Australasian Transport Research Forum 2022 Proceedings
28-30 September, Adelaide, Australia
Publication website: http://www.atrf.info

Network resilience analysis with large streams of mobility data and Open-Source software

Pedro Camargo, Ph.D. [1]

[1]Outer Loop Consulting, Argonne National Laboratories

Email for correspondence: pedro@outerloop.io

Abstract

GPS-based mobility data such as vehicle tracking data and Location-Based Services (LBS) have seen a surge in use in the last decade, but struggles with sample sizes, biases and new privacy laws pose many challenges to its use. This research, based on over three months of vehicle tracking data for the entire commercial vehicle fleet of Vietnam, over 700 thousand vehicles, is a unique opportunity to show how network resilience models can leverage this type of data. We propose a methodology based solely on open-source software and implement it for the province of *Long An* in Vietnam, for which four use cases are analyzed.

1 Introduction

Transport network infrastructure is essential for the growth and prosperity of cities and regions and are also vital for access to critical services such as health and education. Thus, transportation infrastructure that is efficient and reliable is the backbone of all the economic and human value enabled by transportation.

The value of each one of a city's infrastructure components depends on the extent that they are utilized and how critical they are to the system's performance (e.g. whether there are reasonable alternatives to each asset analysed). Although it is reasonable to posit that failure in pieces of infrastructure that represent natural bottlenecks such as bridges and tunnels would cause a lot more impact than the failure of a small street in a residential suburb, it is often difficult to quantify these differences and prioritize maintenance accordingly.

In practice, the impact resulting from the failure of a piece of infrastructure is not only a matter of how many people/vehicles use that piece of infrastructure, but whether reasonable alternative routes are available and the quality of those alternatives, which result in varying levels of additional time, distance & emissions generated by vehicles taking those alternative routes.

Thus, this type of analysis requires not only a complete representation of the network, but also a view of the demand using it that includes its origins and destinations. While the former has become trivial to obtain through Open-Street Maps (OSM) and commercial vendors, the latter has remained dependent on much bigger efforts, be data in data collection or modelling.

1.1 Commercial Vehicle Tracking System data

The Commercial Vehicle Tracking System (CVTS) consists of round-the-clock GPS tracking of **all commercial vehicles** in Vietnam, and provides continuous GPS traces (latitude, longitude and timestamp) for each vehicle that is part of the system, as well as a persistent unique identifier for the vehicle and an aggregate vehicle class (e.g. taxi, bus, etc.). As such, the CVTS data represents both a general mobility class of data assets, but also a unique big data resource in Vietnam, with over 1.2 billion high-frequency vehicle position feeds per day, during the 2020 COVID lockdowns.

Given that the CVTS is a national public sector digital asset, its use for decision support also needed to adhere to prevailing Vietnamese data governance stipulations and privacy protection. This included ensuring that analytics and visualization capabilities could be deployed in such a way that they adhered to Vietnam data localization requirements, but also allowed for use-case applications to be collaboratively refined across domestic and international based experts.

For this reason, all work developed with this data source was performed on a cloud service physically located in Vietnam, although access to the raw data will remain limited to a few consultants and researchers working with the World Bank on projects related to CVTS.

The data used in the study spans approximately 100 days from March to June 2020 and contains tracking data for over 760 thousand unique commercial vehicles across Vietnam, which was leveraged for a series of analytical exercises, including the one presented in this paper. All data and results presented in this paper are therefore limited to the universe of **commercial vehicles** operating in Vietnam.

2 Literature review

2.1 Network resilience models

Network resilience models, often referred to as link criticality models, based on transportation modeling frameworks, particularly the 4-step model, are rather common in the literature and in consulting practice.

A recent practical example can be found in a 2018 study by the World Bank in Mozambique (Espinet, Rozenberg, Rao, & Ogita 2018), where the transportation team used a Network resilience model to evaluate the network resilience to floods in rural areas and their impact on the freight distribution system and overall economy.

In Bangladesh, another study (Das 2020) applied a similar model, this time focused on estimating long-lasting disruptions. In that case, the researchers incorporated an estimation of changes in vehicle demand flows caused by network disruptions, as that is a likely consequence in such long-lasting disruptions.

Other more theoretical studies have also explored the use of transportation planning models for the computation of link criticality indices and network resilience have also demonstrated the relevance of this type of model in ensuring transportation networks are resilient.

One such example was a study (Camargo 2021) done for Dili, Timor Leste, where the author was able to identify one set of streets around a major intersection that, if disrupted, would completely cut off to critical services such as hospitals and schools from thousands of people.

More sophisticated models, such as an economic analysis of network disruptions done a team of researchers from the World Bank for Tanzania (Colon, Hallegatte, & Rozenberg 2019), also have their transportation and network analysis bedrock in transportation planning frameworks such as the 4-step model, confirming the robustness and capability of this type of model.

The weakest point of all these studies, however, was the estimation of the transportation demand, as reliable ground-truth data was not available in any of these cases. This is where CVTS puts Vietnamese jurisdictions in the remarkable position of knowing the entirety of their commercial vehicle travel demand.

2.2 Zoning systems

The problem of geographical resolution and aggregation when modelling spatial phenomena (e.g. transport) has long been the subject of research (Openshaw 1977). In the last couple of decades, however, several studies have focused on the effect that zoning systems may have on particular aspects of the transportation modelling process, including trip generation models (Ghadiri, Rassafi, & Mirbaha 2019), freight models (Sahu, Chandra, Pani, & Majumdar 2020) or the overall modelling process (Martínez, Viegas, & Silva 2009).

The motivation for most of the papers in this area is around the minimization of the impact of the zoning system on the main object of their research (e.g. traffic forecast, freight modelling, etc.), as shown in detail on more comprehensive literature reviews on this topic (Martínez, Viegas, & Silva 2009).
Pure GIS techniques that do not have known transportation phenomena (observed or modelled trips) as input, seem to have fallen out of favour. However, a couple of older papers (Bennion & O'Neill 1994) and (You, Nedović-Budić, & Kim 1998) do provide a good blueprint on practical algorithms.
(You, Nedović-Budić, & Kim 1998) in particular, proposes a clustering-based algorithm that aims to create a zone system with many desirable properties, such as contiguity and homogeneity, although some of its underlying methods would require substantial effort to be implemented.

