

# INAUGURAL – DISSERTATION

zur

Erlangung der Doktorwürde

der

Gesamtfakultät für Mathematik,

Ingenieur- und Naturwissenschaften

der

Ruprecht-Karls-Universität

Heidelberg

Vorgelegt von  
M.Sc. Steffen Knoblauch

Tag der mündlichen Prüfung: 18.12.2024



---

# Harnessing Geospatial Big Data to Guide Local Interventions for *Aedes*-borne Arboviral Infections

---

Supervisors:

Prof. Dr. Alexander ZIPF

Prof. Dr. Joacim ROCKLÖV



# Acknowledgments

Redacted for online publication



## Abstract

*Aedes*-borne arboviral infections, such as dengue, are a major public health threat, with 400 million infections and 40,000 deaths annually. The WHO projects that by 2080, over 60% of the global population will reside in areas at risk from *Ae. aegypti*. This risk is driven by the impacts of climate change and accelerating urbanization, the latter increasing the availability of artificial containers that *Ae. aegypti* prefers for breeding. Effective control requires a thorough understanding of *Ae. aegypti*'s distribution and interactions with humans. Pathogen transmission is complicated by the mosquito's daytime biting behavior and high spatial variability in abundance, which is particularly pronounced in heterogeneous urban settings due to the limited flight range of *Aedes* mosquitoes. Sample-based entomological surveillance methods, such as ovitraps, often considered for guiding interventions, cannot capture these complex dynamics. This work aims to enhance surveillance by integrating geospatial big data to refine disease control.

This thesis developed novel methods for (i) generating environmental suitability indicators for *Ae. aegypti*, to interpolate ovitrap data across heterogeneous urban landscapes considering *Aedes* flight range, and (ii) modeling daytime *Aedes*-human interactions to incorporate biting risk into guidance for local interventions. To enable practical application, all methods were designed to be scalable to the municipal level, as demonstrated in the Rio de Janeiro case study (1,221 km<sup>2</sup>), representing the primary intervention level in many countries, including Brazil.

The results indicate that (i) *Aedes* breeding container density, estimated at scale using satellite and street view imagery, can be a significant indicator ( $p \leq 0.05$ ) for modeling *Ae. aegypti* egg and larval counts, as monitored with 2,700 ovitraps. The significance varies by container type, data source, and the modeled *Aedes* habitat size around ovitrap locations. (ii) When combined with additional suitability indicators derived from geospatial big data, such as water accumulation, urban morphology, and urban climate, these indicators explained up to 75% of ovitrap count variation in this case study. (iii) Modeling daytime *Aedes*-human interactions with mobile phone data improved dengue predictions by 14%, contributing to the explanation of 77% in spatial deviance of 8-year occurrence (2015-2022). Collectively, these findings provide improved guidance for interventions, informed by (a) continuous environmental suitability maps for *Ae. aegypti* and (b) key *Aedes*-human interaction hotspots. Major challenges limiting accuracy remain due to heterogeneous data availability and coarse resolution.

The developed concept can potentially reduce public health costs in two ways: (i) by providing more precise intervention guidance and (ii) by minimizing the number of ovitraps needed, through interpolation of entomological measurements and optimal surveillance positioning. These insights are particularly important as long as vector control measures, such as *Wolbachia* and insecticide spraying, (i) remain among the most effective interventions against *Aedes*-borne arboviral infections due to the challenges in vaccine development and equitable healthcare access, and (ii) continue to be costly or face environmental regulations. Although tailored to the municipality of Rio de Janeiro, these methods may be applicable to other *Aedes*-endemic areas with similar ecological contexts, pending further research for broader application and integration into prescriptive analytics, facilitating the transition from predictive insights to more actionable strategies.



## Zusammenfassung

Arbovirus-Infektionen, die durch *Aedes*-Mücken übertragen werden, wie beispielsweise Dengue, stellen eine erhebliche Bedrohung für die öffentliche Gesundheit dar, mit jährlich 400 Millionen Infektionen und 40.000 Todesfällen. Die WHO prognostiziert, dass bis 2080 über 60% der Weltbevölkerung in Gebieten leben werden, die von *Ae. aegypti* besiedelt sind. Diese Risikozunahme wird durch die Auswirkungen des Klimawandels und die zunehmende Urbanisierung angetrieben, wobei Letzteres die Verfügbarkeit künstlicher Wasserbehälter erhöht, die von *Ae. aegypti* bevorzugt zur Eiablage genutzt werden. Eine wirksame Krankheitsbekämpfung erfordert ein umfassendes Verständnis der Verbreitung von *Ae. aegypti* und der Interaktionen zwischen Mücken und Menschen. Die Modellierung des Krankheitsübertragungsriskos wird jedoch durch das tagaktive Stechverhalten und die begrenzte Flugreichweite von *Ae. aegypti* erschwert, was insbesondere in heterogenen Stadtgebieten, die sich durch unterschiedliche Umweltbedingungen auszeichnen, zu einer hohen Variabilität in der Populationsdichte führen kann. Stichprobenartige entomologische Überwachungsmethoden, wie beispielsweise Ovitrap, die oft zur Steuerung von Interventionen herangezogen werden, können diese komplexen Dynamiken nur unzureichend erfassen. Ziel dieser Dissertation ist es, die Überwachung durch die Integration von georeferenzierten Daten zu verbessern, um die Krankheitsbekämpfung zu optimieren.

In dieser Dissertation wurden Methoden zur Generierung von Umwelt-Eignungsindikatoren für *Ae. aegypti* entwickelt, die aus georeferenzierten Daten extrahiert wurden und eine Interpolation von Ovitrap-Daten über heterogene städtische Strukturen unter Berücksichtigung der Flugreichweite von *Aedes* ermöglichen. Zudem wurde ein Ansatz zur Modellierung stündlicher *Aedes*-Mensch-Interaktionen vorgestellt, um das Stechrisko in lokale Interventionsleitlinien zu integrieren. Die praktische Anwendbarkeit dieser Methoden wurde am Fallbeispiel der Kommune Rio de Janeiro (1.221km<sup>2</sup>) demonstriert, um ihre Eignung für die Interventionsplanung auf kommunaler Ebene zu validieren, die in vielen Ländern eine zentrale Rolle in der Koordination von Interventionen zur Bekämpfung von *Aedes*-übertragenen Krankheiten spielt.

Die Ergebnisse zeigen, dass die mithilfe von Satelliten- und Straßensichtbildern geschätzte Dichte von Brutbehältern für *Aedes* ein signifikanter Umwelt-Eignungsindikator ( $p \leq 0,05$ ) für die Modellierung der mit 2.700 Ovitrap erfassten *Ae. aegypti*-Ei- und -Larvenzahlen sein kann. Die Signifikanz dieses Indikators variiert in Abhängigkeit vom Behältertyp, der Datenquelle und der Größe des modellierten *Aedes*-Lebensraums um die Ovitrap-Standorte. In Kombination mit weiteren Umwelt-Eignungsindikatoren für *Ae. aegypti*, die aus georeferenzierten Daten abgeleitet wurden - wie Wasseransammlungen, städtische Morphologie und klimatische Faktoren - erklärte der Indikator "Brutbehälterdichte" bis zu 75% der Variation der Ovitrap-Zahlen in dieser Fallstudie. Die daraus resultierende Modellierung der stündlich variierenden *Aedes*-Mensch-Interaktionen anhand von Mobilfunkdaten verbesserte die Vorhersage von Dengue-Fällen um 14% und trug zur Erklärung von 77% der räumlichen Verteilung gemeldeter Fälle über den 8-jährigen Zeitraum (2015-2022) bei. Zusammen können diese Erkenntnisse verbesserte Leitlinien für Interventionen liefern, die auf den im Rahmen der Forschung erstellten kontinuierlichen Umwelt-Eignungskarten für *Ae. aegypti*

und den identifizierten Hotspots für *Aedes*-Mensch-Interaktionen basieren. Herausforderungen, die die Modellgenauigkeit einschränken, bestehen weiterhin aufgrund heterogener Datenverfügbarkeit und zu grober Auflösung.

Die entwickelten Methoden zur Erfassung von Umweltfaktoren, die für Arbovirus-Infektionen relevant sind, haben das Potenzial, die Kosten im öffentlichen Gesundheitswesen zu senken, indem sie zur Erstellung präziserer Risikokarten genutzt werden, die eine gezieltere Durchführung von Kontrollmaßnahmen und eine effizientere Platzierung von Ovitrapns ermöglichen. Diese Erkenntnisse sind besonders bedeutsam, solange Vektorkontrollmaßnahmen wie *Wolbachia* und Insektizid-Sprühungen (i) aufgrund von Herausforderungen bei der Impfstoffentwicklung zu den wirksamsten Interventionen gegen *Aedes*-übertragene Arbovirus-Infektionen gehören und (ii) kostenintensiv bleiben oder unter Umweltauflagen stehen. Obwohl die Methoden auf die Kommune Rio de Janeiro zugeschnitten sind, könnten sie auf andere *Aedes*-endemische Gebiete mit ähnlichem ökologischen Kontext übertragbar sein, vorbehaltlich weiterer Forschung zur breiteren Anwendung und Integration in präskriptive Analysen, um den Übergang von vorausschauenden Erkenntnissen zu umsetzbareren Strategien zu erleichtern.

## List of Publications

This cumulative dissertation consists of the following publications:

S. Knoblauch, H. Li, S. Lautenbach, Y. Elshiaty, A. Rocha, B. Resch, D. Arifi, T. Jänisch, I. Morales, A. Zipf (2023). “Semi-supervised water tank detection to support vector control of emerging infectious diseases transmitted by *Aedes Aegypti*”. *International Journal of Applied Earth Observation and Geoinformation (Volume 119) Special Issue: Harnessing Geospatial Big Data for Infectious Diseases*. DOI: 10.1016/j.isprsjprs.2021.06.011 (**peer-reviewed, accepted**).

S. Knoblauch, M. Su Yin, K. Chatrinan, A. Rocha, P. Haddawy, F. Biljecki, S. Lautenbach, B. Resch, D. Arifi, T. Jänisch, I. Morales, and A. Zipf (2024). “High-resolution mapping of urban *Aedes aegypti* immature abundance through breeding site detection based on satellite and street view imagery”. *Scientific Reports: Urbanization and Communicable Diseases*. DOI: 10.1038/s41598-024-67914-w (**peer-reviewed, accepted**).

S. Knoblauch, R. Tadiwa Mukaratirwa, P.F. Paolucci Pimenta, A. Wilder-Smith, J. Rocklöv, S. Lautenbach, S. Randhawa, O.J. Brady, A. Rocha, P. Dambach, T. Jänisch, B. Resch, F. Biljecki, T. Bärnighausen, A. Zipf. “Suitability indicators for immature *Aedes aegypti* to guide vector control in the municipality of Rio de Janeiro”. Submitted to *The Lancet Planetary Health* (**under review**).

S. Knoblauch, S. Gross, S. Lautenbach, A. Rocha, M. C. Gonzalez, B. Resch, D. Arifi, T. Jänisch, I. Morales, A. Zipf (2024). “Long-term validation of inner-urban mobility metrics derived from Twitter/X”. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/23998083241278275 (**peer-reviewed, accepted**).

S. Knoblauch, J. Heidecke; M. Reinmuth, S. Lautenbach, P. Filemon Paolucci Pimenta, J. Rocklöv, O.J. Brady, A. Rocha, T. Jänisch, B. Resch, F. Biljecki, A. Wilder-Smith, T. Bärnighausen, A. Zipf. “Assessing Dengue Risk in Urban Areas: The Role of Daytime *Aedes*-Human Interaction”. Submitted to *Scientific Reports: Urbanization and Communicable Diseases* (**under review**).



