The role of mobility infrastructure at risk for carbon efficiency

Mauricio Rada-Orellana[1] Markus Schläpfer[2]

[1]Department of Urban Planning
Columbia University, New York, USA

[2]Department of Civil Engineering and Engineering Mechanics
Columbia University, New York, USA

Abstract

This paper explores the impact of vulnerable mobility infrastructure on carbon emissions efficiency in four countries in the Global South: Colombia, Mexico, Indonesia, and India. The study assesses how inadequacies in transportation networks—insufficient infrastructure, unreliable connectivity, and risk vulnerability—affect the carbon efficiency of mobility systems. Using a combination of empirical data analysis and comparative assessments, the study identifies key characteristics of infrastructural vulnerability and their relation to both carbon emissions and changes in the spatial interactions of people. Our results suggest that accessibility is a key driver in public transport, but high fare costs could reduce travel and, therefore, people's spatial interactions. By expanding the capacity of reliable transit systems, both spatial interactions and carbon efficiency can be enhanced.

USE-Lab Working Paper #2025–12–003
https://urbansystems.civil.columbia.edu

The Role of Mobility Infrastructure at Risk for Carbon Efficiency

Mauricio Rada-Orellana[1] and Markus Schläpfer[2]
[1]*Department of Urban Planning*
[2]*Department of Civil Engineering and Engineering Mechanics*
Columbia University, New York, USA

This paper explores the impact of vulnerable mobility infrastructure on carbon emissions efficiency in four countries in the Global South: Colombia, Mexico, Indonesia, and India. The study assesses how inadequacies in transportation networks—insufficient infrastructure, unreliable connectivity, and risk vulnerability—affect the carbon efficiency of mobility systems. Using a combination of empirical data analysis and comparative assessments, the study identifies key characteristics of infrastructural vulnerability and their relation to both carbon emissions and changes in the spatial interactions of people. Our results suggest that accessibility is a key driver in public transport, but high fare costs could reduce travel and, therefore, people's spatial interactions. By expanding the capacity of reliable transit systems, both spatial interactions and carbon efficiency can be enhanced.

I. Introduction

While carbon-efficient transportation projects are ubiquitous in developed economies, cities in the Global South mainly rely on informal, vulnerable mobility infrastructure. Therefore, it is important to understand how these systems modulate the citizens' travel patterns, which can lead to higher carbon emissions.

Transportation-related emissions account for $\approx 21\%$ of the total contribution of carbon dioxide, making it the second most significant contributor worldwide [1]. Focusing on the ground transportation of people within cities, the contribution of CO_2 largely depends on the travel choices citizens make [2, 3]. For example, a trip in which the main mode was a (combustion engine) car would generate more CO_2 compared to a trip based on active transportation such as biking or walking. Compared to car-based trips, a public transit-based trip, using a bus or the subway, would have a higher carbon efficiency since the total carbon emitted corresponds to a larger number of passengers. Furthermore, a higher number of cars can lead to a higher trip duration due to congestion, further reducing the carbon efficiency.

These commute choices are shaped by the configuration of the multimodal mobility system, which consists of multiple factors. An example is the degree of vulnerability, since citizens can be more incentivized to choose taxis in a less reliable public transportation system. Another factor is accessibility: if the mass transit system does not reach the peripheries of the cities, citizens can be more inclined to own a car. Although climate resilience is a widely explored subject in academic discourse, significant gaps remain in its study on transportation networks [4, 5].

At the same time, one of the main drivers for cities to exist is to enable spatial connectivity and the resulting interactions of people, accelerating knowledge diffusion, agglomeration economies, and social integration [6, 7]. As a result, policy objectives should not aim to reduce the number of trips per se, but rather to find ways in which carbon emissions are lowered while fostering those spatial interactions. To that end, this study seeks to evaluate the characteristics of the urban transportation infrastructure across cities in the Global South on a daily basis and examine their influence on carbon emissions and spatial interactions by analyzing the intraurban human mobility patterns. By doing so, we address a research gap regarding traveler's behavior and vulnerability, which has not been sufficiently explored in the existing literature [8]. The focus on mobility, carbon, and interaction provides a nuanced understanding of how mobility infrastructure impacts environmental and social dynamics, contributing to the broader discourse on sustainable urban development.