2.3 Deriving vehicle trips from GPS data

Substantial research on deriving vehicle stops from commercial vehicle GPS data has been published in the last 12 to 15 years as large data streams became available for research, particular in the USA, and vehicle stops became one of the main products of this type of data for planning purposes (Bassok, McCormack, Outwater, & Ta 2011).
To this end, a range of techniques have been applied to the identification of truck stops, including sophisticated hidden Markov chains (Taghavi, Irannezhad, & Prato 2019) and kernel density functions (Thierry, Chaix, & Kestens 2013) and simple rule-based processing (Camargo, Hong, & Livshits 2017; Yang et al. 2022), both reporting excellent results.
As this field is extremely broad and deep and only tangential to this research, we recommend (Yang et al. 2022) for a recent and in-depth literature review of this topic.

3 Methodology and data sources

Transportation engineers and planners have developed a range of workhorse resilience analytics models based on the 4-step modelling process, and the methodology presented in this paper builds on its main principles. The specific focus of the specific methodology we propose, however is to obtain a procedure that could be replicated for any region in Vietnam or elsewhere with minimum manual intervention. In other words, we propose automated processes to build the portions of the model relevant to the criticality analysis using the CVTS dataset.
As the model structure is loosely based on the 4-step model (McNally 2008), is useful to revisit its components and identify exactly which portions are relevant to the present methodology:

1. **Trip generation**: How many trips are produced/attracted in each point/zone/region of your region
2. **Trip distribution**: Connects origins and destinations of trips to create actual trips
3. **Mode split/choice**: Separates the trips by the mode they are using (public transport, walk/bike or auto for individual transport, or rail, truck, air, water for freight)
4. **Traffic assignment**: Finds the paths used for each trip between their origins and destinations. What is important to know here is that computational time increases with the number of trip origins and the number of links in the network

As the objective of this analysis is to compute the impact of network disruptions for a given set of **known** vehicle trips derived from the CVTS data, the three first steps of this methodology are not relevant for this analysis. Yet, using the overall modeling and data frameworks regularly applied to simpler 4-step models is computationally convenient if we are to use traffic assignment, particularly the use of Traffic Analysis Zones (TAZs) as a tool to abstract all flows to/from a zone into a single node in the network.

With that in mind, there are two critical steps to the model build after the construction of the road network from Open-Street Maps (OSM) data:

- Definition of a zoning system
- Derivation of vehicle trips from the CVTS data

The approach we have taken for the computation of link criticality per-se is trivial, but we also discuss in some level of detail

3.1 Zoning system

As mentioned above, one simplification used in 4-step models and that is not immediately clear through the description of each step is the use of Traffic Analysis Zones, which consolidate on a single point (their *centroids*, in the modeling jargon) all the trips with origins and destinations within their areas.

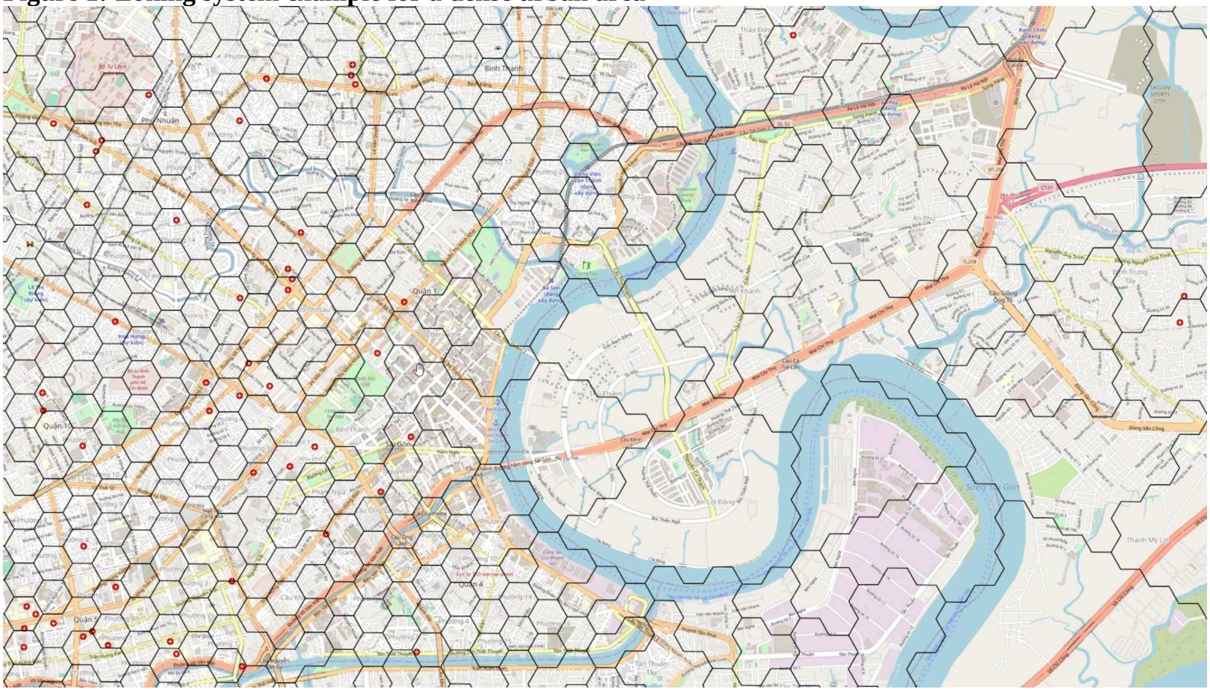
This simplification reduces the number of possible origins and destinations from an arbitrarily large number down to a fixed number of zones, and greatly reduces the computational time of finding paths for all vehicles in the network.