## List of Abbreviations

<b>AIC</b>	Akaike Information Criterion
<b>BI</b>	Breteau Index
<b>BTI</b>	Bacillus Thuringiensis Israelensis
<b>CHIKV</b>	Chikungunya Virus
<b>CSI</b>	Chitin Synthesis Inhibitors
<b>DENV</b>	Dengue Virus
<b>GIS</b>	Geographic Information System
<b>GeoAI</b>	Geospatial Artificial Intelligence
<b>GBD</b>	Geospatial Big Data
<b>GPS</b>	Global Positioning System
<b>GSV</b>	Google Street View
<b>HI</b>	House Index
<b>IDW</b>	Inverse Distance Weighting
<b>INLA</b>	Integrated Nested Laplace Approximation
<b>IGR</b>	Insect Growth Regulator
<b>JHA</b>	Juvenile Hormone Analogs
<b>LIRAA</b>	Larval Infestation Rapid Assay <i>Aedes</i>
<b>LOESS</b>	Locally Weighted Scatterplot Smoothing
<b>NGO</b>	Non-Governmental Organization
<b>M</b>	Method
<b>MET</b>	Mean Eggs Per Trap
<b>ML</b>	Machine Learning
<b>MLT</b>	Mean Larvae Per Trap
<b>MBD</b>	Mosquito-Borne Diseases
<b>NB-GLM</b>	Negative-Binomial Generalized Linear Model
<b>OSM</b>	OpenStreetMap
<b>PCR</b>	Polymerase Chain Reaction
<b>PCA</b>	Principal Component Analysis
<b>QP-GLM</b>	Quasi-Poisson Generalized Linear Model
<b>RO</b>	Research Objective
<b>RQ</b>	Research Question
<b>SSST</b>	Semi-Supervised Self-Training
<b>SPDE</b>	Stochastic Partial Differential Equation
<b>VBD</b>	Vector-borne diseases
<b>WHO</b>	World Health Organization
<b>YFV</b>	Yellow Fever Virus
<b>ZIKV</b>	Zika Virus



# Contents

<b>Acknowledgments</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>List of Publications</b>	<b>vii</b>
<b>List of Abbreviations</b>	<b>ix</b>
<b>Contents</b>	<b>xi</b>
<b>I Synopsis</b>	<b>1</b>
1 Introduction . . . . .	5
2 Background . . . . .	6
2.1 Geospatial Big Data . . . . .	6
2.2 <i>Aedes</i> -borne Arboviral Infections . . . . .	8
2.3 Study Area: Municipality of Rio de Janeiro . . . . .	14
3 Contribution of This Thesis . . . . .	16
3.1 Main Research Objectives and Questions . . . . .	16
3.2 Research Framework and Core Methods . . . . .	23
3.3 Results and Discussion . . . . .	36
3.4 Research in Context . . . . .	54
3.5 Future Directions Towards Practical Application . . . . .	56
3.6 Conclusion . . . . .	63
<b>II Publications</b>	<b>87</b>
Publication I:	
Semi-supervised Water Tank Detection to Support Vector Control of Emerging Infectious Diseases Transmitted by <i>Aedes aegypti</i> . . . . .	89
1 Introduction . . . . .	89
2 Experimental Design . . . . .	92
3 Results and Discussion . . . . .	98
4 Conclusion . . . . .	103

Publication II:	
High-resolution Mapping of Urban <i>Aedes aegypti</i> Immature Abundance through Breeding Site Detection based on Satellite and Street View Imagery . . . . .	
	115
1	Introduction . . . . .
	115
2	Methods . . . . .
	117
3	Results and Discussion . . . . .
	121
Publication III:	
Urban <i>Aedes aegypti</i> suitability indicators . . . . .	
	135
1	Introduction . . . . .
	136
2	Materials and Methods . . . . .
	137
3	Results . . . . .
	142
4	Discussion . . . . .
	147
5	Conclusion . . . . .
	150
Publication IV:	
Long-term Validation of Inner-Urban Mobility Metrics Derived From Twitter/X . . . . .	
	177
1	Introduction . . . . .
	177
2	Materials and Methods . . . . .
	179
3	Results and Discussion . . . . .
	187
4	Conclusion . . . . .
	194
Publication V:	
Assessing Dengue Risk in Urban Areas: The Role of Daytime <i>Aedes</i> -Human Interaction . . . . .	
	207
1	Introduction . . . . .
	207
2	Materials and Methods . . . . .
	209
3	Results . . . . .
	216
4	Discussion . . . . .
	218

### III Declarations

**233**

**Part I**  
**Synopsis**



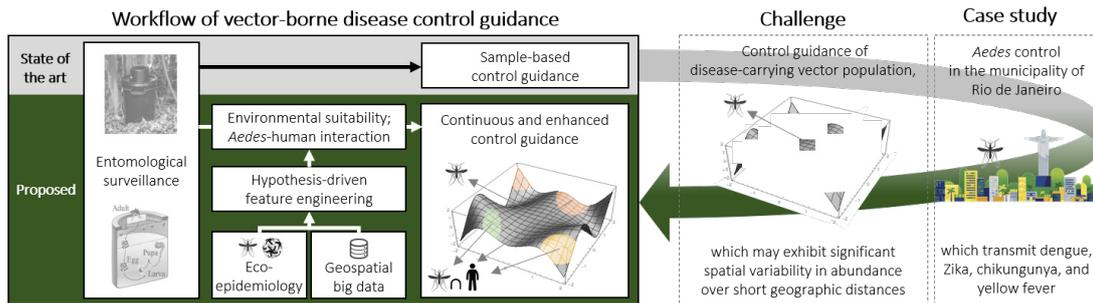
*»All models are wrong, but some are useful.«*

– George E.P. Box (1976)



# 1 Introduction

In recent years, the integration of big data analytics into public health has transformed approaches to disease prevention and control (Dolley, 2018). Among the various data sources being leveraged, geospatial big data (GBD) has emerged as a pivotal tool for understanding and mitigating the spread of infectious diseases (Mir et al., 2022; Oliver et al., 2020; Shah and Patel, 2020). In the field of infectious disease modeling, GBD has shown particular promise for vector-borne diseases (VBDs), where human movement patterns - derivable from GBD sources - not only influence the dissemination of pathogens but also affect the exposure risk to vectors (Iggidr et al., 2017; Ramadona et al., 2019). Traditionally, entomological surveillance has been the cornerstone of VBD risk assessment and intervention guidance, primarily relying on field-based sampling methods for monitoring vector populations (World Health Organization et al., 2016). However, this approach is constrained by high labor costs, which limits both spatial coverage and temporal resolution. These limitations are particularly challenging for VBDs, where vector abundance and associated infection risk can vary significantly over short geographic distances, e.g., due to small vector habitats in combination with high spatial variability in environmental circumstances (Kache et al., 2022a; Killeen et al., 2018).



**Figure 1:** Comparison of the state-of-the-art and proposed workflow for VBD control guidance, highlighting the addressed challenge and the selected case study in this dissertation.

This cumulative dissertation investigates the integration of GBD for improving VBD surveillance and control, with a specific focus on *Aedes*-borne arboviral infections. It addresses the challenges posed by high spatial variability in vector populations and the limitations of traditional surveillance methods by exploring the potential of GBD to enhance the guidance of local intervention strategies. Part I in this thesis is structured into two main sections. The first section presents the background of GBD and *Aedes*-borne arboviral infections and motivates the selection of the municipality of Rio de Janeiro as a suitable case study region for the analysis. The second section outlines the primary research objectives and questions addressed in this thesis, detailing the innovative contributions. It references the relevant publications included in Part II, summarizing key findings, discussions, conclusions, and future perspectives that highlight the impact of this work at the intersection of geoinformatics and public health.

## 2 Background

### 2.1 Geospatial Big Data

GBD refers to the vast volumes of spatial and geographical information that are collected from a variety of sources, including satellite imagery, Global Positioning System (GPS) devices, and remote sensing technologies (Werner, 2021).

The key characteristics of GBD can be described as its high volume, velocity, variety, and veracity, underscoring its complexity and its potential for spatial analysis in various application fields (Werner, 2021). The volume of GBD is exceptionally large, often reaching petabyte scales, due to the continuous collection of information from diverse sources such as satellites, GPS devices, and Internet of Things sensors. The velocity at which this data is generated is another crucial attribute, necessitating efficient management and, in some cases, prompt analysis, particularly when data processing occurs at the edge rather than in centralized systems. GBD also exhibits variety, encompassing various forms such as raster data from satellite imagery, vector data from maps, and textual data from geotagged social media posts. This diversity requires sophisticated data integration techniques to enable comprehensive analysis. Finally, veracity is a key characteristic, as the accuracy and reliability of GBD are contingent on factors such as sensor precision, data resolution, and the methodologies employed in data collection. Ensuring data quality through rigorous validation and error correction is essential for producing reliable and actionable insights.

The sources of GBD are varied and technologically advanced, each offering unique advantages that justify the use of complementary approaches, such as Principal Component Analysis (PCA), which enables the extraction of underlying structures and patterns, facilitating a more comprehensive understanding of complex spatial phenomena, and mitigating potential biases inherent in single data types (Gao et al., 2024). Satellite imagery stands out as a primary source, providing high-resolution images of the Earth’s surface. Ongoing advancements in remote sensing technologies have significantly enhanced the resolution and frequency of these images, increasing their applicability across various domains such as environmental monitoring (Rolf et al., 2021). Furthermore, street view imagery offers ground-level visual data that complements satellite imagery by providing more detailed perspectives of built-up environments, making it particularly useful for urban analytics or studying human perceptions (Hou et al., 2024; Ito et al., 2024; Zhang et al., 2018). In parallel, GPS and mobile devices contribute vast amounts of location data, which are crucial for navigation and location-based services. The ubiquity of these devices ensures a constant flow of GBD that reflects human mobility patterns and spatial dynamics, adding a temporal dimension that other sources might lack (Kwan and Schwanen, 2017; Maddison and Ni Mhurchu, 2009). Additionally, social media represent emerging sources of GBD, where geotagged information from platforms like X (formally known as Twitter) offers real-time insights into human activities (Terroso-Saenz et al., 2022). Further, crowdsourced data, such as that from the volunteered geographic information platform OpenStreetMap (OSM), provides detailed, user-generated geographic data, which is crucial for maintaining up-to-date knowledge, for example, on urban morphology (Boeing, 2021; Sussman and Hollander, 2021).

The fields of application for GBD are diverse, including environmental moni-

toring, disaster management, or public health, where it can be specifically applied to estimate the environmental suitability of disease vectors such as mosquitoes. This approach differs from monitoring adult mosquitoes, as it involves predicting habitats and conditions favorable for their abundance. These insights can be crucial for public health officials, enhancing vector control strategies by enabling more efficient resource management and targeted interventions (Louis et al., 2014). In the same domain, GBD can be instrumental in tracking human movement patterns, enabling the identification of host-vector interactions. This knowledge further supports the design of targeted intervention strategies specifically aimed at disrupting the transmission cycle of vector-borne diseases (Martinez et al., 2021; Wen et al., 2015). In summary, GBD holds the potential to enhance the preparedness and effectiveness of public health efforts, ultimately contributing to improved disease prevention and control.

