sprawl. Methodology C takes into account the full ambit of commercial vehicle activity.



Area	Year	Mean ASD	Mean DOG	Median ASD	Median DOG
Gauteng	2010	15.74	10.57	8.05	3.61
	2014	16.92	11.21	8.59	3.51
	change	+1.18	+0.64*	+0.54	−0.10*
Cape Town	2010	13.93	9.40	7.20	3.67
	2014	13.11	8.86	6.83	3.47
	change	**	**	**	**
eThekwini	2010	13.25	8.99	6.45	3.70
	2014	12.98	8.79	6.73	3.83
	change	**	**	**	**

Fig. 8. Distributions of the ASD from facilities to their stakeholders compared to the distributions of facilities' DOGs.
* A t-test could not reject the null hypothesis that the means of the underlying populations are, in fact, equal. However, a WMW did reject the null hypothesis that the underlying distributions are similar.
** The WMW test and t-test could not reject the null hypotheses.

7. Methodology C: GPS trace calculations

The previous two methodologies were designed for different datasets available to the authors of those studies at the time. The limitations of those datasets were emulated as much as possible in Sections 5 and 6 for the purpose of comparing methodologies in this study. Methodology C is tailored to leverage the full detail of the commercial vehicle activity chains that emanate from the GPS traces.

7.1. Changes to activity-chaining behaviour

As the geography of logistics facilities and market forces adapt, transport operators would as well — always seeking the most efficient way to render their services. The reality of changing logistics behaviour riddles any simplistic analysis with two of the caveats already discussed: simple averages do not represent reality and distances to stakeholders do not really give insight into how vehicles are deployed and routed in the logistics space. The GPS dataset provides insight into how activity-chaining behaviour has changed in the four-year period.

7.1.1. Gauteng

The number of vehicles in the sample increased by 19% between 2010 and 2014. We track these vehicles' *entire* activity chains, which include kilometres travelled inside Gauteng (in-area) as well as kilometres travelled outside Gauteng (out-area). If there was no change in activity-chaining behaviour, a commensurate increase in in-area and out-area kilometres would be expected. However, the total in-area kilometres only increased by 9% while the out-area kilometres increased by 37%. These disproportionate changes suggest that the ratio of out-area facilities to in-area facilities is increasing in many activity chains.

To shed further light on the change in activity chain composition, we make a distinction between *in-area* and *out-area* activity chains. We define a chain as *in-area* if 50% or more of its kilometres were travelled inside Gauteng, otherwise it is an *out-area* chain. In 2010, 74% of all activity chains were exclusively in-area, meaning that 100% of their kilometres were travelled inside Gauteng. A further 9% of all activity chains were non-exclusive in-area chains, meaning that the majority (but not all) of their kilometres were travelled in Gauteng. The remainder of chains (17%) were out-area chains. While the percentage of non-exclusive in-area chains remained stable at 9% in 2014, the percentage of exclusively in-area chains dropped to 71% while out-area chains rose to 20%. Fig. 9 illustrates that out-area chains were typically much longer as they served a broader geographic range. This implies supply chain management in Gauteng is becoming more regionally-focussed instead of city-focussed.

Fig. 9 also shows that in-area activity chains became slightly longer with a statistically significant increase in the mean of 3.36% (119km→123km) while out-area chains showed statistically significant increases in the mean and median of 24.09% (921km→1143km) and 9.35% (622km→680km), respectively. Activity chains are thus becoming longer — especially those chains that serve a broader geographic area. This statistic — kilometres per activity chain — is indicative of a change in activity-chaining behaviour. It is *not* indicative of a change in transport activity. For that we need a different measure. But first we consider what happened in Cape Town and eThekwini.

7.1.2. Cape Town

The number of vehicles in the sample increased by 24% between 2010 and 2014. Similar to the case in Gauteng, the increase in in-area and out-area kilometres was not commensurate. Total in-area kilometres increased by 9% — similar to Gauteng — while the out-area kilometres increased by 19%. However, the underlying trends are different to Gauteng.