The definition of a zone system is, traditionally, a somewhat manual exercise of aggregating adjacent geographical subdivisions such as census tracts into zones such that the level of detail is compatible with the network density and desired precision. Small land subdivisions were not readily available for Vietnam, however, nor is it reasonable to rely on laborious manual exercises.

Motivated by the dimension of the modeling area we are dealing with (the entire Vietnam), we have adopted the methodology developed in a recent study (Camargo 2021) to define our base zoning system.

In a nutshell, this methodology consists in building a mesh of small hexagonal polygons covering the area of interest and computing the estimated population of each polygon from raster data (*Open Spatial Demographic Data and Research - WorldPop* n.d.). With these “micro-zones” in hand, we can recursively apply spatial clustering algorithms to create bigger zones with a maximum population per zone.

Figure 1: Zoning system example for a dense urban area



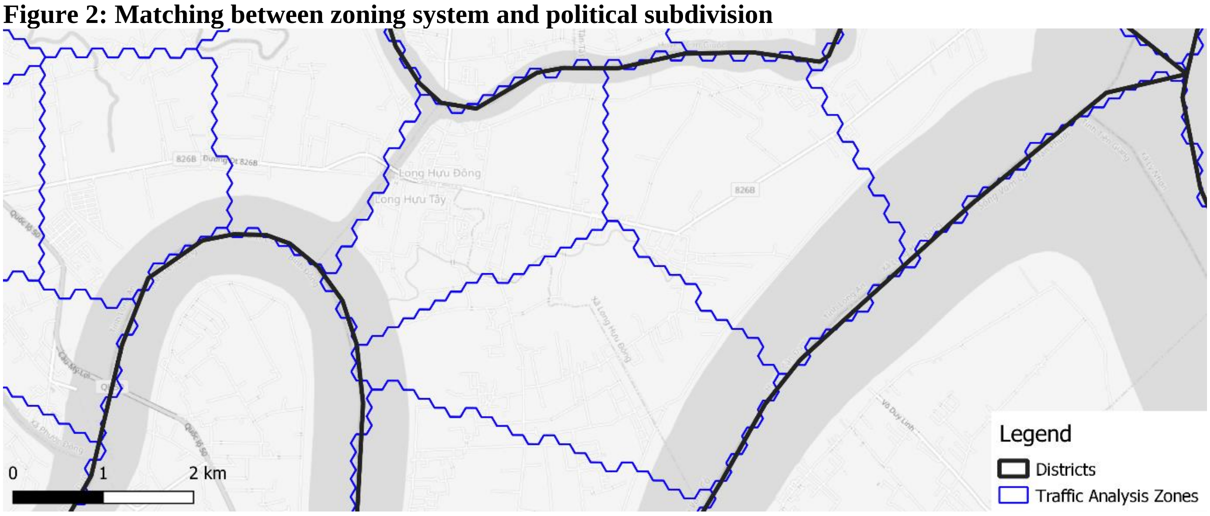
Source: Open-Street Maps, Outer Loop Consulting

With a population-based criteria, we obtain a zoning system that is more detailed in heavily populated areas and much coarser in unpopulated regions of the country (**Figure 1**), which generally reflects very well the amount of vehicle trips starting or ending in a region and the population and activity centers located there.

For this exercise, we established that zones should have a population of up to 10,000 people. Automated experimentation resulted in a base layer of hexagonal polygons (micro-zones) such that their population was, at most, 10,000 when overlaid with the WorldPop population layer.

When applying the spatial clustering procedures, we limited the clustering to base hexagons that were (mostly) in the same district (or equivalent political subdivision) and prevent disconnect zones from existing. As a result, many zones with much smaller populations, as low as 500, were also created, but zone aggregations do a remarkable job in matching the zones., as shown on **Figure 2.**

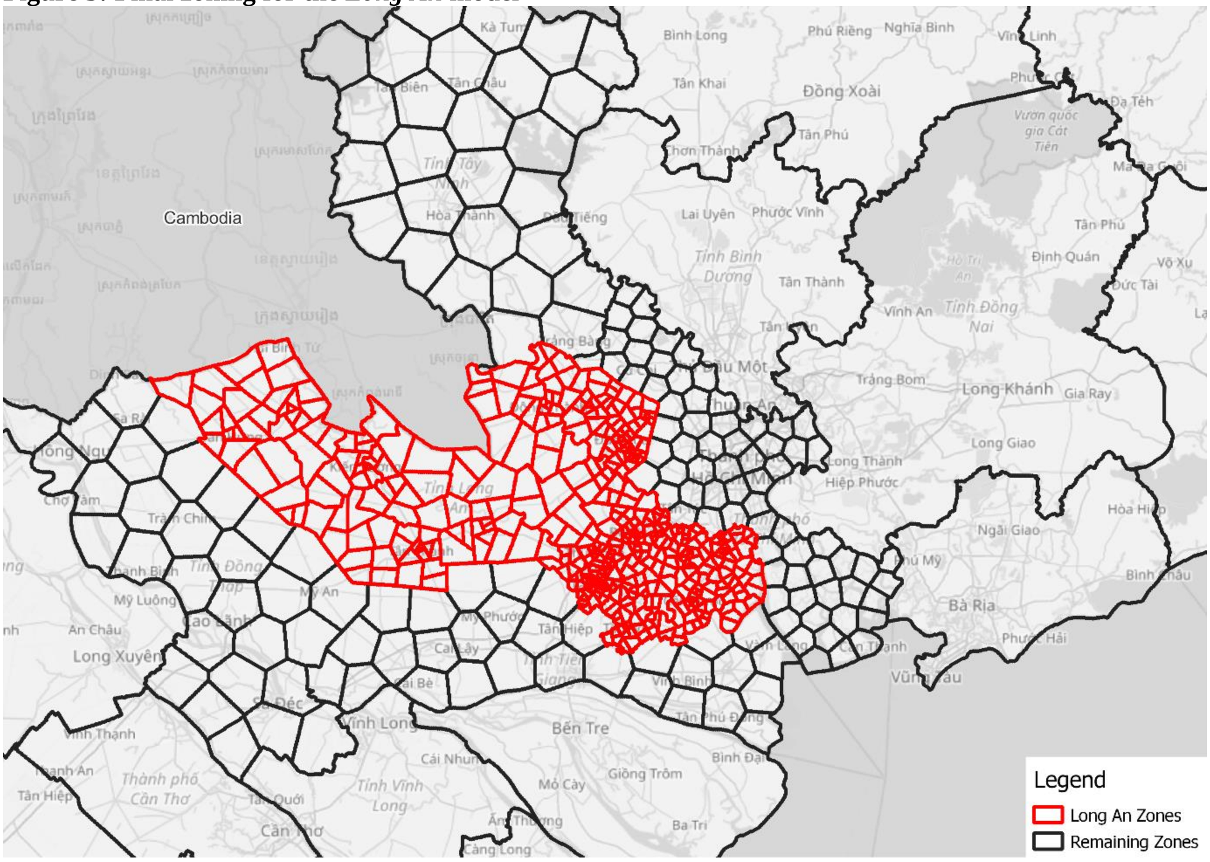
Figure 2: Matching between zoning system and political subdivision