However, harnessing GBD comes with several challenges. One of the primary difficulties is the substantial computational power and advanced analytical techniques, such as parallel processing, required to manage and analyze the vast volumes of data generated (Gao et al., 2024). Another challenge lies in data integration, where combining data from multiple sources with varying formats, resolutions, and quality can be complex and time-consuming. Effective integration is crucial for generating accurate insights, yet it requires sophisticated methods to ensure consistency and compatibility across datasets (Werner, 2021). Additionally, privacy and security concerns are paramount, especially when dealing with personalized datasets. Ensuring that personal data is properly anonymized and protected is essential for maintaining public trust in innovative solutions and complying with legal regulations, making it a critical aspect of any GBD project.

The potential of GBD is vast, with advancements in technology expected to further enhance the resolution, frequency, and accuracy of GBD collection, thereby offering increasingly detailed and timely data (Werner, 2021). Moreover, the integration of artificial intelligence (AI) into GBD analysis holds the promise of revolutionizing the field, enabling more insights. These technologies will enhance pattern detection, trend prediction, and the automation of complex analyses. The proliferation of Internet of Things devices, such as smart mosquito traps that can automatically monitor mosquito densities using optical and acoustic sensors, represents another significant area of growth, though these devices are still relatively costly and not widely rolled out yet. However, they enable continuous monitoring of environmental conditions and vector populations, providing near real-time data that can potentially enhance vector control strategies (Liu et al., 2023). Finally, open data initiatives are expected to gain momentum, with growing efforts to make GBD publicly accessible. Projects like TensorFlow, a free and open-source software library for machine learning (ML) developed by the Google Brain team (Alphabet Inc., 2024), exemplify the trend of sharing advanced research tools. These initiatives promote innovation, collaboration, and the establishment of new data standards, such as those seen with the Overture Maps Foundation, a mapping platform supported by the competing corporate members AWS, Meta, Microsoft, and TomTom. By making GBD available to researchers, businesses, and governments, these efforts enable broader use of open-source development to expand the applicability of GBD and, in our context, provide more detailed, timely information for societal benefit.

In conclusion, GBD holds immense potential for addressing a broad spectrum of challenges, with its development still in the early stages, where scientific research plays a pivotal role. Through pioneering studies and proof-of-concept initiatives, science can lay the groundwork for leveraging advanced technologies and open-source analytical methods, ultimately driving informed decision-making in critical areas.

## 2.2 *Aedes*-borne Arboviral Infections

*Aedes* mosquitoes, belonging to the family Culicidae, are a diverse genus comprising over 950 species (Harbach, 2007). Among them, *Ae. aegypti* and *Ae. albopictus* are the most prominent due to their role as vectors of human arboviruses, including dengue, Zika, chikungunya, and yellow fever viruses (Panigrahi et al., 2024). These species exhibit distinct ecological and behavioral traits that facilitate their ability to thrive in various environments. The life cycle of *Aedes* mosquitoes includes four stages: egg, larva, pupa, and adult. Females typically lay eggs on the inner, damp walls of containers with water, where the larvae then develop. The larvae and pupae are aquatic, requiring standing water to complete their development. Adult *Aedes* mosquitoes are adapted to urban and suburban environments, often breeding in artificial containers such as flower pots, discarded tires, and water storage containers (Sallam et al., 2017). *Aedes* mosquitoes exhibit diurnal feeding patterns, with peak biting activity occurring during early morning and late afternoon (Muhammad et al., 2020; Mutebi et al., 2022a; Zahid et al., 2023). Biting activity is generally low during the day and, when present, tends to occur under low-light conditions, such as in shaded areas. However, some studies have observed increased biting behavior under artificial light conditions at night, even though *Aedes* mosquitoes are typically inactive during night hours (Rund et al., 2020). Additionally, the flight range of these species is estimated to be less than 1,000 meters without the assistance of wind. *Aedes* mosquitoes are highly adaptable and have expanded their geographic distribution significantly over the past few decades, largely due to global trade and travel (Kraemer et al., 2019; Semenza et al., 2014; Swan et al., 2022; Willoughby et al., 2024). *Ae. aegypti* is predominantly found in tropical and subtropical regions, while *Ae. albopictus* has a broader range, extending into temperate zones. Understanding the taxonomy, life cycle, and global distribution of *Aedes* mosquitoes is crucial for developing effective vector control strategies and mitigating the spread of MBDs.

*Aedes* mosquitoes are primary vectors for several arboviruses that pose significant public health threats globally (Leta et al., 2018). Dengue virus (DENV), with its four distinct serotypes (DENV-1 to DENV-4), is transmitted primarily by *Ae. aegypti* and *Ae. albopictus*. Dengue is endemic in over 100 countries, leading to millions of infections annually, characterized by fever, severe joint and muscle pain, and in severe cases, hemorrhagic fever and shock syndrome (Guzman and Harris, 2015; Halstead, 2007; Wilder-Smith et al., 2019). Zika virus (ZIKV), first identified in Uganda in 1947, gained global attention during the 2015-2016 epidemic in the Americas. It is transmitted by *Aedes* mosquitoes and can cause mild symptoms such as fever and rash, but it is particularly concerning due to its association with congenital abnormalities like microcephaly and neurological disorders such as Guillain-Barré syndrome (Musso and Gubler, 2016; Petersen et al.,

2016). Chikungunya virus (CHIKV) also relies on *Aedes* mosquitoes for transmission and causes an illness marked by severe, debilitating joint pain, fever, and rash, with outbreaks frequently occurring in Africa, Asia, and the Indian subcontinent (Pialoux et al., 2007; Sudeep and Parashar, 2008; Vu et al., 2017). Yellow fever virus (YFV), which can be transmitted in both urban and sylvatic cycles by *Aedes* and other mosquito species, causes a spectrum of disease from mild symptoms to severe liver damage and hemorrhagic fever; it is preventable by vaccination, which is crucial for outbreak control (Monath and Vasconcelos, 2015). In addition to these major pathogens, *Aedes* mosquitoes also transmit other arboviruses such as Mayaro and Rift Valley fever viruses, which are less common but still significant due to their potential to cause outbreaks and severe disease. Understanding the epidemiology, transmission dynamics, and impact of these arboviruses is essential for developing targeted public health interventions and controlling their spread.

The epidemiology of *Aedes*-borne arboviral infections reveals significant global and regional trends marked by periodic outbreaks and shifting endemic zones (Chang et al., 2016; Guo et al., 2017). Current statistics indicate a rising incidence of diseases like dengue and Zika, with dengue alone affecting an estimated 390 million people annually, predominantly in tropical and subtropical regions (Vasconcelos, 2017). Historical trends show a dramatic increase in the geographical range of these viruses, influenced by the impacts of accelerating urbanization, globalization, and climate change (Messina et al., 2019) (cf. Figure 1 in publication IV). Key risk factors for the spread include environmental conditions such as warm temperatures and stagnant water bodies, demographic factors like high population density and international travel, and behavioral aspects such as inadequate waste management and water storage practices. Effective surveillance and reporting systems are crucial for monitoring these infections, employing methods like geographic information systems (GIS) for mapping outbreaks, polymerase chain reaction (PCR) for the accurate diagnosis of transmitted pathogens, including serotypes, and integrated disease surveillance programs to ensure timely reporting and response. The epidemiology of Dengue is particularly challenging due to complex cross-immunity between subsequent serotypes (DENV-1 to DENV-4). In brief, past infections with a heterologous serotype confer short-term cross-immunity, while past infections with a homotypic serotype confer long-term immunity to the same serotype. The duration and effect size of heterologous cross-immunity, as well as potential enhancement, depend on the time interval between infections and the specific sequence of serotypes and their genotypic similarity (Guzman et al., 2016; Katzelnick et al., 2015; Simmons et al., 2012).

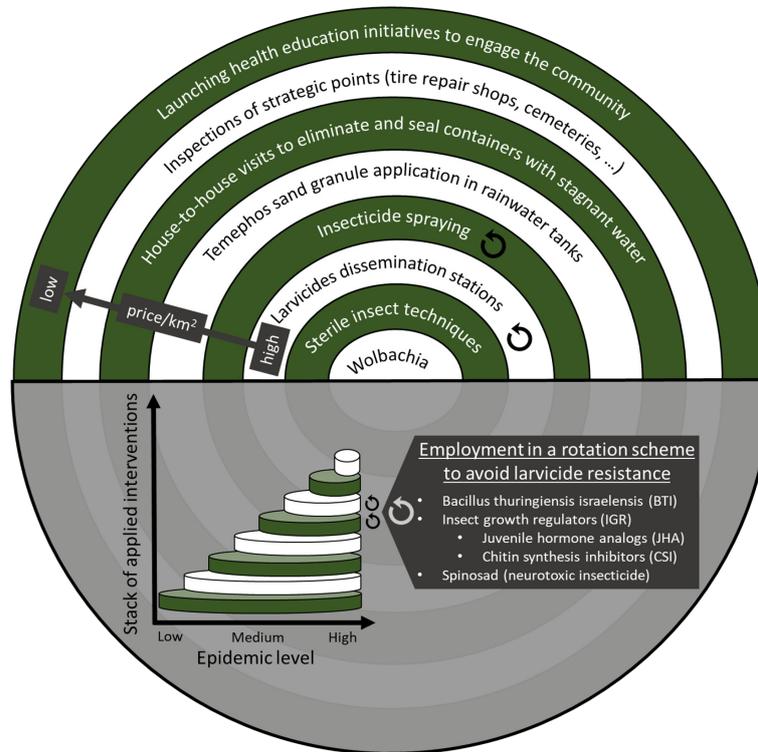
Prevention and control of *Aedes*-borne diseases rely on a multi-faceted approach involving two main strategies: (i) vector control, which focuses on reducing mosquito populations and minimizing human exposure, and (ii) vaccines and therapeutics, which focus on protecting humans.

Vector control strategies are selected based on their relevance at different stages of an epidemic. Depending on the epidemic level, a specific combination of interventions is typically chosen: some strategies are implemented continuously, while others are considered only during high epidemic levels due to their cost (cf. Figure 2). Vector control strategies designed for constant application may include (i) community education programs to raise awareness about prevention methods,

(ii) legislative actions to enforce vector control regulations and public health policies that promote sustainable environmental management and urban planning to reduce mosquito habitats (Buhler et al., 2019; Horstick et al., 2010; Luz et al., 2011), (iii) house-to-house visits to eliminate stagnant water in common breeding sites, (iv) inspections of strategic locations such as cemeteries, tire repair shops, junkyards, scrap metal or building materials depots, and bus garages, and (v) residual household spraying targeting *Aedes* mosquitoes. When launching health education initiatives to engage the community, it is crucial to consider additional socio-economic factors. Individuals from different socio-economic backgrounds may have varying priorities, such as safety, food security, access to clean water and sanitation facilities, healthcare services, education, and employment opportunities. When the epidemic situation becomes more severe, additional measures can be implemented collectively, incorporating more technical methods such as (i) the application of the larvicide Temephos to rainwater tanks, (ii) dissemination stations containing larvicides, (iii) sterile insect techniques, or (iv) the *Wolbachia* method. These approaches can be crucial for addressing structural problems related to socio-economic inequalities in water supply and solid waste management. *Wolbachia*, a naturally occurring bacterium, can be introduced into mosquito cells to reduce the transmission of viruses by *Ae. aegypti* and influence mating outcomes, thereby promoting its spread and sustainability within natural mosquito populations. Larvicides should be employed in a rotational scheme to avoid larvicide resistance. In Brazil the products used include *Bacillus thuringiensis israelensis* (BTI), insect growth regulators (IGRs) such as juvenile hormone analogs (JHAs) and chitin synthesis inhibitors (CSIs), and more recently, Spinosad, a neurotoxic insecticide (Valle et al., 2019). The use of insecticides to reduce adult mosquito populations and biological control methods, such as introducing natural predators or genetically modified mosquitoes, becomes particularly relevant during epidemic situations, as these strategies have demonstrated high efficacy in vector control (Vinhai Frutuoso and Barbosa Duraes, 2023). However, most methods are either costly to implement or face environmental constraints, which is why they are not applied on a larger scale year-round. To reduce associated costs, these interventions are usually deployed in a targeted manner, guided by entomological surveillance. Their use may be triggered only when certain thresholds of egg and larval density are reached.