In 2010, 61% of all activity chains were exclusively in-area while 21% of chains were non-exclusive in-area chains. The remaining 18% were out-area chains. By 2014, 64% of activity chains were exclusively

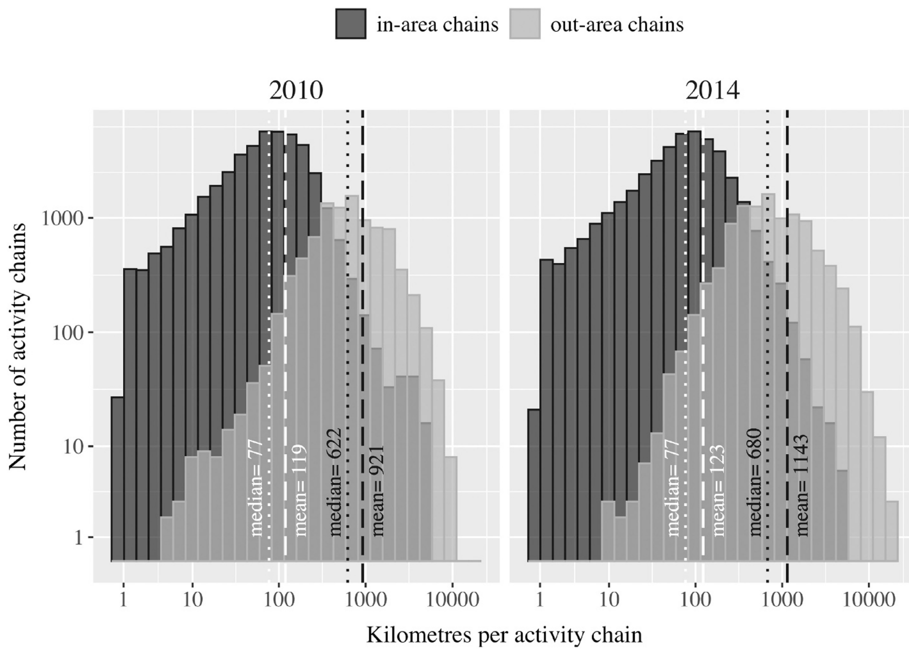


Fig. 9. The change in kilometres travelled per activity chain for in-area and out-area chains in Gauteng during the four-year period.

in-area with 18% of chains non-exclusive in-area chains. The percentage of out-area chains remained stable at 18%. Contrary to Gauteng, this suggests that more activity chains became exclusively city-focussed. Fig. 10 shows the very slight (but statistically insignificant) decrease in the kilometres per activity chain for in-area chains. But, similar to Gauteng, the out-area chains certainly got longer with increases in the mean and median of 17.6% (921km→1081km) and 12.09% (479km→537km), respectively.

7.1.3. eThekwini

eThewkini showed the smallest increase (13%) in the number of vehicles between 2010 and 2014. Similar to the other two study areas, out-area kilometres grew much more (22%) than in-area kilometres (4%).

One key difference between eThekwini and the other study areas is that most of the activity chains are out-area activity chains. In 2010, 6% of activity chains were exclusively in-area and only 14% were non-exclusive in-area chains. These values remained stable in 2014 with 5% exclusive and 14% non-exclusive in-area chains. Therefore, more than three quarters of the activity chains active in the study area perform the majority of their activities outside of the area. This observation partly explains why no logistics sprawl is observed in eThekwini. The bulk of logistics activity seems to happen outside the study area altogether.

The length of the out-area activity chains in eThekwini showed an increase in the mean and median of 19.78% (824km→987 km) and