The objectives of this study are:

1. Providing a conceptual and quantitative characterization of vulnerability-related features of the mobility infrastructure focused on cities in the Global South.
2. Evaluating the influence of the mobility infrastructure characteristics on ground-transportation carbon emissions and spatial interaction, independently and as a composite ratio.
3. Building a data-driven approach that integrates infrastructure risk and vulnerability to carbon emissions, aiming to complement the more explored, inverse relationship.

To accomplish the objectives, the research framework combines spatial calculations, network analysis, and regression techniques. First, we identify nine features that through a process of data collection, spatial processing, and network simulation are numerically defined. Second, using mobile data location and carbon estimations, we quantify a composite ratio that measures the performance of carbon emissions and spatial interaction. This ratio is then studied with the risk characteristics through statistical analysis.

This report is divided into five sections. The following describes the various datasets that were used to construct the study variables. Section 3 focuses on the methodological framework to build each of the explanatory and dependent variables, as well as the process of describing the relationship between them. Section 4 presents the results of the regression techniques and comparative patterns between the urban agglomerations. Finally, section 5 presents an overview of the results in light of other current studies and avenues for further research.

II. Datasets

The datasets used include aggregated origin-destinatio data for Colombia, Mexico, Indonesia, and India [11], flood risk satellite imagery [12], urban center boundaries [13], daily gridded ground-transportation CO_2 emissions for November and December 2019 [14], WorldPop population counts [15], and OpenStreetMaps [16]. For this study, we focus on urban agglomerations greater than one million citizens ($n = 111$) on a daily measurement between November 1st 2019 and December 31, 2019.

A. Transportation Network Data

To estimate the features that derive from the public transit network, we used OpenStreetMap [16] as the data source and processed with OSMnx [17] to retrieve a collection of edges and nodes for each city. OpenStreetMap is an open-source collective dataset that is continuously updated, as a result, the extracted instance depends on the date and time when the data collection occurs. For each urban agglomeration, we obtain a single connected undirected graph representing an aggregation of public transit routes. Out of the 111 urban agglomerations we are studying, only 2 have an empty graph due to the absence of public transportation data.

B. Carbon Emissions

For this study, we focus on transportation-driven carbon emissions, specifically the volume generated from ground transit, including private and public vehicles. We used the GRACED dataset [14] which provides high-resolution (0.1 degrees of spatial resolution), global daily CO_2 emissions data which relies on the Carbon Monitor, EDGAR, and TROPOMI NO_2 satellite measurements. GRACED datasets have independent subtypes focused on different sectors like aviation, residential, energy, or ground transportation. It is important to mention that we focused only on ground transportation for the months of November and December of 2019, which allows us to reduce potential biases that could come from non-transportation sectors. The original rasters were combined for each agglomeration boundary and aggregated into a single daily CO_2 emission volume.

C. Flood Hazard Raster

As a measure of the level of hazardousness of public transit in these four countries, we selected the flood hazard dataset provided by the European Joint Research Centre [12]. The raster, of a spatial resolution of 30 arc-seconds, indicates the water depth in a 100-year return period flood. For this study, we selected a flood exposure of 1 meter as a minimum.

D. Mobility Data Patterns

The main dataset for this study is the mobility data patterns estimated from GPS tracking. This dataset is obtained from Spectus [11] and provides the origin-destination statistics across four countries in a gridded geographical boundary. The dataset is spatially divided into geohashed cells of level 5 resolution with a daily temporal dimension. It describes the descriptive statistics of the patterns which include the average distance, average time, standard deviation, median, and counts, which are based on mobile phone data tracing for a day of duration between November 1st 2019 and December 31, 2019.