Vaccines and therapeutics are another vital components in the fight against *Aedes*-borne arboviral infections. Currently, vaccines are available for yellow fever, which has been highly effective in preventing outbreaks, and the dengue vaccine (Dengvaxia), approved in several countries for individuals with prior dengue infection (Aguiar et al., 2016; Schwartz et al., 2015; Thomas and Yoon, 2019; Tully and Griffiths, 2021). Ongoing research is focused on developing more universally effective vaccines for dengue, as well as for Zika and chikungunya, which are in various stages of clinical trials. Antiviral treatments remain limited, with supportive care being the primary management for most *Aedes*-borne diseases. Experimental treatments and research into specific antiviral drugs are ongoing, aiming to provide targeted therapies to reduce viral load and improve patient outcomes in the future. In addition to conventional vector control methods and vaccines, are personal protection measures essential for the prevention and control of *Aedes*-borne arboviral infections. These measures include using repellents,

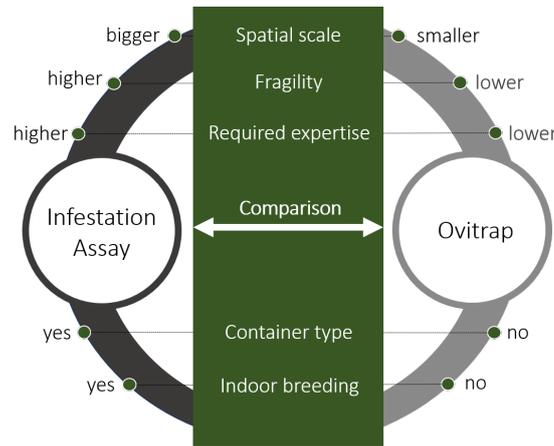
wearing protective clothing, and sleeping under insecticide-treated bed nets to prevent mosquito bites. Targeted public education about personal protection can, in turn, be an integral part of vector control strategies.



**Figure 2:** Catalog of local intervention strategies for the control of *Aedes*-borne arboviral infections, where interventions in the outer circles represent those that should be conducted universally during both non-pandemic and pandemic periods, and interventions in the inner circles represent more costly interventions or those that face environmental constraints and, due to limited financial resources, are mostly implemented in a targeted manner during high epidemic levels. While the figure shows a gradient from lower to higher costs per km<sup>2</sup> from the outer to inner circles, it is important to note that the overall implementation costs are not solely determined by the size of the application area but also by the frequency with which interventions must be applied to reach and maintain effectiveness. This nuance is crucial to avoid potential misinterpretation of the cost gradient.

Targeted prevention and control need sophisticated guidance to be effective and feasible, especially when financial resources are limited or environmental regulations must be adhered to. For *Aedes*-borne arboviral infections, intervention guidance is primarily informed by either disease occurrence or entomological surveillance. Entomological surveillance may include monitoring *Aedes* egg and larval counts using ovitraps (cf. Figure 3 in publication I), tracking adult mosquitoes with smart traps based on acoustic or optical sensors, or conducting infestation assay to collect entomological indices that can guide targeted interventions (cf. Figure 3). Ovitrap refers to fixed measurement stations specifically designed to function as breeding sites. These stations are characterized by water retention and a dark color to attract mosquitoes. Infestation assay indices are block-level indicators derived from manual sampling processes and can include: (i) the House Index (HI), which measures the proportion of houses infested with larvae relative to the total number of houses surveyed, and (ii) the Breteau Index (BI), which represents the number of positive containers per 100 houses inspected.

While infestation assay indices depend on the active search for breeding sites by health agents, ovitraps are 'sought' by mosquitoes, making ovitrap counts a more sensitive indicator. Additionally, infestation assay indices typically yield low values due to the transient nature of positive containers and their limited persistence over time. Therefore, to obtain a comprehensive and representative sample, it is necessary to include a large number of properties, as most may not have positive containers. However, from an operational perspective, ovitraps cannot fully replace infestation assays, as the latter often provide additional indices, such as the 'container type index,' which monitors infestation levels by container type. Beyond structural components, the reliability of infestation assay indices also depends on human factors, such as the dedication and expertise of field workers, including their understanding of vector biology and index calculation methods (Valle and Aguiar, 2023a). Another factor contributing to spatial discrepancies is that ovitrap-based surveillance does not account for indoor breeding sites. Moreover, ovitrap surveillance focuses on egg and larval counts, whereas infestation assay indices usually also account for infestation by *Ae. aegypti* pupae, which have different lifespans and lower mortality rates.

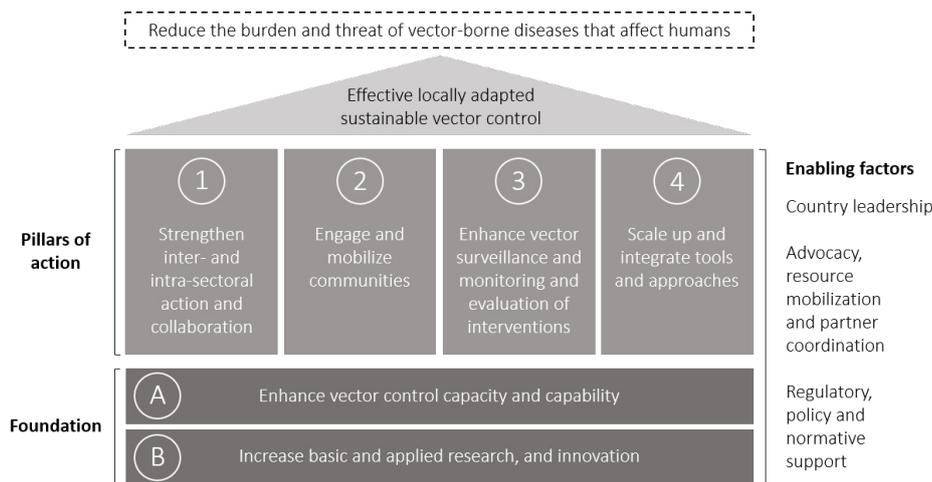


**Figure 3:** Comparison of infestation assay and ovitraps for entomological surveillance.

The global burden of *Aedes*-borne arboviral infections on public health is profound, manifesting in significant morbidity and mortality. These diseases cause a considerable health burden, with millions of cases reported annually, leading to severe symptoms, long-term complications, and thousands of deaths, particularly in endemic regions (Puntasecca et al., 2021; Shepard et al., 2016; Stanaway et al., 2016). The economic impact is substantial, estimated to be at least 94.7 billion US dollars (cumulative for the period 1975–2020), encompassing direct costs such as medical care and hospitalization, and indirect costs like lost productivity and strain on healthcare systems (Roiz et al., 2024). Socioeconomic factors play a critical role in the spread and impact of these diseases, with impoverished communities often experiencing higher transmission rates due to inadequate housing, limited access to healthcare, and lack of resources for effective vector control measures. Addressing these factors is crucial for mitigating the public health impact of these infections.

Future directions in combating *Aedes*-borne arboviral infections focus on addressing current research gaps, developing innovative strategies, and enhancing

global health initiatives. Identifying knowledge gaps, such as understanding the genetic and environmental factors influencing virus transmission and mosquito behavior, is essential for advancing research. Innovative strategies include the development of novel vector control technologies like *Wolbachia*-infected mosquitoes, which reduce mosquito lifespan and viral transmission, and next-generation vaccines that provide broad protection against multiple arboviruses. Global health initiatives emphasize international collaboration, improved surveillance systems, and coordinated response efforts to manage outbreaks and prevent the spread of these diseases. In this context, harnessing GBD for improved guidance on interventions targeting *Aedes*-borne arboviral infections could play a crucial role (Li and Dong, 2022). Strengthening inter- and intra-sectoral collaboration, enhancing vector surveillance and monitoring, and scaling up and integrating tools and approaches were listed as key pillars in the World Health Organization’s (WHO) Global Vector Control Response Framework for 2017–2030 (WHO, 2017). Strengthening these areas of action from a scientific perspective will be pivotal in reducing the burden of *Aedes*-borne arboviral infections worldwide (Makepeace and Gill, 2016; McCall and Lenhart, 2008; Muñoz et al., 2020; Murray et al., 2013; Werren et al., 2008).



**Figure 4:** WHO global vector control response framework (WHO, 2017).

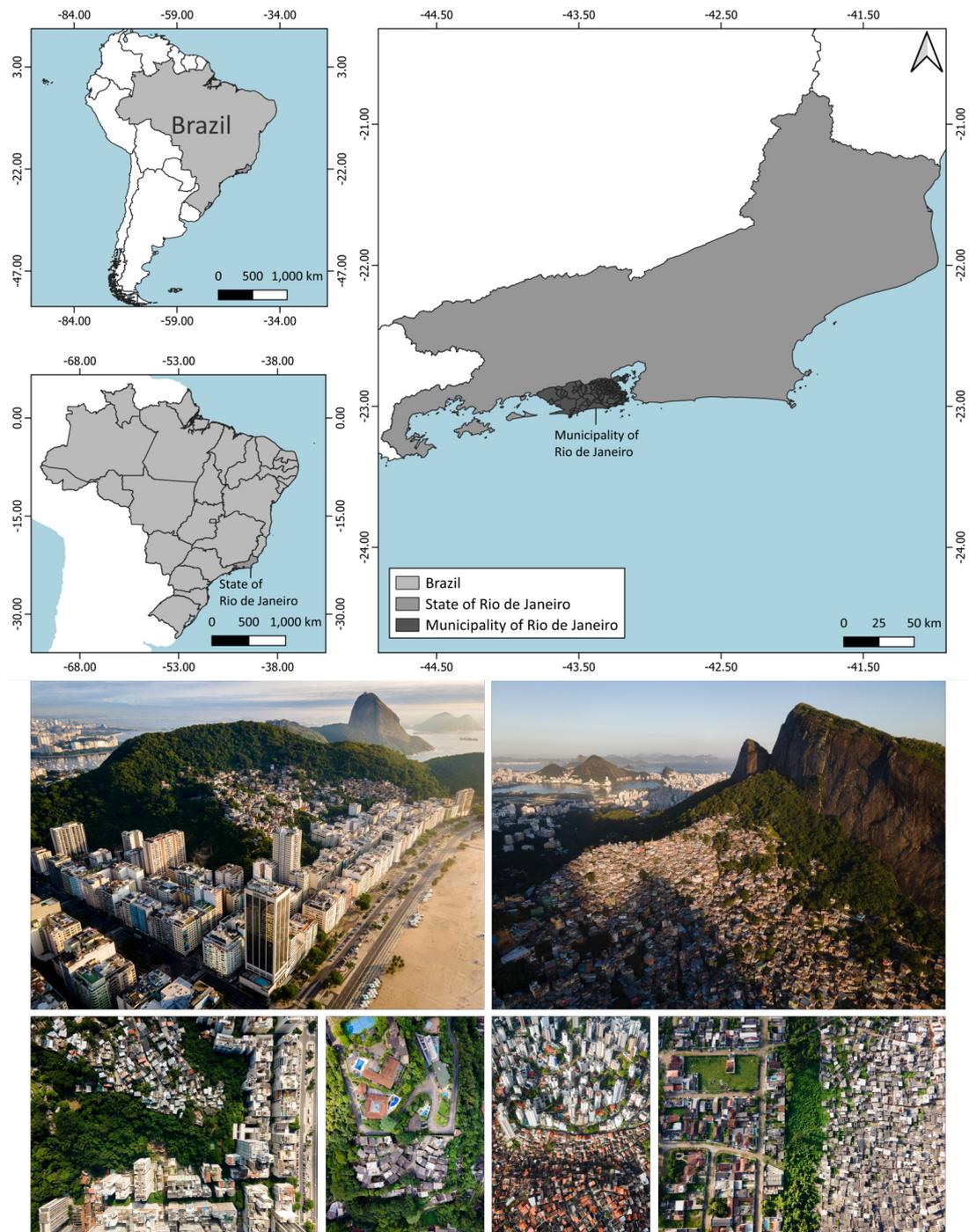
In Brazil, which consistently ranks highest in global dengue prevalence, including in 2024, the “National guidelines for the prevention and control of arboviruses” prioritize interventions in large cities. Between 2013 and 2022, 52% of probable dengue cases were reported in municipalities with populations exceeding 100,000. The Brazilian Ministry of Health recommends several key strategies for controlling *Aedes*-borne arboviruses: (i) entomological monitoring using ovitraps, (ii) household residual spraying (BRI-*Aedes*), (iii) the deployment of larvicide spraying stations, (iv) the release of mosquitoes carrying *Wolbachia*, and (v) the utilization of sterile insect techniques to control *Ae. aegypti*. These interventions are implemented based on an local action plan that require ovitrap-based risk stratification and are supplemented by household visits and community engagement, particularly in high-risk areas (Vinhai Frutuoso and Barbosa Duraes, 2023).