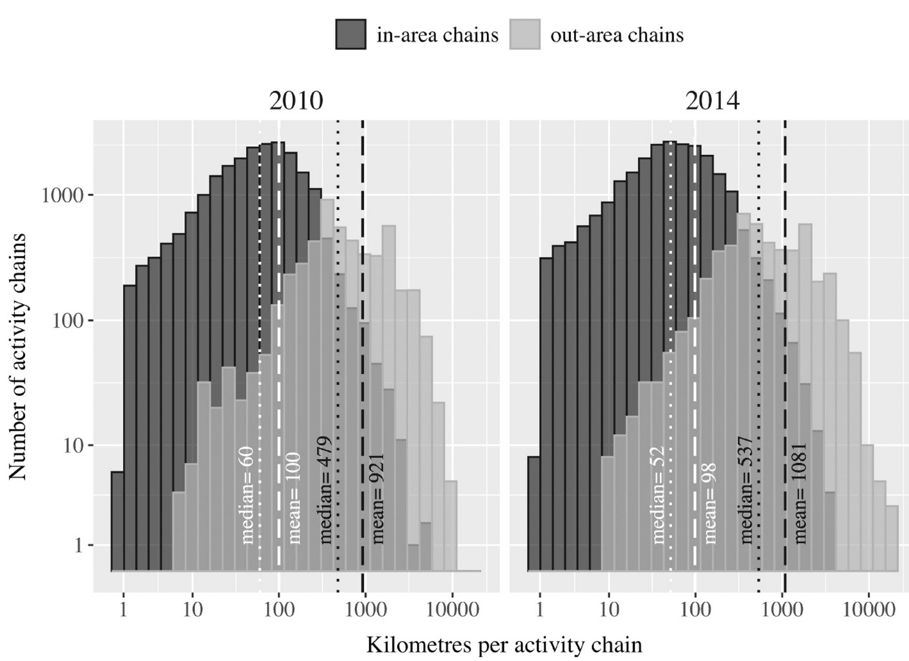


Fig. 10. The change in kilometres travelled per activity chain for in-area and out-area chains in Cape Town during the four-year period.

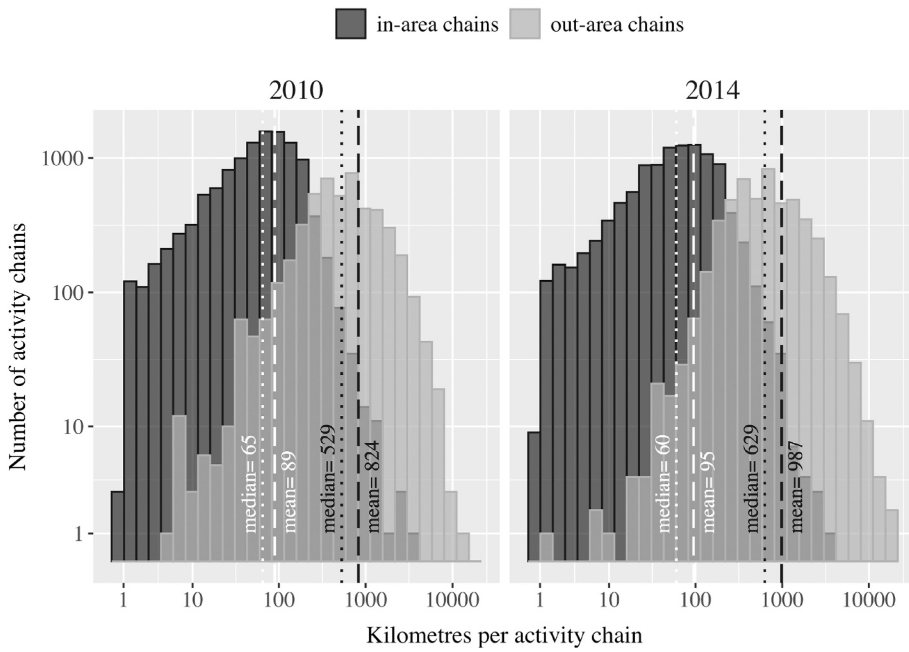
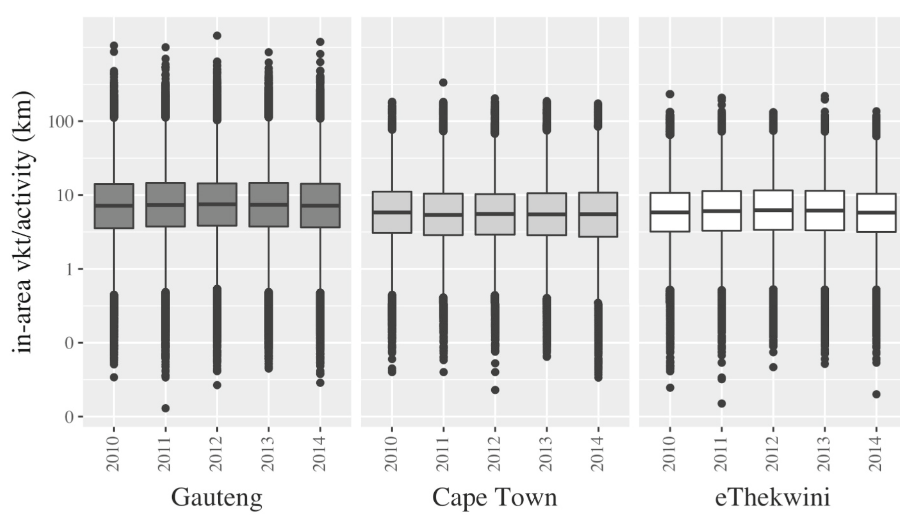