III. Methodology

The study aims to explain the relation of the mobility infrastructure with carbon emission and spatial interaction within the cities. The analysis is conducted at the city level in its geographical dimension which is composed of: (1) Mobility infrastructure characterization, whose quantitative dimensions are depicted in Table I, (2) Calculation of the human interaction - carbon emission efficiency (HI-CE efficiency), and (3) Estimation of the infrastructure effects in (1) on the spatial interaction, HI-CE efficiency, using statistical analysis and machine learning techniques. For this study, we focus on urban agglomerations above one million citizens ($n = 111$) on a daily measurement between November, 1st 2019 to December, 31st 2019.

The first stage involved preparing the datasets to ensure spatial and temporal consistency across the urban agglomerations. For this purpose, we defined the urban agglomerations based on the catchment regions of continuous urbanized areas developed by Cattaneo et al. (2024) [18]. To provide a sufficient breadth and sample size, we selected Tier 1 urban centers which includes 1430 city-regions worldwide. These are then spatially merged with the GADM datasets [19] to filter by country and population and ensure consistency with the administrative delimitations per country. The outcome is shown in Figure 1.

With the generated boundaries, we associated the origin-destination daily data at the geohash level 5 to the urban agglomerations to calculate the static characteristics in Table I. These are service quality, variability, and speed efficiency. Capacity is calculated with the

Characteristic	Measurement	Dataset	Temporality
Service Quality	Average travel time to destination	NetMob OD	Daily
Capacity	Peak trips count (all modes) during typical travel days	NetMob OD	Static
Fare Burden	Ratio of monthly transit fare to national minimum income	Various	Static
Hazardousness	Percentage of public transit network length in flood-prone area	JRC Flood Hazard	Static
Accessibility	Population within access to public transit services	OSM, WorldPop	Static
Variability	Aggregated standard deviation of travel time	NetMob OD	Daily
Speed Efficiency	Ratio of trip distance and travel time	NetMob OD	Daily
Robustness	Ratio of removed critical edges before public transit network failure[a]	OSM	Static
Connectivity	Network global efficiency of public transit network [9]	OSM	Static

[a]Edge order calculated using betweenness centrality until network shrinks to 50% of the total length [10]

TABLE I: Characterization of mobility infrastructure.

maximum number of trips that occur in the city when the average travel time is within a ± 5% of its median. Hazardousness is calculated by estimating the proportion of the length of the public transit system that is located within flood-prone areas based on the European Joint Research Center Flood Hazard dataset [12]. Fare burden is expressed as a ratio between monthly public transit fare, assuming a daily round-trip travel during a 30-day period, and minimum country income. Transit fare information was obtained using ChatGPT [20], for preliminary collection, and manually cross-referenced with the cost of living by Numbeo [21]. Accessibility represents the total population that lives near the transit system using the WorldPop grid cells of 3 arc seconds (≈100 m).

For each agglomeration, transportation networks were extracted and processed using OSMnx [17], which provided graph representations of public transit routes. The graph allowed us to calculate Robustness and Connectivity. Robustness represents how many network links would have to fail to cause a substantial failure of the entire system. It is calculated by sorting the links by their importance in the system, using the betweenness centrality, and removing continuously until the length of the largest connected component drops to 50% of the complete graph. Connectivity was obtained through OSMnx [17] that represents the average efficiency of the graph defined by Latora and Marchiori (2001) [9].

For the second stage, we clipped emissions and mobility data using urban boundary files, allowing for the analysis of spatial relationships and infrastructure features. The Human Interaction - Carbon Emission (HI-CE) efficiency serves as a key indicator of how efficiently human activities (visits) are managed in relation to the carbon emissions generated by ground transportation. The daily visit count variable captures the number of trips that ended within each geohash cell during a 24-hour period. We calculate the HI-CE efficiency for each city by relating the aggregated trip count across all cells to the total daily CO_2 emissions. The total daily CO_2 emissions are the aggregation of all the cells within the urban boundary. Equation 1 reflects the operationalization of this variable.