### 2.3 Study Area: Municipality of Rio de Janeiro

The municipality of Rio de Janeiro was selected as the case study region for this research on harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections due to several compelling factors. The city's suitable climate for *Aedes* mosquitoes is characterized by hot, humid summers and mild, wet winters, providing optimal conditions for the proliferation of *Aedes* mosquitoes, which thrive in warm, humid environments with plentiful water containers for breeding (Neiva et al., 2017). Additionally, Rio de Janeiro is an endemic region for *Aedes*-borne diseases, with a history of frequent and severe outbreaks, such as the dengue emergency during Carnival 2024, where the local health infrastructure was stretched to its limits (CNN World, 2024). Between January 2015 and December 2022, approximately 95,000 dengue cases, 42,000 Zika cases, and 68,000 chikungunya cases were reported, emphasizing the critical need for research and improvement of control measures in this region. The city's heterogeneous urban landscape and socio-economic disparities add significant complexity to guiding interventions effectively. Unlike cities with more homogeneous landscapes or evenly distributed topography, Rio de Janeiro's urban fabric includes dense urban areas, sprawling informal settlements often scattered throughout the city on steep hill-sides, and relatively undeveloped green spaces (O'Hare and Barke, 2002). This varied and complex topography makes it more challenging to predict and manage the distribution and availability of *Aedes* breeding sites, thereby complicating efforts to implement targeted interventions and increasing the necessity for advanced geospatial analysis techniques.

With a large population, Rio de Janeiro is the second-largest city in Brazil, making it a significant area at risk for *Aedes*-borne diseases. The city's size (1,221 km<sup>2</sup>) and high population count (6.2 million (Instituto Brasileiro de Geografia e Estatística, 2024) heighten the need for effective vector control strategies, aligning with the Brazilian Ministry of Health's focus on cities with populations over 100,000 inhabitants for concentrated intervention measures (Vinhali Frutuoso and Barbosa Duraes, 2023). Moreover, Brazil is recognized for its innovative public health strategies, and the city of Rio de Janeiro has shown strong interest in integrating GBD into their vector control efforts. This research was developed in close cooperation with Fiocruz and Rio's Center for Epidemiological Intelligence, ensuring that the findings are directly applicable and beneficial to ongoing public health initiatives in the city. Furthermore, Rio de Janeiro's membership in the C40 Cities Network reflects its commitment to innovative and sustainable urban solutions, including addressing public health threats posed by climate change and vector-borne diseases. The municipality's open data initiative, particularly through the Data Rio portal, provided a wealth of datasets necessary for this research. Importantly, the city also shared non-public data, including mobile phone records, entomological surveillance data, and detailed clinical records, with ethical approval and restricted use specifically for this research. These clinical records are particularly valuable, containing PCR test results to determine dengue serotypes, lists of symptoms, patient age, and precise home addresses. The availability of such detailed and sensitive data, coupled with proper anonymization, was crucial in enabling the comprehensive analysis conducted in this study. The science-friendly data sharing culture from the Brazilian Health Ministry, the support from the Brazilian World Mosquito Program, and close collaboration with the

Universidade Federal Fluminense as well as the Non-Governmental Organization (NGO) project “Heróis Contra Dengue” to conduct science communication and understand community needs provided a favorable environment for this research, aligning with their shared interest in combining geoinformatics and public health to tackle contemporary challenges. This institutional backing, coupled with the city’s commitment to data sharing and innovative public health strategies, underscores the relevance and impact of this work in the municipality of Rio de Janeiro.

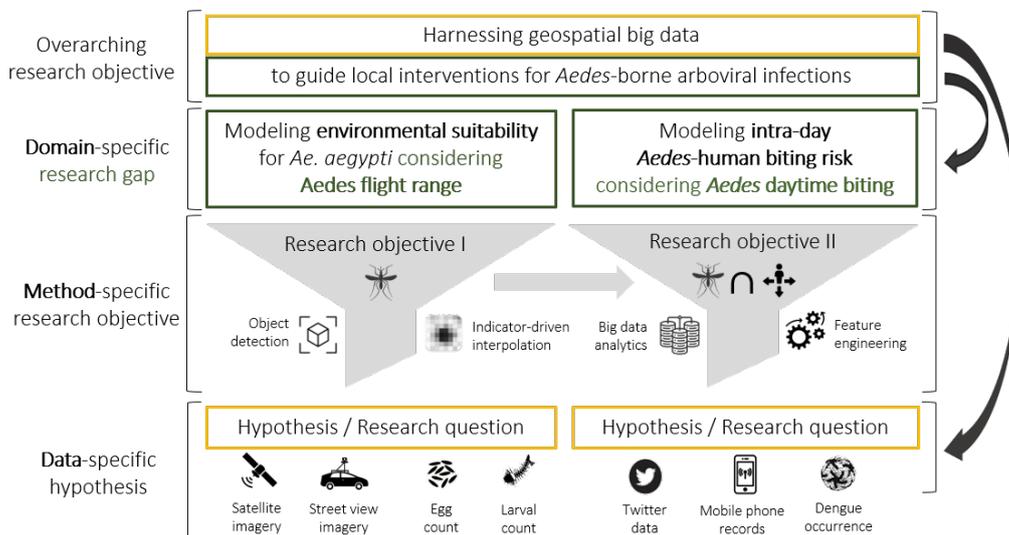


**Figure 5:** Geographical location of the municipality of Rio de Janeiro (top) and representative images illustrating the diverse urban landscape within the study area (bottom). © iStock

### 3 Contribution of This Thesis

#### 3.1 Main Research Objectives and Questions

The overarching research objective of this thesis is “*harnessing geospatial big data to guide local interventions for Aedes-borne arboviral infections.*” To fulfill this objective, two more specific research objectives were identified through a thorough literature review, each addressing distinct aspects of this topic, as detailed in the following paragraphs. For each research objective, a specific hypothesis and a corresponding research question were formulated, emphasizing the most promising strategies for leveraging GBD in guiding local interventions for *Aedes*-borne arboviral infections, taking into account current GBD availability and practical implementation potential. While various research paths could have been explored, the chosen approach was carefully selected to target specific research gaps, thereby maximizing the potential impact of this study.



**Figure 6:** Breakdown of the overarching research objective into two specific research objectives, followed by the derivation of individual hypotheses, from which corresponding research questions were formulated. The two specific research objectives were formulated to address the overarching objective of harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections. These objectives, along with their associated hypotheses and research questions, were chosen to explore the most promising research pathways, considering the available GBD at the time of analysis. Each research question was formulated to test a specific hypothesis.

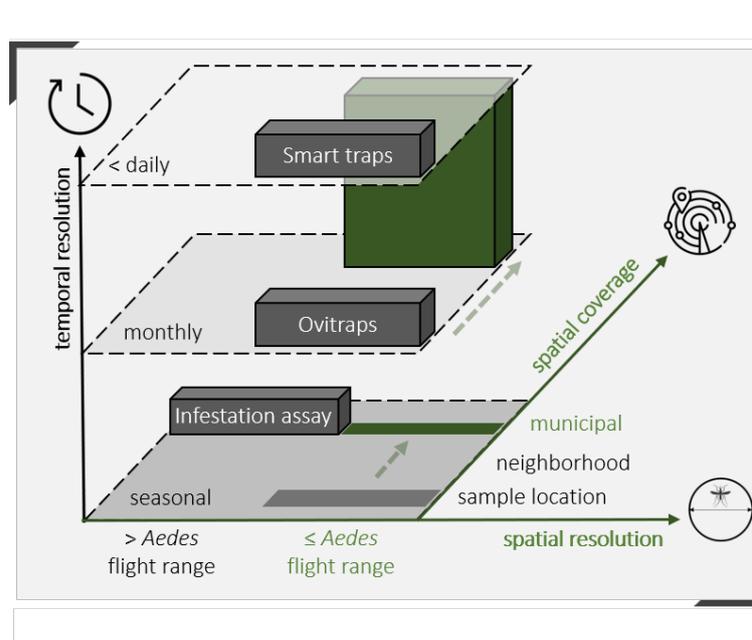
#### Derivation of Research Objective I



Development and validation of a scalable method for capturing an urban landscape feature, that is significant in modelling *Ae. aegypti* egg and larval counts, as monitored by ovitraps.

**Figure 7:** Research objective I.

Entomological surveillance, which involves the monitoring and study of insect populations, constitutes a crucial component in guiding interventions for *Aedes*-borne arboviral infections. Various methods are employed in entomological surveillance, including: (i) ovitraps, which are artificial water containers designed to attract mosquitoes and used to monitor *Aedes* eggs and larval counts, (ii) infestation assay that calculate block-level and container type immature mosquito infestation indices, and (iii) smart adult mosquito traps that enable more frequent and less labor-intensive mosquito monitoring through automated visual or acoustic detections. Each of these methods entails distinct costs, benefits, and limitations (cf. Section 2.2). However, collectively, they often lack sufficient spatial resolution or coverage for precise mosquito mapping relevant for implementing interventions efficiently (cf. Figure C. 1 in publication III). In other words, the current state of the art solutions hamper the ability to capture the potential high spatial variability in mosquito abundance that can occur in heterogeneous urban environments due to the limited flight range of *Aedes* mosquitoes (Kache et al., 2022a; Li and Dong, 2022) (cf. Section 2.2).



**Figure 8:** Schematic visualization of the research gap, visualized as a green box on the space time axis, leading to research objective I. Municipal coverage is of special interest as it is usually the lowest level for decision making. Spatial coverage in high resolution ( $\leq$  *Aedes* flight range) is necessary to capture high spatial variability that can occur in heterogeneous urban landscape due to limited *Aedes* flight range.