Fig. 11. The change in kilometres travelled per activity chain for in-area and out-area chains in eThekwini during the four-year period.

18.90% (529km→629km), respectively (see Fig. 11). This is in keeping with the trend towards longer out-area activity chains observed in the other two study areas. The mean of the lengths of the in-area chains increased slightly (6.74%: 89km→95km) while the median decreased (−7.69%: 65km→60 km). While these change for the in-area chains are small, they are statistically significant.

In summary, all three study areas showed a change in activity-chaining behaviour over the four-year period. In all three areas, the out-area chains were becoming longer. The proportion of in-area and out-area chains differed across the three areas as a result of their dissimilar urban and economic contexts. In Gauteng, more of the activity chains were performing the majority of their activities outside the study



Area	Year	Mean vkt/activity (km)	Median vkt/activity (km)
Gauteng	2010	13.36	7.16
	2014	13.55	7.17
	change	**	**
Cape Town	2010	10.37	5.82
	2014	9.76	5.52
	change	−0.62	−0.30
eThekwini	2010	9.98	8.95
	2014	5.83	5.78
	change	−1.04	−0.05

Fig. 12. Distributions of in-area vkt/activity.
** A WMW test could not reject the null hypothesis that the 2010 and 2014 population distributions were similar, nor could a t-test reject the null hypothesis that the means of the underlying populations are, in fact, equal.

area in 2014. By comparison, Cape Town's activity chains saw a widening split between city-focussed and region-focussed activity chains. The proportion of out-area chains remained stable while exclusively in-area chains grew. In eThekwini, the proportions of in-area and out-area chains remained stable. The next section investigates whether this change in activity-chaining behaviour was accompanied by a change in transport activity.

7.2. Change in transport activity

The length of activity chains is not an indicator of overall transport activity, it is merely an indicator of how transport operators execute their business. While our dataset does not contain load factors (to calculate vehicle-kilometres per tonne), we can calculate vehicle-kilometres per activity (vkt/activity). This is the measure we use to quantify transport activity.

Fig. 12 plots the distributions of the in-area vkt/activity over the four-year period. Statistical tests confirm that the distributions for Gauteng are nearly identical. This implies that there has been no change in transport activity in Gauteng over four years, despite the fact that logistics facilities spread further away from each other.

In the case of Cape Town and eThekwini, the statistical tests confirmed that the in-area vkt/activity decreased over four years. But Cape Town experienced logistics sprawl whereas eThekwini did not experience conclusive sprawl or concentration. The results for Cape Town contradicts the idea that logistics sprawl results in increasing transport activity while the results for eThekwini questions the relationship altogether. But even this methodology is not without its caveats.

7.3. Caveat to Methodology C

The key caveat to this methodology is the lack of information about trip purpose and loading factors. This impacts the analysis in two ways. Firstly, knowledge about loading factors would offer a further layer of insight as vkt/tonne and tonne-kilomentres could then be evaluated in conjunction with vkt/activity. Secondly, trip purpose data would refine the facility locations extracted from the GPS traces.

The three methodologies generated different results when evaluating the impact of logistics sprawl on transport activity. In the next section, we discuss these contradictions and their implications to the study of logistics sprawl.

8. Discussion and conclusion

This paper set out to address two research questions:

1. *How does the relationship between the change in transport activity and logistics sprawl differ depending on the methodology used?*
2. *What are the implications of using different methodologies to evaluate the relationship between logistics sprawl and a change in transport activity?*

To this end, we applied three different methodologies to the same dataset of GPS traces to quantify the relationship between logistics sprawl and a change in transport activity.