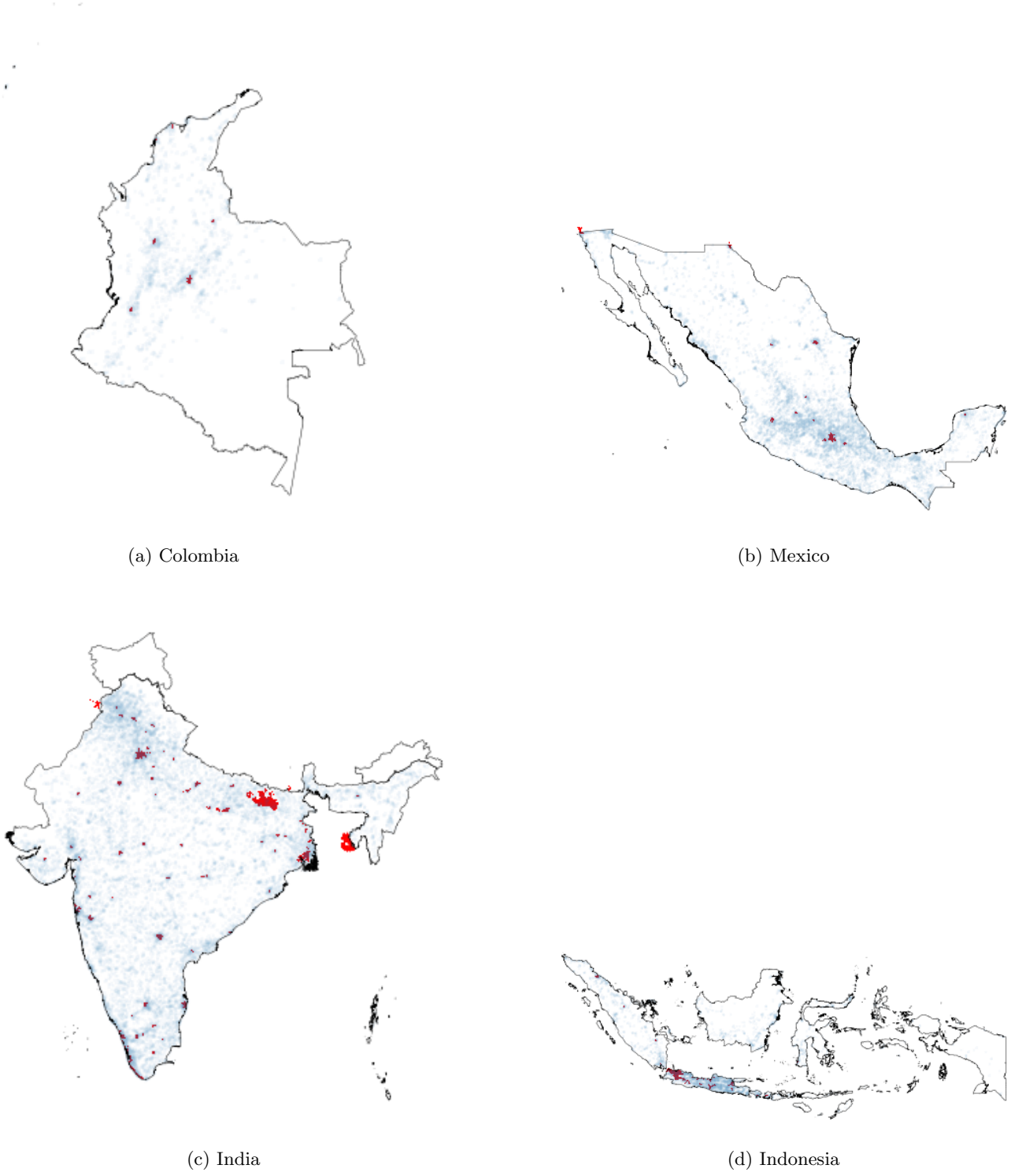
$$\text{HI-CE}_{Uij} = \frac{\sum_k^{U_i} \text{Visits}_{kj}}{\sum_l^{U_i} \text{Daily CO}_2 \text{ Emissions}_{lj}} \tag{1}$$

Where Visits_{kj} is the end visits to the cell k during the day j and Daily CO_2 Emissions_{lj} is the aggregation of the volume of CO_2 ground-transportation emissions in the cell l during the day j. U_i represents the urban agglomeration i.