Source: Open-Street Maps, Outer Loop Consulting

Since this study was a Proof-of-Concept limited to the province of *Long An*, we have modified the clustering rules for micro polygons covering provinces other than *Long An*, clustering neighboring provinces into 50 physically homogenous zones regardless of population, and aggregating entire provinces as a single zone for all other provinces, for a total of just under 700 zones for the entire *Long An* model, as shown on **Figure 3**, below.

Figure 3: Final zoning for the *Long An* model



Source: Open-Street Maps, Outer Loop Consulting

3.2 Deriving a trip matrix from CVTS data

As briefly discussed in the literature review, there are many suitable methods for the identification of vehicle stops from GPS feeds. One issue not often discussed in the literature is computational performance and implementation details, as data sets are frequently in the few hundred million of GPS points.

The data set we were working with, however, is 3 orders of magnitude bigger than other real-world applications we have seen in the literature, while the cloud-based server available to process this data could be considered small for today’s standard, at 32 logical CPUs and 64Gb of RAM.

With this in mind, more complex procedures that required too much memory or were potentially too slow would not be feasible for this application. Therefore, we have opted for the rule-based approached described in (Camargo, Hong, & Livshits 2017) with a few minor changes to the original code for better computational performance. This method allowed us to process the complete dataset in under 70h.

The result of this procedure was stored with the original coordinates for each stop, allowing them to be aggregated into any arbitrary zoning system. Although this aggregation process was substantially slower than expected (about 2h of processing time), it allows for its final result to

feed into any zoning system, potentially feeding into multiple models for individual provinces without requiring re-identification of vehicle stops.
The creation of demand matrices from the CVTS data was handled by arranging each pair of sequential stops by a vehicle on a trip table and did not require relevant computational effort.

3.3 Building the model

One of the advantages of using an established framework such as the 4-step model is that there is a multitude of software designed for this specific use. Among other things, using an off-the-shelf software simplifies the decisions otherwise required around the data model, scenario handling and file structure to be utilized.
In our case, we have chosen AequilibraE as the transportation modeling software, and therefore we have adopted its data model for both the network, zoning system and traffic data (i.e. vehicle OD matrix). The documentation for its data model can readily be accessed in AequilibraE's online documentation (AequilibraE Developers 2022).
Having established the procedures for generating the demand data and the zoning system, it is still necessary to build the transportation network for the model, for which we have used Open-Street-Maps (OSM).
The decision to use OSM as the base network for this study was the lack of official dataset and the apparent good state of completeness of the OSM network for Vietnam as a whole. Consequently, this study also shows the feasibility of performing this type of analysis based on OSM data, even though it may not be perfectly reflective of the actual extent of the road network in some cases.
The import and conversion of the OSM into the AequilibraE data model is handled by AequilibraE itself in just a few hours. From the original Vietnam OSM data import we have also maintained another 300,000 links with a total of 32,000km to provide connectivity to every other province in the country.

3.4 Computing link criticality

Once the transport model has been assembled, one can compute an estimate of traffic loads on each link of the network by applying a *traffic assignment* procedure to load the vehicle demand onto the network, and its result can be understood as the base-case.
With the base-case on hand, one can compute the total vehicle-kilometers (VKT) and vehicle-hours (VHT) in the network by multiplying the vehicle loads in each link by the links' respective distances and estimated travel times.
Should the demand be stratified by vehicle type, day of the week or time of the day, VKT and VHT results can be computed for each segment of the data.
With the base case computed, the impact of any change to either the network or the demand can be easily measured by re-applying the process used for the base-case to this demand/network scenario.
In the case of link criticality, the objective is to compute the impact that removing each link of the network would have, given that original known vehicle demand. So, by removing one link at a time and re-applying the process used in the base case, we can measure the total VKT and VHT for the disrupted network and the differences in comparison with the base case and identify the most critical links.
As no source for personal travel demand was available, the current model does not address congestion effects and relies on simple *all-or-nothing* traffic assignment.
Finally, the decision of modelling static demand was intentional, as the focus of the study was on short-term network disruptions.