8.1. Logistics sprawl in the study areas

Over the four-year period from 2010 to 2014, the centrality of logistics activity shifted in each of the study areas as indicated by a shift in the centroid. Although most studies in the field are limited to locating an unweighted centroid, GPS traces allowed us to locate a centroid weighted by outgoing trips per facility. This additional layer of detail proved useful in analysing the shift of the logistics activity.

Logistics sprawl was quantified using the change in the mean distance to the weighted centroid. The extent of sprawl observed in Gauteng and Cape Town is comparable to the trends in other global cities. Also similar to global trends, facilities that generate more logistics activity (and are thus presumably larger) are sprawling further than the rest, most likely in search of available and affordable land. In eThekwini, there was no conclusive evidence of logistics sprawl, but the results did show that more productive facilities also tend to the outskirts.

Given that logistics sprawl did occur in two of the country's three largest urban areas, we proceeded to address the first research question.

8.2. Comparing methodologies

Methodology A assumes a priori that an increase in the distance to the weighted centroid results in an increase in transport activity as the former is used to calculate the latter. Thus, according to that methodology, transport activity increased in Gauteng and Cape Town. This was the first methodology that empirically measured the transport-related impact of logistics sprawl. It can be argued that the methodology was suited to the dataset and the context of the Paris study, but it is not an appropriate methodology when studying the logistics industry in general and especially not in polycentric urban areas.

Methodology B improves on Methodology A by considering facility-specific impacts of sprawl. Instead of using the change in the distance to the centroid of *all* facilities, they monitor the change in the distances between specific facilities and their stakeholders. When adapted for the dataset of GPS traces, Methodology B finds that, on average, the transport activity (measured by ASD) did increase for each facility in Gauteng. In Cape Town and eThekwini, the change in transport activity was not statistically significant. While this is expected for eThekwini (where there was no sprawl), Cape Town's result contradicts the general assumption that logistics sprawl leads to increased transport activity.

Methodology C leverages the full capability of the GPS dataset. This allows an insight into the changes in activity-chaining behaviour. There are differences in activity-chaining across the three areas. These are driven by the individual geographic and economic realities. But, overall, activity chains are becoming longer and more regionally-focussed.

Methodology C measures transport activity in terms of vkt/activity and the results question the link between logistics sprawl and a change in transport activity altogether. In Gauteng, there was no significant change in transport activity despite logistics sprawl; in Cape Town, transport activity decreased despite logistics sprawl; and in eThekwini, there was a significant reduction in transport activity even though there was no conclusive logistics concentration.

Answering the first research question, there is a vast difference in the conclusions drawn regarding the link between logistics sprawl and a change in transport activity depending on the methodology used. This leads us to a discussion of the second research question.

8.3. The implication of using different methodologies

A pivotal assumption of Methodology A is that all transport activity is drawn towards the centroid of logistics facilities. While this assumption is defendable in the context in which it was developed (parcel deliveries from outlying terminals into a densely populated monocentric urban area), it is not applicable to studying the link between logistics sprawl and transport activity in a general sense. Applied generally, it ignores that: not all cities are monocentric; the geographic spread of a facilities' suppliers and customers can differ between industry sectors and is not necessarily around the centroid of the logistics facilities; and that vehicle utilisation, fleet compositions, and routing are complex and adaptable in response to a sprawling freight landscape. When applied to the three study areas in South Africa, Methodology A suggests that logistics sprawl results in greater transport activity and should thus be regulated.

Methodology B takes a facility-specific approach to quantifying transport activity which accommodates dissimilar urban forms (i.e. monocentric or polycentric) and distinctions between industries. In this methodology, the measurement of the change in transport activity is effectively de-coupled from the measurement of logistics sprawl which allows a truer reflection of the relationship between the two phenomena. The metrics developed in this methodology are based on shipment data, which, to a great degree, accounts for decisions regarding vehicle utilisation and fleet composition. The implications of these differences are well-illustrated when comparing the results of Methodology A and B for Cape Town.