The third stage consists in ordinary least squares models to quantify the relationship between the mobility variables (such as service quality, robustness, efficiency, among others) and two key outcomes: carbon emissions and spatial interaction efficiency or HI-CE. Independent variables include infrastructure-related metrics, while dependent variables are the carbon emissions (Model 1) and spatial efficiency (Model 2). Since some variables have large scales, the variables are standardized to make the comparison between regression coefficients more interpretable. The linear regressions is supplemented with a machine learning technique (random forest) to capture potential non-linear relationships. The residuals histograms for both models is shown in Figure 2.

IV. Results and Discussion

The relations between transportation infrastructure, ground transport carbon emissions (kg C/h), and interaction efficiency show a consistent trend among ordinary least squares (OLS) as described in Table II. The random forest regressions show a similar trend with R^2 scores of 0.99 for the test sets for both models. For Model 1, it is worth mentioning that a positive value represents a positive association between the infrastructure variable and higher emissions of CO_2. As mentioned in the datasets section, the CO_2 information only includes ground-transportation emissions in its inventory. In the case of Model 2, a positive coefficient value suggests that the variable is positively correlated with HI-CE or increased spatial interactions per kg of carbon emissions. Considering the urban objectives described in the introduction, cities should strive for more spatial interactions



(a) Colombia

(b) Mexico

(c) India

(d) Indonesia

FIG. 1: Urban agglomerations (in red) for each country obtained from the contiguous catchment regions, the GADM administrative boundaries and the mobile phone data (MPD) as described in the first stage of the methodology. The blue areas represent the MPD at the geohash level 5.