Meanwhile, novel, highly effective disease control programs, such as the release of *Wolbachia*-infected mosquitoes or Spinosad containing dissemination stations (cf. Section 2.2), are emerging. These programs increase the need for more precise intervention guidance due to their high associated costs and sometimes associated environmental pollution. To address this current limitation in local intervention guidance for *Aedes*-borne arboviral infection, a complementary approach integrating conventional field measurements with high-resolution remote monitoring techniques is needed (Kache et al., 2022a; Li and Dong, 2022; Louis et al., 2014; Moraga, 2024; Sallam et al., 2017). In accordance with the WHO's

strategic pillar of action to “scale up and integrate tools and approaches for vector control” (WHO, 2017), this thesis proposes an urban landscape indicator-driven interpolation technique for entomological surveillance data. This complementary approach aims to assess the environmental suitability for hosting *Aedes* mosquito populations on a more continuous scale, than entomological surveillance alone. It enables broad coverage while maintaining detailed resolution of *Aedes* habitat size (cf. Section 2.2), which is valuable for intervention guidance.

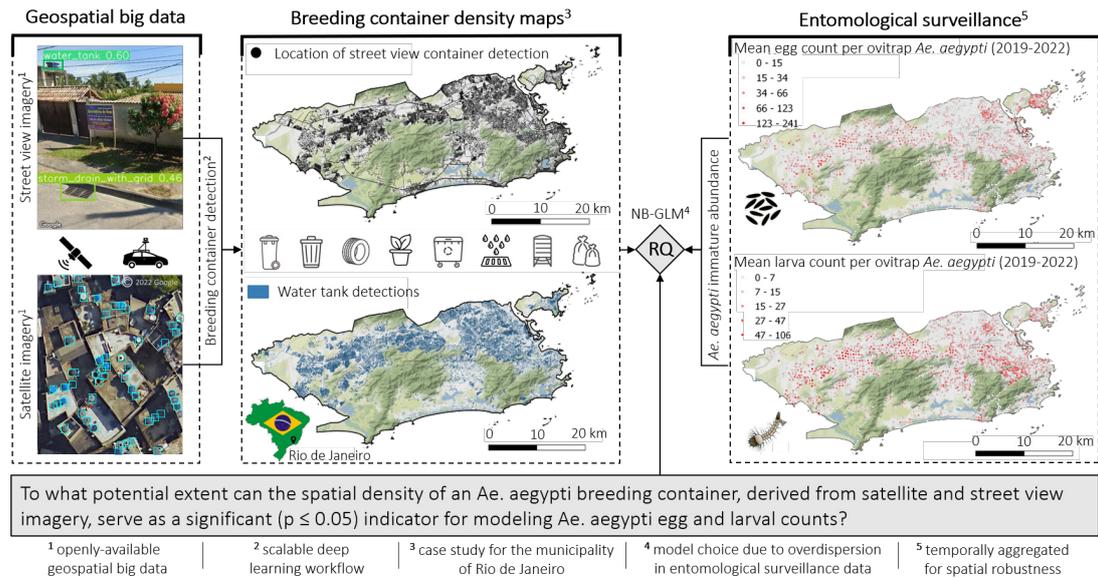
Environmental suitability, in this context, refers to a relative measure of how conducive an environment is to supporting *Aedes* mosquito populations compared to other environments. This differs from the concept of carrying capacity, which denotes an absolute measure of the maximum potential population size that an environment can sustain (Levin, 2013). Comprehensive literature reviews (Li and Dong, 2022; Louis et al., 2014; Sallam et al., 2017) indicate that scalable methods to generate significant ( $p \leq 0.05$ ) proxies for large-scale mosquito suitability modeling are a prerequisite for implementing urban landscape indicator-driven interpolations, particularly for its real-world applicability in vector control guidance. Therefore, this thesis focused on hypothesis-driven urban landscape features at a resolution finer than the *Aedes* flight range, to minimize the risk of erroneous conclusions during interpolation and to rely on causality rather than on correlations that may vary with different datasets. Considering this research gap and the broader research goal of harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections, we formulated the following hypothesis to guide our research:



*Ae. aegypti* breeding container density, detected from satellite and street view imagery, represents a significant urban landscape feature to model *Ae. aegypti* egg and larval counts, as monitored by ovitraps.

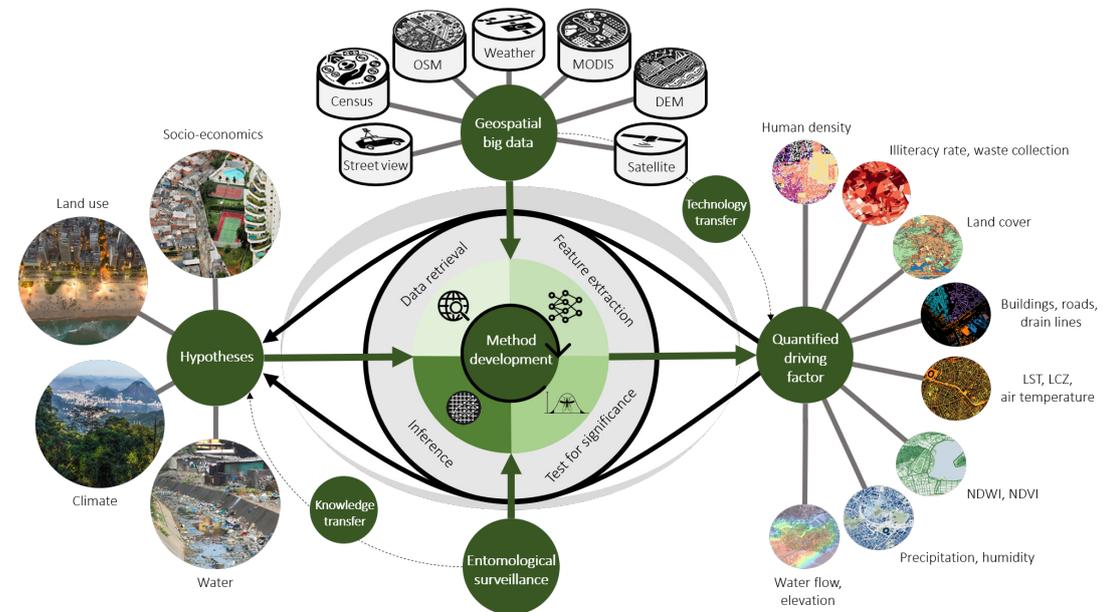
**Figure 9:** Hypothesis I.

This specific hypothesis (cf. Figure 9) was chosen because satellite and street view imagery are becoming increasingly openly available at higher resolutions and broader scales (cf. Section 2.1). Combined with modern deep learning frameworks for computer vision tasks, these geospatial data sources provide a promising means for capturing urban landscapes in a scalable manner, applicable across various research domains and case study locations. Initial studies have explored the potential of these openly available geospatial data sources for detecting *Aedes* breeding sites (Haddawy et al., 2019; Su Yin et al., 2021). However, documentation on workflows for large-scale applications ( $\geq$  municipality, representing the primary intervention level in many countries, including Brazil (Vinh Frutuoso and Barbosa Duraes, 2023)) remains scarce, hampering real-world applications of proof-of-concept findings and method transfer to various locations at risk of *Aedes*-borne arboviral infections. To test the aforementioned hypothesis (cf. Figure 9), we designed the following research concept to achieve the outlined research objective (cf. Figure 7):



**Figure 10:** The research concept developed to test hypothesis I in order to achieve research objective I, incorporating the formulated research question (RQ).

In a subsequent step, we expanded our set of breeding container density indicators by incorporating additional digital indicators related to socio-economics, urban morphology, climate, and water accumulation (cf. Figure 11). This aimed to refine the environmental suitability models for *Ae. aegypti* beyond what could be achieved with the indicators of breeding container density alone. Although this expansion is not directly aligned with Research Objective and Hypothesis I, it addresses the second part of our overarching research objective: “Harnessing geospatial data to guide local interventions for *Aedes*-borne arboviral infections”.



**Figure 11:** Conceptual idea of retrieving hypothesis-driven *Ae. aegypti* environmental suitability indicators from open GBD to infer entomological surveillance.

The additional indicators were selected based on their documented causal relationships with *Aedes* suitability (cf. Table A. 1 in publication III) to improve

the accuracy of ovitrap count modeling. By prioritizing causality over correlation, which can fluctuate with different entomological data collection periods, we aimed to reduce the risk of overfitting and better estimate environmental suitability in areas without direct ovitrap data by drawing on similarities with areas where data were available. This hybrid approach, combining entomological surveillance with a comprehensive set of digital landscape indicators representing environments conducive to *Ae. aegypti* breeding, aimed to produce a continuous map of environmental suitability at the mosquito habitat scale, extending beyond ovitrap locations. Such mapping can potentially inform vector control strategies by identifying priority regions for cost-benefit analysis, which are typically conducted before the deployment of interventions (cf. Section 2.2). Detailed methodologies and results are provided in Section 3.2.1.

## Derivation of Research Objective II



Development and validation of a method for capturing daytime *Aedes*-human interactions to enhance (i) spatial estimates on dengue occurrence and (ii) intervention guidance for *Aedes*-borne arboviral infections.

**Figure 12:** Research objective II.

While mosquito abundance, frequently monitored through entomological surveillance, serves as an actionable metric for guiding local interventions, it does not directly indicate the risk to humans of contracting *Aedes*-borne arboviral infections. The spatiotemporal transmission of these infections is influenced by several additional factors, including (i) pathogen prevalence within the *Aedes* mosquito population, (ii) levels of human immunization, and (iii) daytime variations in *Aedes*-human interactions (cf. Section 2.2). The complex interplay of these epidemiological factors is a key reason why the risk to humans of contracting *Aedes*-borne arboviral infections has been the subject of extensive research for several decades (Guzman and Harris, 2015; Murray et al., 2013).

In the context of the overarching research objective of “harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections”, investigating daytime variation in *Aedes*-human interactions presents the most promising research pathway among the aforementioned factors. While data on pathogen prevalence within mosquito populations and human immunization levels is limited, there is an increasing amount of GBD on human trajectories collected via mobile devices, enabling the modeling of daytime human movement. Numerous research studies have harnessed human mobility data from mobile devices to analyze and elucidate the significance of human movement in modeling the spatiotemporal dynamics of *Aedes*-borne arboviral infections across various scales (Chen et al., 2022b; Kraemer et al., 2018; Ramadona et al., 2019; Stoddard et al., 2009). The primary outcome of these studies indicates that inferring and incorporating human movement patterns is crucial for accurately predicting disease outcomes. This relevance arises from the limited flight range of *Aedes* mosquitoes, which confines

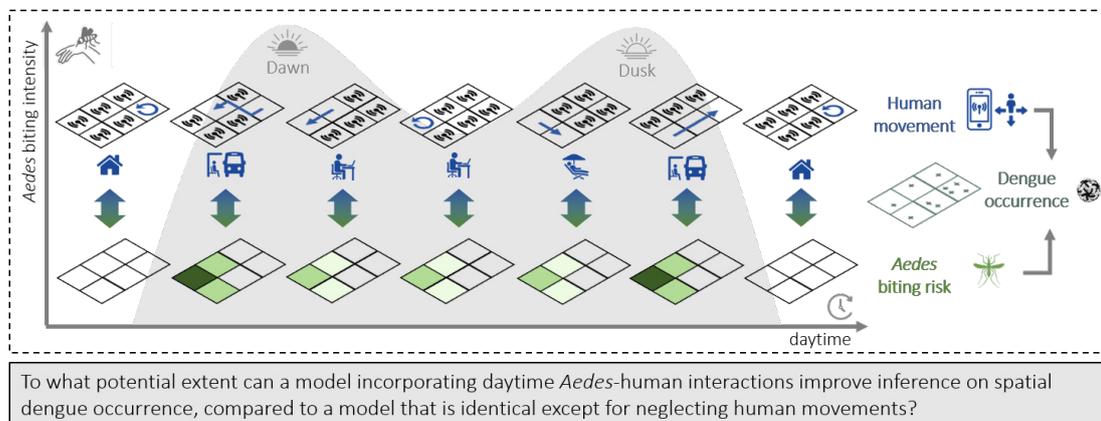
them to localized habitat areas. Conversely, humans, through their daily activities, traverse multiple mosquito habitats, facilitating the dissemination of the pathogen across different spatial regions and mosquito populations (Iggidr et al., 2017). Only a few studies (Kraemer et al., 2018) have considered the preference of *Aedes* mosquitoes for biting during daylight hours, particularly in the early morning and late afternoon (cf. Section 2.2). Studies that do not account for this biological behavior may be limited in their ability to accurately predict the risk of infection and effectively guide interventions (Kraemer et al., 2018). Accordingly, the following hypothesis was formulated to guide the research in this thesis:



Human movement patterns, derived from mobile phone records, can enhance estimates on sub-neighborhood dengue occurrence by modelling daytime *Aedes*-human interactions.