Cape Town is a polycentric agglomeration of independent towns. While the city centre itself developed around the port, the other towns developed around agriculture and tertiary education. Industrial sectors are clustered in distinct areas of the urban landscape, for example: the city centre bustles with service-related businesses spanning finance, media, hospitality and conferencing; manufacturing and distribution activities are pushed to the outskirts of the densely populated city and gravitate around the regional transport arteries; a concentration of technology firms exists around Stellenbosch; wholesale and retail centres appear in the northern suburbs; and agriculture and agriculture processing is a mainstay of the areas that were, initially, farming communities. In contrast to Methodology A, Methodology B can accommodate this variation in urban form and industrial organisation when measuring transport activity. Methodology B captures the impacts of supply chains adapting and reorganising in concert with their customers and suppliers. In so doing, it suggests that the potential increases in transport activity were offset as facilities repositioned themselves with regard to their stakeholders. Therefore, industry seems self-regulating in this regard and policy intervention may be unnecessary.

Methodology C extends the capabilities of Methodology B by incorporating explicit data regarding vehicles' activity-chaining behaviour and routing. This is a final level of detail that is made possible by the use of GPS traces. This additional level of detail has a marked impact on the results. Once again using the case of Cape Town as an example, Methodology B showed that there was no significant increase in transport activity despite logistics sprawl whereas Methodology C shows that there was a significant *decrease* despite logistics sprawl. This difference in results shows that the adaptation of transport operators, in addition to the reorganisation of supply chain stakeholders, resulted in efficiencies that outweighed the impact of sprawling facilities. The policy signal in this case would be that regulating logistics sprawl to reduce transport activity is probably counterproductive.

Three different methodologies using the same dataset for the same study areas resulted in three markedly different conclusions. This is a call to researchers (the authors included) to take extreme care when choosing methodologies. These insights also underline the value of including empirical data of actual vehicle movements when studying the link between logistics sprawl and transport activity.

8.4. Limitations of the study and future work

Two important limitations of this study should be kept in mind. Firstly, the sample size of the commercial vehicle GPS traces is relatively small. Although the dataset tracks many thousands of vehicles, this is a small percentage of the national fleet. In addition, we cannot say with certainty that the sample is a representative cross-section of the vehicle population in terms of size, type, and industry. Secondly, the identification of logistics facility locations from GPS traces was scrutinised at length in Section 3.

In response to these limitations, we point out that the same limitations impacted each of the methodologies applied. All three methodologies used the same sample of data and all used the set of facility locations as they are. We also point out that approximations in terms of facility location, type, and size are not exceptional in logistics sprawl studies as discussed in Section 3. Notwithstanding, these limitations present opportunities for future work.

In Section 3.1 we outline a strategy for refining how facilities are identified from GPS traces by considering the ratio of activities per unique vehicle at each facility cluster. Another extension of this work is to increase the sample size, representativity, and longitudinal span of the data. While the dataset in this study is sufficient to address the research questions, a study aimed at providing extensive urban planning and transport policy input for South Africa requires a richer sample.

Moving beyond extensions of this study, this work starkly questions whether there is an empirical link between logistics sprawl and transport activity at all. This opens avenues for future research. These findings should be confirmed or challenged by similar studies conducted in other urban areas around the world. Furthermore, it would be worthwhile to understand the underlying logistics behaviour in urban areas and how this adapts to sprawl.

Advances in technology (such as smart sensors and Internet-of-Things) make it possible to study deeper layers of logistics behaviour — assuming that the data would be made available by industry. Adding data about axle loads to GPS traces would illuminate vehicle utilisation decisions. Combining process-related data from the loading docks of factories, warehouses, and retail facilities would elucidate the efficiencies of logistics activities. Accessing data that track driver behaviour would show how the human element shapes transport activity. These are the most immediate ideas of how these technologies could advance this work.

Some of these areas of future work will be pursued by the authors, but the scope of opportunities presented here are an open call for researchers in the urban logistics domain.

References

Aljohani, K., Thompson, R.G., 2016. Impacts of logistics sprawl on the urban environment and logistics: taxonomy and review of literature. J. Transp. Geogr. 57, 255–263.

Andreoli, D., Goodchild, A., Vitasek, K., 2010. The rise of mega distribution centers and the impact on logistical uncertainty. Transport. Lett. 2 (2), 75–88.