4 Use cases

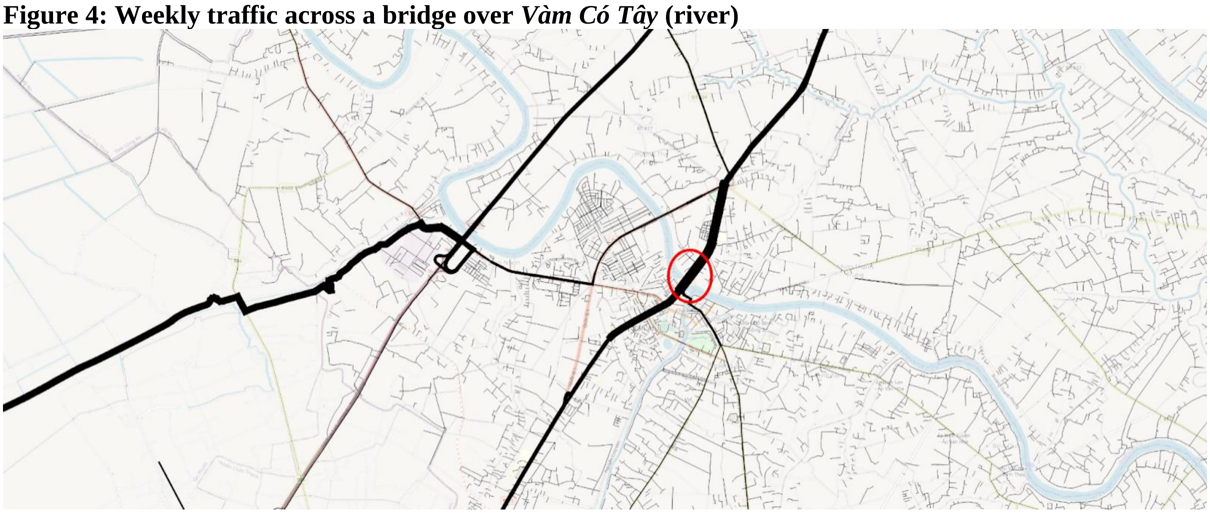
As a proof of concept, this study focused on performing a variety of analysis that could show the capabilities of this type of model, rather than explore the nuances of each analysis. Thus, details such as time profile and vehicle segmentation are not explored.

4.1 Disrupting a single network link

A bridge disruption most clearly illustrates the potential cost of a network disruption. For example, we could analyze what would have happened if one bridge over the *Vàm Có Tây* river (circled in red in the map on **Figure 4**) had been closed for a week during the lockdowns in the first half of 2020. For the sake of this example, we could get all the commercial vehicles operating in the general vicinity of the *Long An* province as tracked by the CVTS.

Looking into the base-case results, we verify that that there are 83,600 **commercial vehicle** trips used this *Vàm Có Tây* bridge per week.

Figure 4: Weekly traffic across a bridge over *Vàm Có Tây* (river)

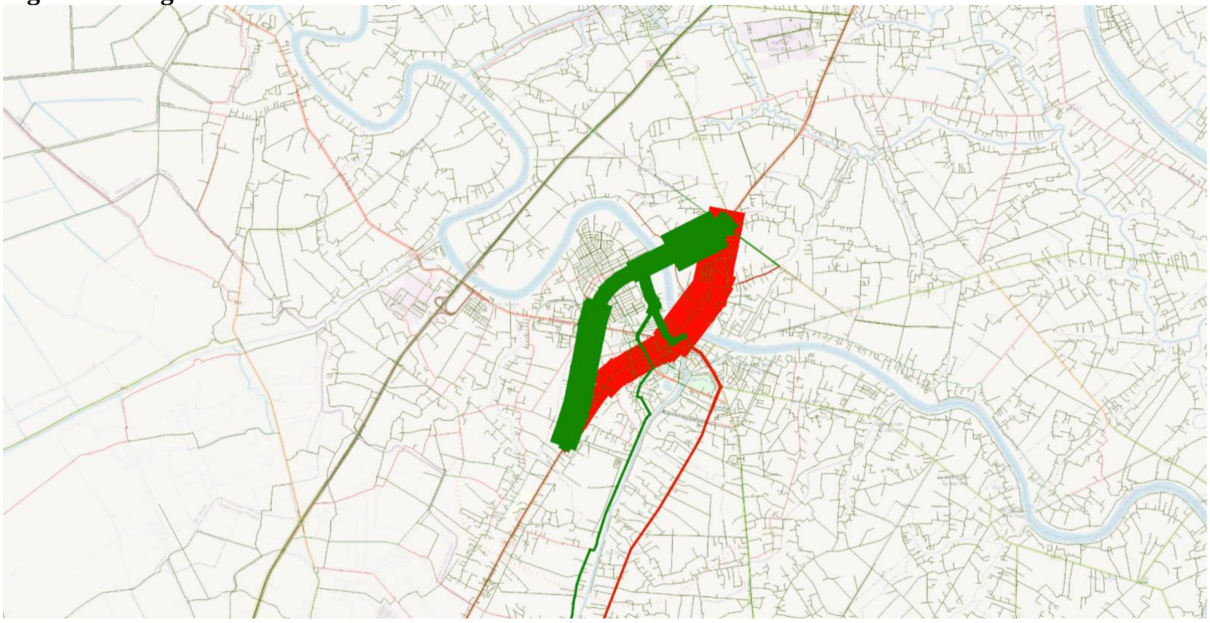


Source: Open-Street Maps, CVTS, Outer Loop Consulting

Source: Open Street Maps, CVTS

Re-routing these nearly 84,000 vehicles/week shows that the detouring is largely limited to the nearby bridge (**Figure 5**), but there are changes in virtually all the entire network, as trips that cross the bridge do indeed start and end throughout the entire *Long An* province and beyond. Quantitatively, re-routing these vehicles generates an additional 50,000 commercial.vehicle.km/week, and additional vehicle.hours, including added congestion to the rest of the network could as well be derived from this analysis.

Figure 5: Long An Commercial Traffic Detour Scenario



Source: Open-Street Maps, CVTS, Outer Loop Consulting

4.2 Comparative bridge disruption impact analysis

Competent asset management is one of the most efficient ways to maximize the value we can extract from our assets, and that is also true when our assets are the transportation network.

In that setting, effective asset management means to apply maintenance resources to the pieces of the network where there will be the most overall benefits, and to ensure that critical pieces of the network are well maintained and that disruptions in individual assets don't have catastrophic systemwide consequences.

To allocate resources, however, one needs to have an appropriate measure of the benefits and costs of a large number of resource allocation scenarios such that decision makers can be objective in such allocation.