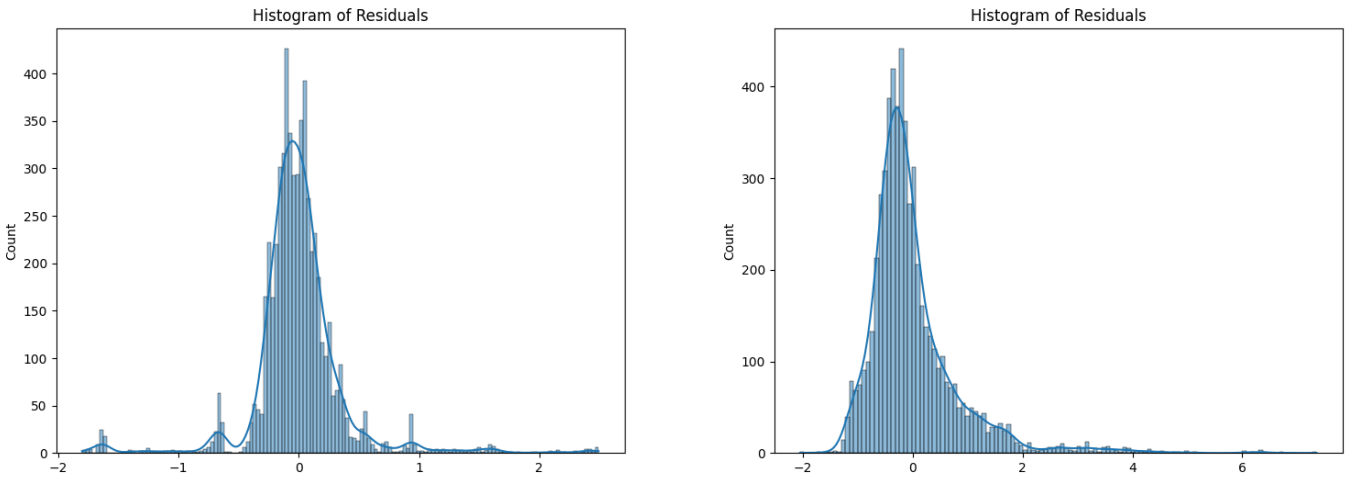


FIG. 2: Histogram of residuals for both models tested in the study. The left plot shows the distribution for the model 1 of CO_2 emissions. The right plot shows the distribution for the model 2 which shows the HI-CE metric. Both show bell-shaped distributions to hold the normality assumption of residuals for the OLS regression.

keeping lower CO_2 emissions, or equivalently, a higher HI-CE metric.

The most important factor in the public transport variables is accessibility, but managing fare affordability is crucial in developing economies. Considering the capital costs of building infrastructure, increased accessibility to the mobility system could be reflected in higher fares. While they may reduce emissions by suppressing travel demand, they negatively impact the spatial interaction metric (coefficient of -0.1204), as we found in both Model 1 and 2 of this study. In consequence, the positive effects of increased transit accessibility on interaction efficiency can be diminished when the cost burden is too high. This result supports that environmental sustainability depends both on good economics and social equality for a policy to produce satisfactory results proposed by Cervero [22].

When considering all the mobility characteristics, not only those related to public transit, the total capacity of the infrastructure, whether roads, parking, or public transport, is of strong statistical significance for both the carbon and HI-CE models. Interestingly, higher capacity is associated with increased carbon emissions (coefficient of 0.1765), implying that without a broader strategy, a capacity expansion policy induces more travel demand, including CO_2-intensive informal transport [23]. This relationship, coupled with the influence of robustness of the public system on both target metrics, suggests that boosting reliable transit capacity can improve overall spatial interaction-carbon efficiency. Our models suggest that increased robustness leads to lower CO_2 and higher interaction efficiency.

One possible explanation is that in the event of a disruption, such as floods or maintenance closures, residents in cities with less resilient public transit systems are more likely to resort to private transportation, which results in higher carbon emissions. Alternatively, some residents may forego travel altogether, leading to reduced levels of interaction and a decline in the Human Interaction-Carbon Emission Efficiency (HI-CE) metric. This illustrates how the public transport robustness can influence both environmental and social efficiency.

Figure 3 shows a comparison between two Colombian cities prone to flood hazards with different robustness and HI-CE efficiencies. Cali shows a higher interaction efficiency than Barranquilla. This indicates that despite the flood vulnerability, Cali's transit infrastructure facilitates more spatial interactions per carbon emissions. In the robustness simulation shown in Figure 3, the black line indicates a baseline in which the effect of edge removal is constant linear relationship with the largest connected component. Urban agglomerations above the line, like Cali, have a more robust public system than those under the baseline.

Together, our work explores an important, yet less studied trade-off in the growth of transit networks and the resulting CO_2 emissions: building more expensive sparse networks that cover larger urban areas but are less resilient, versus denser, affordable networks that (initially) cover smaller areas but are more resilient. As such, we expect that our results will have direct implications for climate-sensitive and inclusive transit planning in the Global South.

	OLS Model 1 - CO_2			OLS Model 2 - Efficiency		
Variable	Coefficient	St Error	t-stat	Coefficient	St Error	t-stat
Constant	-2.776e-17	0.005	-5.51e-15	2.45e-16	0.011	2.25e-14
Service Quality	-0.0527	0.010	-5.418*	-0.0267	0.021	-1.267
Capacity	0.1765	0.009	20.150*	0.4126	0.019	21.758*
Accessibility	-0.6546	0.010	-63.920*	0.1329	0.022	5.992*
Variability	0.0154	0.009	1.626	0.0295	0.021	1.435
Speed Efficiency	-0.0048	0.005	-0.940	-0.0328	0.011	-2.990*
Robustness	-0.0278	0.006	-4.857*	0.0657	0.012	5.313*
Connectivity	-0.0255	0.006	-4.456*	0.0035	0.012	0.286
Fare Burden	-0.0772	0.006	-12.826*	-0.1204	0.013	-9.234*
Hazardousness	-0.0395	0.006	-6.491*	-0.0470	0.013	-3.567*
Population	1.6216	0.026	62.996*	-0.6036	0.056	-10.829*
Area	-0.4904	0.018	-27.042*	0.1876	0.039	4.776*
Density	-0.1770	0.008	-23.454*	0.5223	0.016	31.955*