Automotive Fleet, 2015. Telematics Market in South Africa to Reach 840k by 2017. Automotive Fleet. Available online. https://www.automotive-fleet.com/128848/telematics-market-in-south-africa-to-reach-840k-by-2017 (Accessed: 9 November 2018).

BergInsight, 2015. Fleet management in South Africa. In: Technical report. BergInsight.

Bowen, J.T., 2008. Moving places: the geography of warehousing in the US. J. Transp. Geogr. 16 (6), 379–387.

Cidell, J., 2010. Concentration and decentralization: the new geography of freight distribution in US metropolitan areas. J. Transp. Geogr. 18 (3), 363–371.

Cidell, J., 2015. Distributions centers as distributed places: Mobility, infrastructure and truck traffic. In: Birtchnell, T., Savitzky, S., Urry, J. (Eds.), Cargomobilities: Moving Materials in a Modern Age. Routledge, New York.

Dablanc, L., Rakotonarivo, D., 2010. The impacts of logistics sprawl: how does the location of parcel transport terminals affect the energy efficiency of goods' movements in Paris and what can we do about it?. In: Procedia - Social and Behavioral Sciences, 2, pp. 6087–6096.

Dablanc, L., Ogilvie, S., Goodchild, A., 2014. Logistics sprawl differential warehousing development patterns in Los Angeles, California, and Seattle, Washington. Transp. Res. Rec. 2410, 105–112.

De Beer, D. and Joubert, J. W. (n.d.). Evolutionary optimisation of large-scale activity clustering with increased automation. Working paper 082: Centre for Transport Development, University of Pretoria.

Electronic national administration traffic information system (eNaTIS), 2014. Vehicle population statistics for January/February 2014. In: eNaTIS, South Africa. Available online from. http://www.enatis.com/index.php/statistics/13-live-vehicle-population/383-vehicle-population-statistics-for-januaryfebruary-2014 (Accessed 9 November 2018).

Gardrat, M., 2021. Urban growth and freight transport: from sprawl to distension. J. Transp. Geogr. 91, 102979.

Guerin, L., Vieira, J.G.V., de Oliveira, R.L.M., de Oliveira, L.K., de Miranda Vieira, H.E., Dablanc, L., 2021. The geography of warehouses in the São Paulo metropolitan region and contributing factors to this spatial distribution. J. Transp. Geogr. 91, 102976.

Gupta, S., Garima, 2017. Logistics sprawl in timber markets and its impact on freight distribution patterns in Metropolitan City of Delhi, India. In: Transportation Research Procedia, 25, pp. 965–977.

He, M., Shen, J., Wu, X., Luo, J., 2018. Logistics space: a literature review from the sustainability perspective. Sustainability (Switzerland) 10 (8), 2815.

He, M., Zeng, L., Wu, X., Luo, J., 2019. The spatial and temporal evolution of logistics enterprises in the Yangtze River Delta. Sustainability (Switzerland) 11 (19).

Heitz, A., Dablanc, L., Tavasszy, L.A., 2017. Logistics sprawl in monocentric and polycentric metropolitan areas: the cases of Paris, France, and the Randstad, the Netherlands. Region 4 (1), 93–107.

Heitz, A., Launay, P., and Beziat, A. (2019). Heterogeneity of logistics facilities: an issue for a better understanding and planning of the location of logistics facilities. Eur. Transp. Res. Rev., 11(1):1–20.