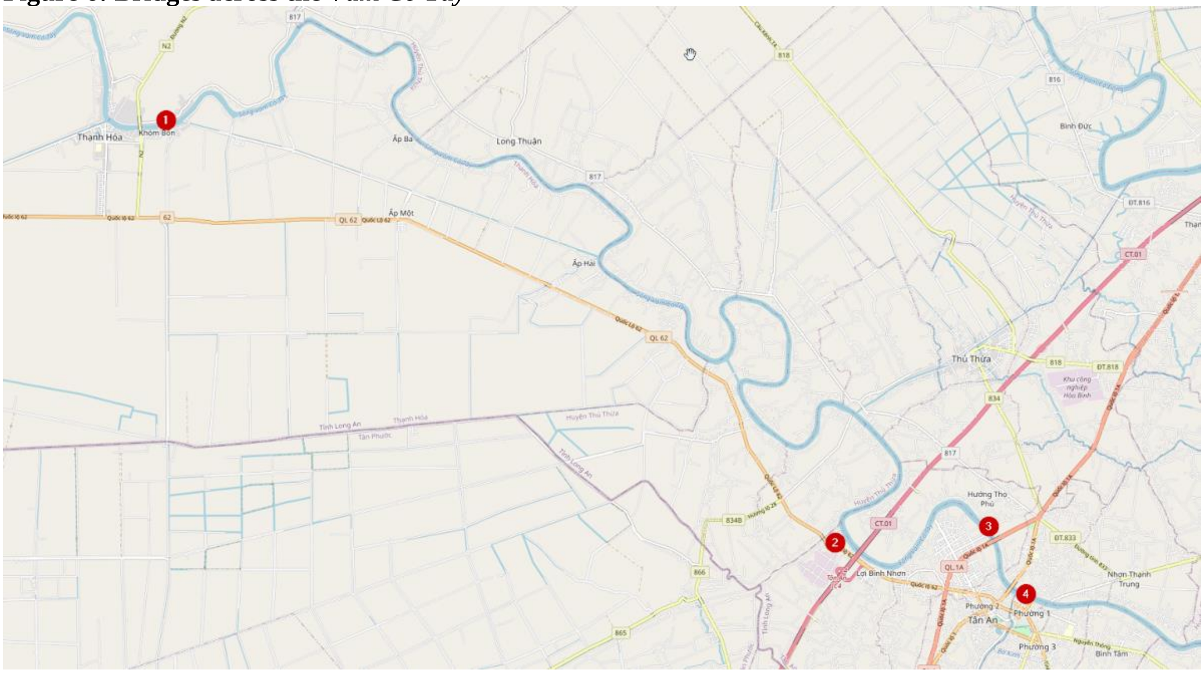
As a way to illustrate this process, we can once again look at the case of the crossings of the *Vàm Có Tay* river as a clear example of how difficult it would be to estimate the impacts of removing any one of the bridges over that river.

The difficulty in estimating the impact of disruptions for each one of these bridges comes from the fact that there are two very distinct sets of bridges in this case (**Figure 6**). The first set is the sole bridge along the *AL.N2* route, in the top-left corner of the map, where the second set is composed of bridges 2, 3 and 4, on the right side of the map, which are all part of the *CT01* and *QL1A* corridors.

By looking at the map it is possible to see that the second set of bridges is much more relevant for the North-South connectivity in the network as a whole and would be responsible for a much larger share of urban transport. However, all these bridges are reasonable alternatives to each other, while the bridge number 1, in the first set, has no alternative within 50 km.

Traffic flow numbers can be extracted with field counts or derived from the CVTS data, but understanding re-routing patterns for all vehicles should each one of these bridges were to be disrupted would provide much more solid evidence on potential impact.

Figure 6: Bridges across the *Vàm Có Tay*



Source: Open-Street Maps, Outer Loop Consulting

From the baseline assignment of the CVTS onto the network, we verify that bridge 1 is indeed substantially less used than the other bridges and is responsible for roughly 10% of the traffic across the river on an average day, as shown on **Table 1**.
However, when we look at the impact on transportation generation metrics (commercial veh.km) for scenarios where each bridge is rendered inactive, and the results on are overwhelming, priving bridge 1 substantially more critical than the remaining three bridges.

Table 1: Traffic flow and disruption estimates for bridges

Metric	Bridge 1	Bridge 2	Bridge 3	Bridge 4
Total Flow	2,850	8,980	3,090	11,900
Daily disruption impact (commercial veh.km)	32,761	6,846	2,433	7,136
Avg. detour per vehicle (km)	11.5	0.75	0.80	0.60