**Figure 13:** Hypothesis II.

Mobile phone records, representing temporally ordered antenna connections made during phone calls, mobile internet usage, or text messaging, have been applied in various research domains to infer human trajectories and their influence on different phenomena (Althouse et al., 2015; Finger et al., 2016; Kogan et al., 2021; Kraemer et al., 2016; Lenormand et al., 2014; Panigutti et al., 2017; Sattenspiel, Lisa and Lloyd, Alun, 2009). Due to the limited open availability of such anonymized GBD, which is typically collected by private companies, openly-available social media posts are often used as an alternative proxy for human movement patterns, with human trajectories reconstructed from sequential geo-tagged posts. While the relevance of such datasets and their inference on human movement patterns have been shown to be significant in modeling *Aedes*-borne arboviral infection risk, integrating these data streams into intervention planning maps, particularly at the municipal scale, has not been adequately documented. This gap hampers the real-world application of reported scientific insights. As this thesis aims to “harness GBD to guide local interventions for *Aedes*-borne arboviral infections”, the following research concept was developed to address this gap and test the aforementioned hypothesis (cf. Figure 13):

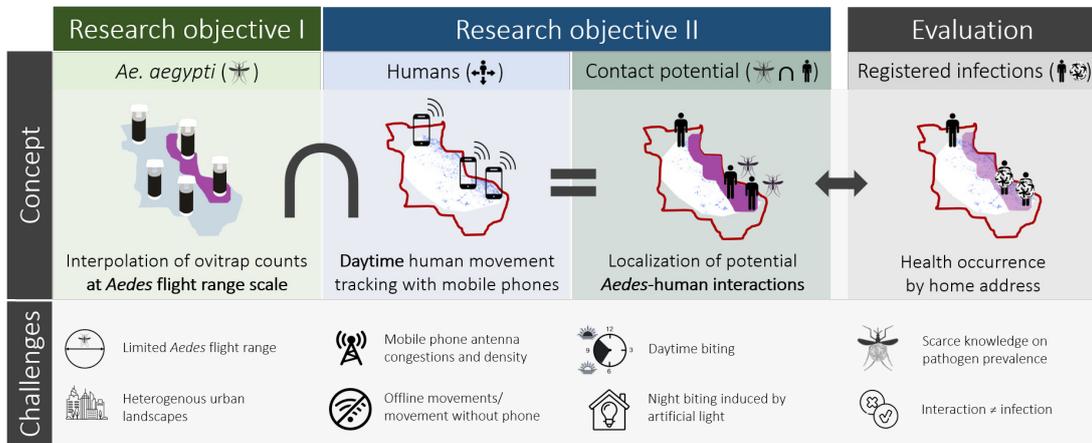


**Figure 14:** The research concept developed to test hypothesis II in order to achieve research objective II, incorporating the formulated research question (RQ).

To enhance the applicability of our research findings and address the second part of our overarching research objective, namely “the guidance of local interventions for *Aedes*-borne arboviral infections”, we incorporated estimates of likely transmission areas, considering *Aedes*-human interactions, into the environmental suitability map generated for local intervention planning. This integration aims to inform decision-makers to focus not only on high “hazard” zones, indicated by high environmental suitability for *Aedes* mosquitoes, but also on high “risk” areas of transmission, indicated by high connectivity to areas with high infection rates weighted by *Aedes* daytime biting. These high “risk” areas may be underestimated when relying solely on entomological surveillance data for planning local interventions. The methodologies and results addressing research objective II are detailed in Section 3.2.2.

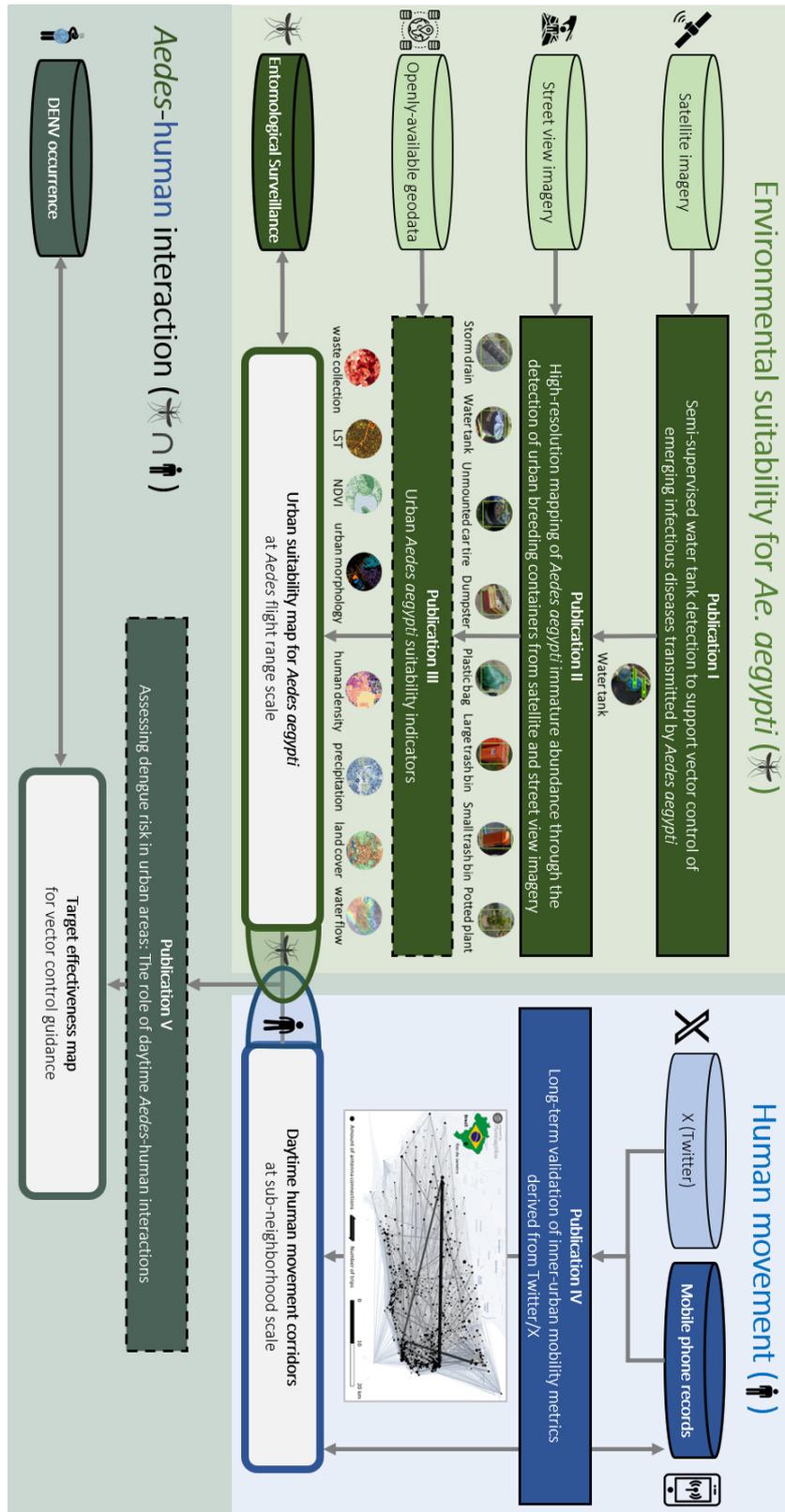
### 3.2 Research Framework and Core Methods

The relationship between research objectives I and II, along with the proposed concept and key challenges addressed in this study, is schematically illustrated in Figure 15. Given the fine spatiotemporal scale of *Aedes*-borne arboviral infections, where transmission risk can vary over short distances and time due to the limited flight range and daytime biting behavior of *Aedes* mosquitoes, this research seeks to enhance local intervention guidance by refining (i) *Ae. aegypti* environmental suitability mapping to the scale of *Aedes* habitats and (ii) the analysis of *Aedes*-human interactions considering peak *Aedes* activity hours. Addressing these ecological complexities introduces a novel perspective to the targeted research field. Beyond harnessing GBD to model these complex interactions and improve dengue occurrence estimates, this study also aims to demonstrate the critical importance of integrating ecological knowledge into local intervention strategies, proposing practical solutions for informed decision-making.