Heitz, A., Dablanc, L., Olsson, J., Sanchez-Diaz, I., Woxenius, J., 2020. Spatial patterns of logistics facilities in Gothenburg, Sweden. J. Transp. Geogr. 88, 102191.
Jaller, M., Pineda, L., Phong, D., 2017. Spatial analysis of warehouses and distribution centers in Southern California. Transp. Res. Rec. 2610 (1), 44–53.
Joubert, J.W., Axhausen, K.W., 2011. Inferring commercial vehicle activities in Gauteng, South Africa. J. Transp. Geogr. 19, 115–124.
Joubert, J., Axhausen, K.W., 2013. A complex network approach to understand commercial vehicle movement. Transportation 40 (3), 729–750.
Joubert, J.W., Meintjes, S., 2015a. Computational considerations in building inter-firm networks. Transportation 42 (5), 857–878.
Joubert, J.W., Meintjes, S., 2015b. Repeatability & reproducibility: implications of using GPS data for freight activity chains. Transp. Res. B 76, 81–92.
Kahle, D., Wickham, H., 2013. ggmap: spatial visualizations with ggplot2. R J. 5 (1), 144–161.
Kang, S., 2020a. Relative logistics sprawl: measuring changes in the relative distribution from warehouses to logistics businesses and the general population. J. Transp. Geogr. 83, 102636.
Kang, S., 2020b. Warehouse location choice: a case study in Los Angeles, CA. J. Transp. Geogr. 88, 102297.
Kang, S., 2020c. Why do warehouses decentralize more in certain metropolitan areas? J. Transp. Geogr. 88, 102330.
Krzysztofik, R., Kantor-Pietraga, I., Spórna, T., Dragan, W., Mihaylov, V., 2019. Beyond ‘logistics sprawl’ and ‘logistics anti-sprawl’. Case of the Katowice region, Poland. Eur. Plan. Studies 27 (8), 1646–1660.
Kumhálová, J., Ružicka, M., Lopéz, E.C., Chyba, M., 2019. Logistics sprawl in Prague’s suburb from satellite images. In: TAE 2019 - Proceeding of 7th International Conference on Trends in Agricultural Engineering 2019, pp. 319–324.
Lindsey, C., Mahmassani, H.S., Mullarkey, M., Nash, T., Rothberg, S., 2014. Industrial space demand and freight transportation activity: exploring the connection. J. Transp. Geogr. 37, 93–101.
Ma, X., Wang, Y., McCormack, E., Wang, Y., 2016. Understanding freight trip-chaining behavior using a spatial data-mining approach with GPS data. Transp. Res. Rec. J. Transp. Res. Board 2596, 44–54.
Sakai, T., Kawamura, K., Hyodo, T., 2015. Locational dynamics of logistics facilities: evidence from Tokyo. J. Transp. Geogr. 46, 10–19.
Sakai, T., Kawamura, K., Hyodo, T., 2016. Logistics facility distribution in Tokyo Metropolitan Area: experiences and policy lessons. Transp. Res. Proc. 12, 263–277.
Sakai, T., Kawamura, K., Hyodo, T., 2017. Spatial reorganization of urban logistics system and its impacts: case of Tokyo. J. Transp. Geogr. 60, 110–118.
Sakai, T., Kawamura, K., Hyodo, T., 2020. Logistics facilities for intra and inter-regional shipping: spatial distributions, location choice factors, and externality. J. Transp. Geogr. 86, 102783.
Statistics South Africa, 2016. Community survey 2016 provinces at a glance. In: Technical Report ISBN 978–0–621-44666-1. Statistics South Africa, Pretoria, South Africa.
Strale, M., 2020. Logistics sprawl in the Brussels metropolitan area: toward a socio-geographic typology. J. Transp. Geogr. 88.
Trent, N.M., Joubert, J.W., Gidofalvi, G., Kordnejad, B., 2020. A matching algorithm to study the evolution of logistics facilities extracted from GPS traces. Transp. Res. Proc. 46, 237–244.
Viljoen, N.M., 2018. Quantifying supply chain vulnerability using a multilayered complex network perspective. PhD thesis. University of Pretoria, South Africa.
Viljoen, N.M., Joubert, J.W., 2019. Supply chain micro-communities in urban areas. J. Transp. Geogr. 74, 211–222.
Wagner, T., 2010. Regional traffic impacts of logistics-related land use. Transp. Policy 17 (4), 224–229.
Wygonik, E., Goodchild, A.V., 2018. Urban form and last-mile goods movement: factors affecting vehicle miles travelled and emissions. Transp. Res. Part D: Transp. Environ. 61, 217–229.
Yang, X., Sun, Z., Ban Xuegang, J., Holguín-Veras, J., 2014. Urban freight delivery stop identification with GPS data. Transp. Res. Rec. J. Transp. Res. Board 2411, 55–61.
Yuan, Q., 2018. Environmental justice in warehousing location: state of the art. J. Plan. Lit. 33 (3), 287–298.