TABLE II: Model 1: Regression for Carbon Emission (kgC/h). $R^2 = 0.853$. Model 2: Regression for HI - CE (Interactions per kg C/h). $R^2 = 0.309$. * represents a p-value < 0.05.

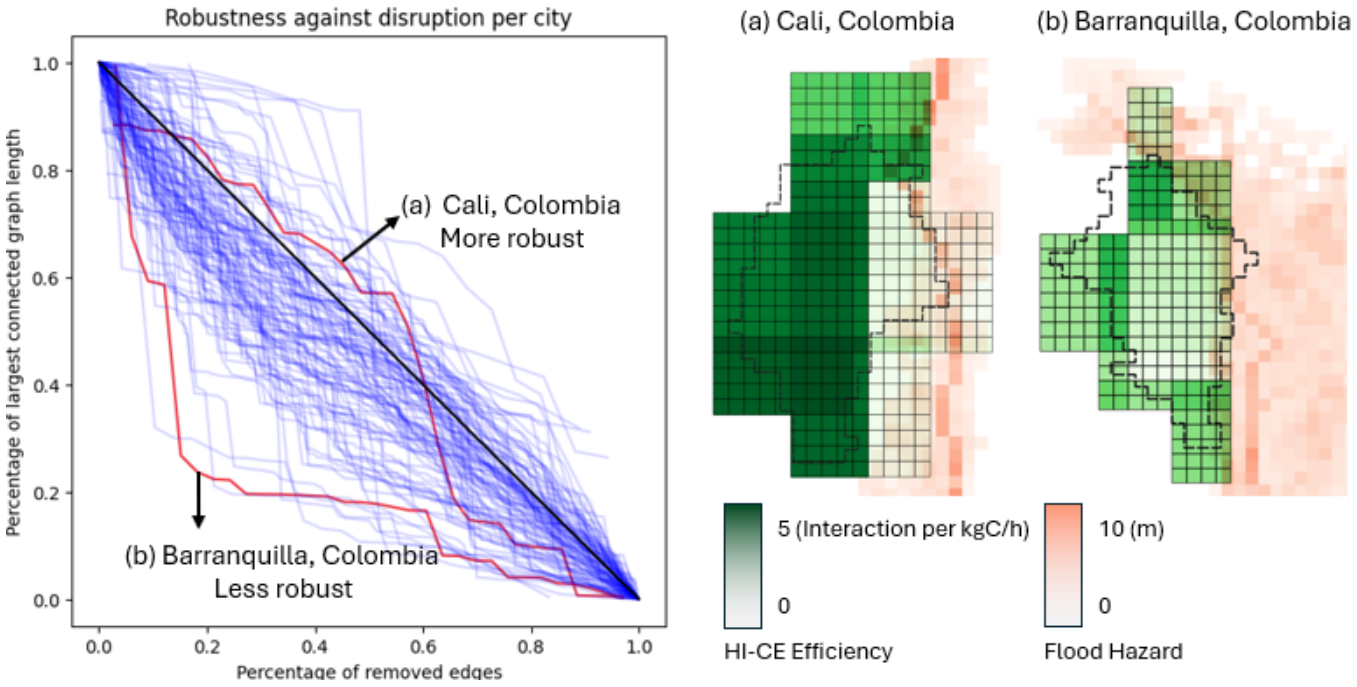


FIG. 3: Comparison of robustness simulation across cities. The left plot shows the relationship between the percentage of removed edges in the public transit system and the length of the largest connected component, obtained from the robustness simulation. Each line plot represents an urban agglomeration. The right plot shows an example of the spatial distribution of HI-CE and flood hazard for Cali and Barranquilla.