Source: Open-Street Maps, CVTS, Outer Loop Consulting

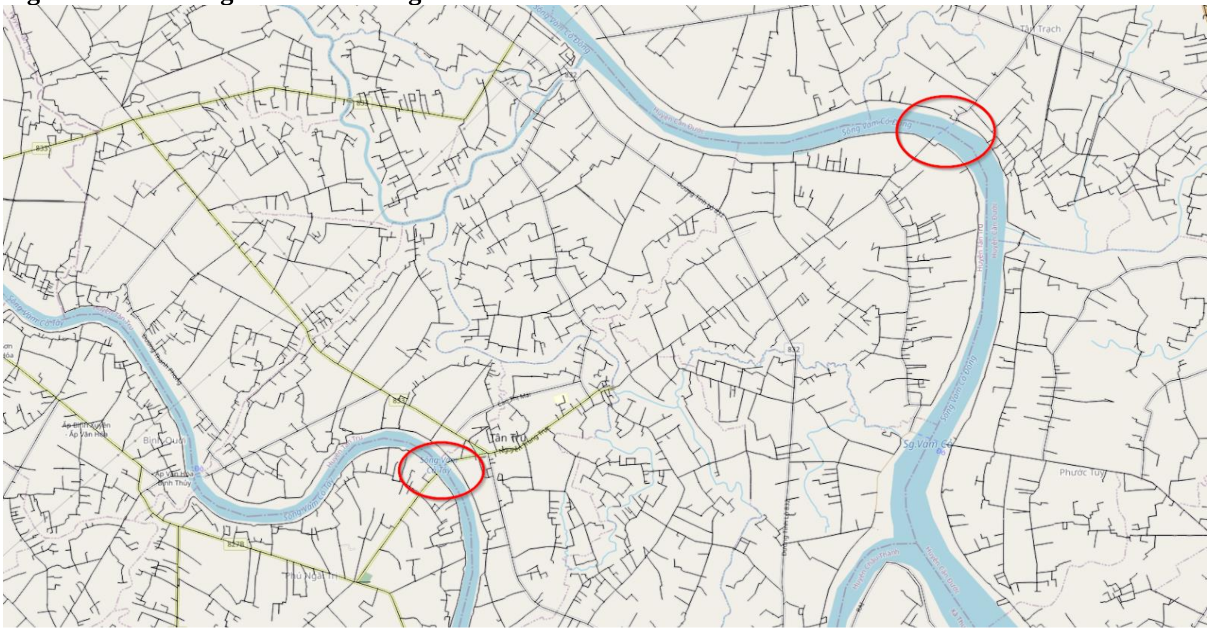
The average detour of each vehicle using bridge 1 is 14 to 19 times the average detours resulting from the disruption of any one of the other 3 bridges, bringing the total disruption to be at least 4.6 times the impact from the disruption of those other bridges.
In this case, the lack of local alternatives to bridge 1 has proven a decisive factor in establishing that bridge as the most critical piece of infrastructure among all 4 bridges analyzed, and the evidence that led to this conclusion can be a powerful piece in prioritizing maintenance resources throughout the region.

4.3 Marginal impact of infrastructure construction

It is also possible to estimate the impact of adding new facilities to the network, such as a new bridge[1]. As an example of this facility, we could look into adding two new bridges on the *DT833* across the *'Vàm Co Dong'* and *'Vàm Có Tay'* rivers, shown in the map on **Figure 7**.

[1] This type of analysis is better done in variable demand settings, but this fixed demand example provides a lower-bound estimate for the reduction in transportation production (veh.km) that this bridge construction would provide.

Figure 7: New bridge locations along the *DT833* route



Source: Open-Street Maps, Outer Loop Consulting

By adding the new bridges to the model network and using the exact same trip matrix used for the criticality analysis, it is quick to derive an estimate of approximately 3,180 vehicles per day crossing the busiest of bridges added (**Figure 8**).

It also becomes clear that such a connection would mostly benefit the vehicles traveling between *Tân Túc* (North) and *My Tho* (South), and that other Origins/destinations along the corridor connecting both cities would be responsible for much smaller trips using these new bridges, as shown on the map of (**Figure 8**).

Figure 8: New bridge locations along the *DT833* route

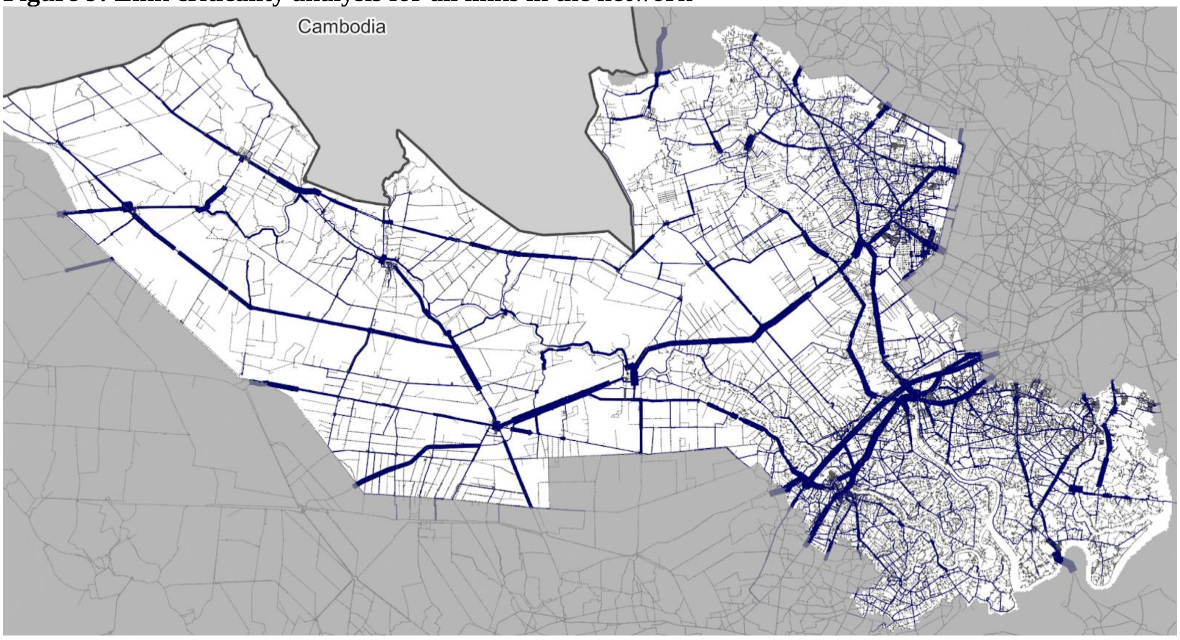
Source: Open-Street Maps, CVTS, Outer Loop Consulting

Finally, the daily reduction of vehicle kilometers in the network would be in excess of 13,670 commercial.veh.km, for an average of almost 4.3km for each vehicle using the new corridor.

4.4 A comprehensive disruption analysis view

One of the advantages of working with smaller models, such as one for a single province, is the ability of performing the disruption analysis for every link in the network within reasonable computational time to generate a comprehensive view of the road network with respect to their relevance in face of existing vehicle trips.

Figure 9: Link criticality analysis for all links in the network



Source: Open-Street Maps, CVTS, Outer Loop Consulting

The results of this analysis, presented on **Figure 9,** show that the most critical corridors are those that connect *Ho Chi Minh City*, Northeast to *Long An* to the southernmost portion of the country, which would be an expected result. However, a few isolated links throughout the network, particular in the more rural West portion of the province also stand out.

Average aggregate results per facility type are also possible, as we see that the average total impact of disrupting a bridge is over 6 times higher than the disruption for the average link in the network, as shown on **Table 2**.

Table 2: Average effects for link disruption

	Number of Sections	Length (km)	Average impact per link disrupted (total commercial veh.km)
Bridges	223	15.8	3,140
Total network	85,494	13,640	503

Source: Open-Street Maps, CVTS, Outer Loop Consulting

5 Conclusions

The results presented in this paper are in line with our expectations and seem to indicate the potential of this methodology. The challenge, however, is still in the amount of computational power, software development effort and specialized skill required for converting over 100 billion GPS records into useable information, as well as the research nature of the methodology implemented.

In that sense, this application shows *what* criticality analytics requires in terms of processing steps and data and illustrates *how* it can be accomplished.
The key result from this work is certainly the comparison between the two sets of bridges, as it highlights how looking at total traffic flow can be misleading when trying to estimate the impact of disruption.

5.1 Future steps

As this research has shown the feasibility and value in developing a network resilience model based on the CVTS data and implementing a system to continuously update the results based on incoming data feeds may be a reasonable direction for decision makers. In that case, web-based dashboards may be required to support local technical staff.
This type of information can then be used to (i) prioritize road and bridges infrastructure assets for pre-emptive investment and maintenance for resilience reviews, (ii) generate cost-benefit analysis metrics for investing in individual links.
Incorporation of personal travel demand into the model though demand modelling or LBS data feeds may also be an important research direction, as it allows the incorporation of congestion into the analysis, making it substantially more accurate.

5.2 Acknowledgement

The work that gave basis to this paper was developed with funding from the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.
The author would also like to thank Kai Kaiser at the World Bank, Jamie Cook from Outer Loop Consulting and Adam Mckeown and Simon Knapp from CSIRO for the collaboration in the larger Vietnam project.

6 References

AequilibraE Developers, 2022, ‘AequilibraE documentation’, viewed 18 July 2022, http://aequilibrae.com.

Bennion, MW & O’Neill, WA, 1994, ‘Building transportation analysis zones using geographic information systems’, *Transportation Research Record*, no. 1429.

Camargo, P, 2021, ‘A Model from nothing. How far can get (and what can you do), with Open-Source and Open-Data alone’, Presented at the 18th TRB Conference on Transportation Planning Applications, USA, Transportation Research Board, Portland, Oregon.

Camargo, P, Hong, S, & Livshits, V, 2017, ‘Expanding the Uses of Truck GPS Data in Freight Modeling and Planning Activities’, *Transportation Research Record*, vol. 2646, no. 1, pp. 68–76, doi: 10.3141/2646-08.

Colon, C, Hallegatte, S, & Rozenberg, J, 2019, *Transportation and Supply Chain Resilience in the United Republic of Tanzania: Assessing the Supply-Chain Impacts of Disaster-Induced Transportation Disruptions*, World Bank, doi:/10.1596/31909.

Das, R, 2020, ‘Approach for measuring transportation network resiliency: A case study on Dhaka, Bangladesh’, *Case Studies on Transport Policy*, vol. 8, no. 2, pp. 586–592, doi: 10.1016/j.cstp.2020.04.001.

Espinet, X, Rozenberg, J, Rao, KS, & Ogita, S, 2018, *Piloting the Use of Network Analysis and Decision-Making under Uncertainty in Transport Operations: Preparation and Appraisal of a Rural Roads Project in Mozambique under Changing Flood Risk and Other Deep Uncertainties*, World Bank, doi:/10.1596/1813-9450-8490.

Ghadiri, M, Rassafi, AA, & Mirbaha, B, 2019, ‘The effects of traffic zoning with regular geometric shapes on the precision of trip production models’, *Journal of Transport Geography*, vol. 78, pp. 150–159, doi: 10.1016/J.JTRANGEO.2019.05.018.

Martínez, LM, Viegas, JM, & Silva, EA, 2009, ‘A traffic analysis zone definition: a new methodology and algorithm’, *Transportation*, vol. 36, no. 5, pp. 581–599, doi: 10.1007/s11116-009-9214-z.

McNally, MG, 2008, ‘The Four Step Model’, Retrieved from https://escholarship.org/uc/item/0r75311t.

‘Open Spatial Demographic Data and Research - WorldPop’, n.d., viewed 17 July 2022, https://www.worldpop.org/.

Openshaw, S, 1977, ‘A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling’, *Transactions of the Institute of British Geographers*, vol. 2, no. 4, pp. 459–472, doi: 10.2307/622300.

Sahu, PK, Chandra, A, Pani, A, & Majumdar, BB, 2020, ‘Designing freight traffic analysis zones for metropolitan areas: identification of optimal scale for macro-level freight travel analysis’, *Transportation Planning and Technology*, vol. 43, no. 6, pp. 620–637, doi: 10.1080/03081060.2020.1780711.

Taghavi, M, Irannezhad, E, & Prato, CG, 2019, ‘Identifying Truck Stops from a Large Stream of GPS Data via a Hidden Markov Chain Model’, in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2265–2271, doi: 10.1109/ITSC.2019.8917156.

Thierry, B, Chaix, B, & Kestens, Y, 2013, ‘Detecting activity locations from raw GPS data: a novel kernel-based algorithm’, *International Journal of Health Geographics*, vol. 12, no. 1, p. 14, doi: 10.1186/1476-072X-12-14.

Yang, Y, Jia, B, Yan, X-Y, Li, J, Yang, Z, & Gao, Z, 2022, ‘Identifying intercity freight trip ends of heavy trucks from GPS data’, *Transportation Research Part E: Logistics and Transportation Review*, vol. 157, p. 102590, doi: 10.1016/j.tre.2021.102590.

You, J, Nedović-Budić, Z, & Kim, TJ, 1998, ‘A GIS-based traffic analysis zone design: technique’, *Transportation Planning and Technology*, vol. 21, no. 1–2, pp. 45–68, doi: 10.1080/03081069708717601.