[1] IEA, *Energy Statistics Data Browser* (IEA, Paris, 2023) https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser.

[2] A. Javaid, F. Creutzig, and S. Bamberg, Determinants of low-carbon transport mode adoption: systematic review of reviews, Environmental Research Letters **15**, 103002 (2020).

[3] D. Liu, H. Du, F. Southworth, and S. Ma, The influence of social-psychological factors on the intention to choose low-carbon travel modes in Tianjin, China, Transportation Research Part A: Policy and Practice **105**, 42 (2017).

[4] S. Hayes, C. Desha, M. Burke, M. Gibbs, and M. Chester, Leveraging socio-ecological resilience theory to build climate resilience in transport infrastructure, Transport Reviews **39**, 677 (2019).

[5] C. Wan, Z. Yang, D. Zhang, X. Yan, and S. Fan, Re-

silience in transportation systems: a systematic review and future directions, Transport reviews **38**, 479 (2018).
[6] J. Jacobs, *The Death and Life of Great American Cities* (Vintage Books, 1961).
[7] A. Bertaud, Order without design: How markets shape cities, Town and Regional Planning **79**, 2 (2021).
[8] S. Pan, H. Yan, J. He, and Z. He, Vulnerability and resilience of transportation systems: A recent literature review, Physica A: Statistical Mechanics and its Applications **581**, 126235 (2021).
[9] V. Latora and M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. **87**, 198701 (2001).
[10] R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, Nature **406**, 378 (2000).
[11] S. Milusheva, Z. W., M. Nunez del Prado Cortez, and V. Gauthier, The NetMob24 dataset: A high resolution dataset of density and mobility for Colombia, India, Indonesia and Mexico (2024).
[12] F. Dottori, L. Alfieri, P. Salamon, A. Bianchi, L. Feyen, and F. Hirpa, *Flood hazard map of the World - 100-year return period* (European Commission, Joint Research Centre (JRC), 2016) [Dataset].
[13] A. Cattaneo, S. Girgin, R. A. de By, T. McMenomy, A. Nelson, and S. Vaz, Worldwide delineation of multi-tier city-regions, 10.5281/zenodo.11187634 (2024).
[14] X. Dou, J. Hong, P. Ciais, F. Chevallier, F. Yan, Y. Yu, Y. Hu, D. Huo, Y. Sun, Y. Wang, *et al.*, Near-real-time global gridded daily CO2 emissions 2021, Scientific data - Nature **10**, 69 (2023).
[15] A. J. Tatem, Worldpop, open data for spatial demography, Scientific data - Nature **4**, 1 (2017).
[16] OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org (2017), `https://www.openstreetmap.org`.
[17] G. Boeing, Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, Computers, environment and urban systems **65**, 126 (2017).
[18] A. Cattaneo, S. Girgin, R. de By, T. McMenomy, A. Nelson, and S. Vaz, Worldwide delineation of multi-tier city–regions, Nature Cities **1**, 469 (2024).
[19] R. Hijmans, N. Garcia, and J. Wieczorek, GADM: database of global administrative areas, version 4.1, GADM Maps and Data (2022).
[20] OpenAI, ChatGPT (GPT-4o version) [Large language model] (2024), accessed via https://chatgpt.com.
[21] Numbeo, Cost of living 2024 (2024), available online at: https://www.numbeo.com/common/ (accessed September 15, 2024).
[22] R. Cervero, *Transport infrastructure and the environment: Sustainable mobility and urbanism* (IURD, Institute of Urban and Regional Development, University of California, 2013).
[23] L. Al Otary, M. Abou-Zeid, and I. Kaysi, Modeling car ownership and use in a developing country context with informal public transportation, Transportation , 1 (2022).