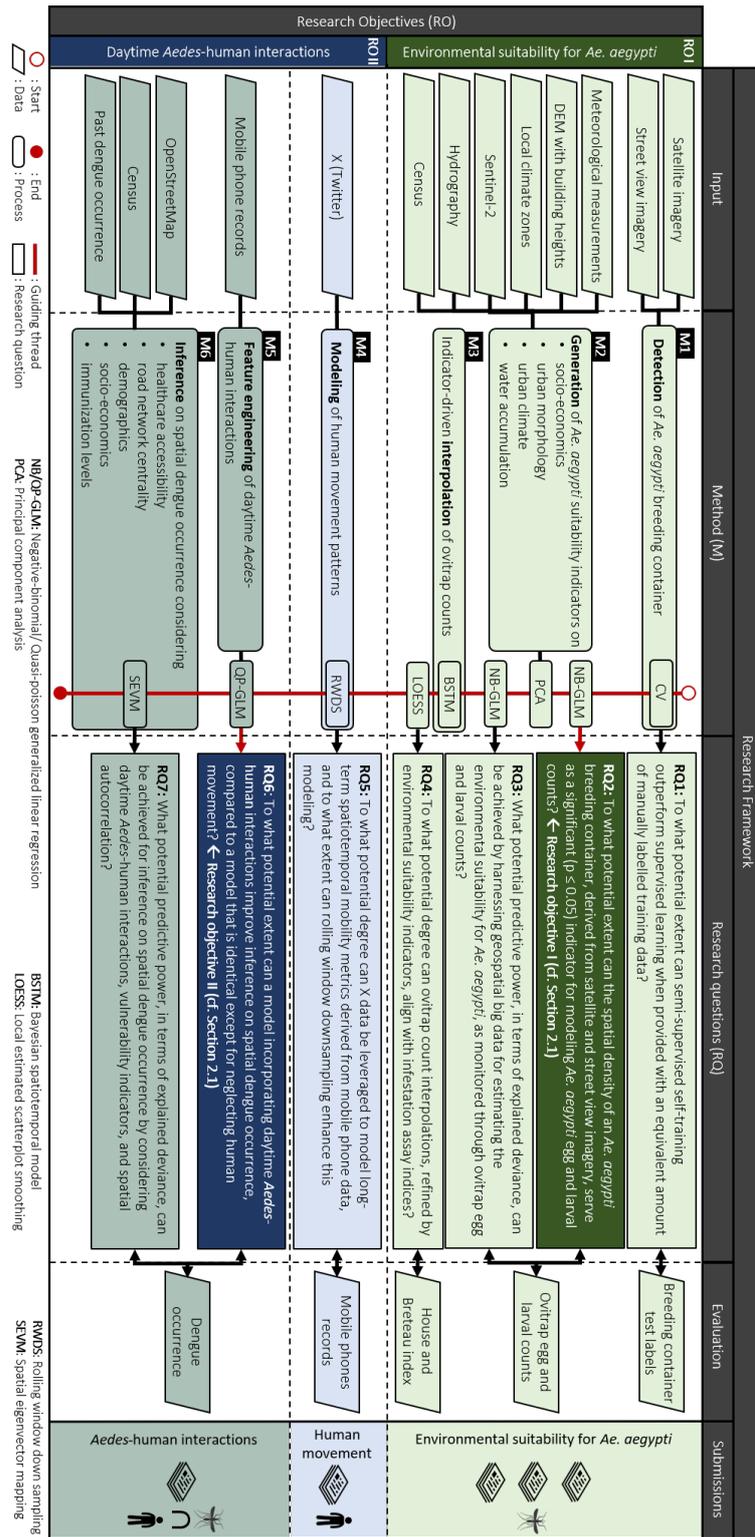


**Figure 15:** High-level schematic overview of the core research concept including key challenges.

This research was organized into five work packages (cf. Figure 16), each corresponding to a scientific manuscript submission. These submissions can be grouped into three main topics: (i) environmental suitability for *Ae. aegypti*, (ii) human movement, and (iii) *Aedes*-human interactions. Figure 17 depicts a more detailed view on the research framework, which was developed to address the two main research objectives and questions. It also incorporates five additional sub-research questions that emerged during the course of this study, focusing on methodological concepts, predictive capabilities, and alternative evaluation techniques. Key findings related to each of these seven research questions are presented in Section 3.3. A summary of applied methods, tools, and data sources is provided in Figure 19.



**Figure 16:** Integrated network of scientific manuscript submissions for the research objective of harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections. Each of the three subtopics included steps for data collection, modeling, and validation, resulting in key outputs potentially relevant for guiding local interventions. Evaluation datasets (indicated by darker colors) were retrieved via ethical approval from the municipal health ministry of Rio de Janeiro to ensure compliance with ethical standards and participant confidentiality. The methodological focus of this cumulative dissertation lies in the generation of urban suitability indicators for *Ae. aegypti* to address the challenge outlined in Figure 1.



**Figure 17:** Research framework, consisting of six sequential core methods: M1) detection of *Ae. aegypti* breeding container, M2) generation of additional indicators for environmental suitability of *Ae. aegypti*, M3), indicator-driven interpolation of ovitrap counts, M4) modeling of human movement patterns, M5) feature engineering of daytime *Aedes*-human interactions, and M6) inference on spatial dengue occurrence considering vulnerability indicators and spatial autocorrelation, all conducted at the scale of the municipality of Rio de Janeiro to enhance applicability for local intervention guidance. The red line indicates the guiding thread through the analysis, illustrating how research question 6 and 7 build upon the preceding ones. The research questions highlighted in a darker color align with those motivated in section 3.1.

### 3.2.1 Environmental Suitability for *Ae. aegypti* (RO I)

Part one of the research framework focuses on the *Ae. aegypti* suitability mapping. This part consists of two sub-methods: (i) the generation of *Ae. aegypti* suitability indicators, and (ii) the indicator-driven interpolation of ovitrap counts. For the evaluation of the results, we acquired two types of entomological surveillance data: egg and larval counts from ovitraps, as well as the BI and HI from the Larval Infestation Rapid Assay *Aedes* (LIRAA) (Ministério da Saúde Brazil, 2013a), both collected by the municipal health department of Rio de Janeiro (cf. Figure C. 1 in publication III).

Ovitrap data allowed for an examination of urban suitability indicators within the vicinity of ovitrap locations, acknowledging that inferences on mosquito abundance for locations beyond the *Ae. aegypti* flight range may be invalid. Ovitrap data was gathered at 2 698 ovitrap locations on a monthly basis, from January to December 2019. LIRAA data provided a broader assessment to evaluate continuous immature *Ae. aegypti* suitability maps generated from ovitrap data and suitability indicators together. Correspondingly, the HI and BI were periodically collected in 250 predefined city strata, representing homogeneous urban characteristics, during February 3-9, 2019; May 5-11, 2019; August 4-10, 2019; and October 13-19, 2019 (Secretario Municipal de Saúde Rio de Janeiro, 2024). The HI gauged the number of larva-infested houses relative to the total number of visited buildings during the survey, while the BI represented the number of positive containers per 100 houses inspected. The municipal health ministry of Rio de Janeiro categorized the house index into three risk classes:  $HI < 0.9$  (minor),  $0.9 \leq HI \leq 3.9$  (significant), and  $HI > 3.9$  (severe). According to the study design, one LIRAA stratum consisted of 8 000 to 12 000 properties, of which 20% were inspected following a structured schema. Field agents assessed the number of eggs and larvae in all water containers present in each household surveyed. The most prominently affected container types, ranging from water tanks and ground-level deposits to furniture such as plates and vases, as well as the class of fixed deposits, tires, garbage, and natural plants such as bromeliads, were also recorded. For each container with mosquito larvae, the agents collected a sample that was sent to the Vector Laboratory of the Agency for the Control of Endemic Diseases for larval identification. Further details regarding the placement of ovitraps and the entomological surveillance in households are currently lacking in our knowledge, but could potentially be provided by the municipal health ministry of Rio de Janeiro upon request.

It is crucial to note that all entomological surveillance data was gathered manually, introducing potential biases due to human error, observer variability, and limitations in sampling frequency and coverage. These biases may lead to inaccuracies in estimating spatial suitability for *Ae. aegypti*, as well as errors in temporal trends, impacting the reliability of our analysis. Additionally, micro-scale factors, such as the positioning of an ovitrap in shaded or unshaded areas, can impact the observation values. The positioning of ovitraps was done in a systematic manner, more or less uniformly across the built-up areas of the municipality of Rio de Janeiro. This positioning resulted in an average distance between the two closest ovitraps of 330.38m.

## Generation of *Ae. aegypti* Suitability Indicators (M1, M2)

As motivated in previous sections, we hypothesized that the generation of *Ae. aegypti* suitability indicators can enable the interpolation of ovitrap data into continuous spatial resolution, while considering the limited flight range of *Aedes* mosquitoes and capturing the potential high spatial variability in *Ae. aegypti* abundance that can occur in heterogeneous urban landscapes. This, in turn, would allow for more targeted guidance of local interventions against *Aedes*-borne arboviral infections. This section elaborates on the generation of *Ae. aegypti* suitability indicators.

Urban suitability indicators for immature *Ae. aegypti* were selected based on availability and a priori expectation of factors influencing immature *Ae. aegypti* abundance. Spatial as well as spatiotemporal covariates with differing resolutions were retrieved to interpolate entomological surveillance data considering the limited mosquito flight range. The generation of *Ae. aegypti* suitability indicators was divided into micro-habitat and macro-habitat indicators. The generation of micro-habitat indicators referred to the detection of typical *Ae. aegypti* breeding sites from satellite and street view imagery. Detected breeding sites included artificial water containers such as water tanks, potted plants, trash bins, unmounted car tires, or dumpsters often found in close vicinity to human settlements. All these containers can harbor stagnant water after rainfall, which is highly suitable for *Ae. aegypti* oviposition and subsequent adult population development. We hypothesized that their spatial distribution and occurrence, in the form of container density, could serve as a reliable indicator for the abundance of immature *Ae. aegypti* in urban environments. In addition to micro-habitat urban suitability indicators, we hypothesized a range of additional macro-habitat urban suitability indicators for immature *Ae. aegypti*, mostly collected at a coarser spatial scale. These indicators encompass a broad range of spatiotemporal proxies describing urban landscape in terms of demography, socio-economy, land use, climate, weather, green spaces, and water availability. The corresponding hypotheses were derived from previous literature (cf. Appendix A. 1 in publication III). An extended description of the applied methods and the hypothesized spatiotemporal influences of suitability indicators for the abundance of immature *Ae. aegypti* is provided in the following paragraphs.

The generation of water tank counts as a micro-habitat indicator derived from satellite imagery was extensively described in publication I (Knoblauch et al., 2023). Here, we conceptualized a semi-supervised self-training (SSST) algorithm to minimize the manual labeling effort for automated water tank detection in urban areas based on satellite imagery. We used a Single-Stage Object Detection network consisting of Inception-ResNet-V2 as a feature extractor and a multi-layer detector with a Non-Maximum Suppression layer pretrained on the Microsoft COCO dataset (Lin et al., 2014). We fine-tuned this model using 4,000 manually labeled water tanks along with 10,400 pseudo water tank labels, encompassing various urban structure types, generated by the model during the training process. In our case, pseudo labels represented the results of model inference at 20,000 training iterations, applying a confidence threshold of 0.8. In total, the neural network was trained for 40,000 iterations: 20,000 initial iterations using manual labels only and 20,000 subsequent iterations using both manual and pseudo labels, which refers to a SSST procedure. We evaluated

model performance using precision, recall, and F1-score, with particular attention to generalization across different urban structures. The model with the best F1-score was then used for large-scale water tank prediction across the entire metropolitan area of Rio de Janeiro, processing over 10 million satellite image patches and storing predictions in a PostGIS database for further analysis. This process resulted in the *Ae. aegypti* suitability indicators of “water tank density” covering the whole municipality of Rio de Janeiro (cf. Figure 8 in publication I).

In addition to the city-wide water tank detection from satellite imagery, we derived the density of trash cans, catch basins, manholes and water valves from the Mapillary API (Mapillary, 2023a) For the mapping of further common *Ae. aegypti* breeding sites from street view imagery such as potted plants, small and large trash bins, dumpsters, storm drains, unmounted car tires, and plastic bags, we fine-tuned a YOLOv5 model using google street view (GSV) imagery. The applied methods were extensively described in publication II (Knoblauch et al., 2024). Here, we fine-tuned a multi-class object detector to map *Ae. aegypti*-specific habitats as an extension of prior research of publication I. To detect *Ae. aegypti* breeding containers, we used street view images retrieved from Google’s Street View Static API (Google LLC, 2023). A 50 m downloading interval for 360-degree street view images calculated from the OSM road network was deemed appropriate for the detection of mosquito breeding sites, following the approach used in other studies (Haddawy et al., 2019; Su Yin et al., 2021). As of August 8th, 2023, this method yielded a total of 467,605 available street view images, which were utilized for labeling and city-wide container detection. The timestamps of the retrieved images ranged from January 2010 until 2023, with a share of 51% for images taken between 2022 and 2023, 15% from 2021, 19% from 2020 and 15% from before 2020. The downloaded image resolution was 600x500 pixels. For the supervised training of our multi-class object detector we manually labeled 7,578 breeding containers on 3,979 images using the graphical image annotation tool ‘labelImg’ (TuzuTa Lin, 2023). To minimize the manual labeling effort we implemented additional data augmentation techniques for instances of the ‘dumpster’ container class, which were observed infrequently within our dataset. We applied PCA color augmentation, horizontal flip and 180 degree rotation. The labeled dataset was then randomly divided into 80% for training, 10% for validation, and 10% for testing, resulting in 3,152, 454, and 373 image subsets, respectively (cf. Table 1 in publication II).

The indicators of human population density and building density were selected as macro-habitat proxies of human influence on the *Ae. aegypti* population, considering that human activities provide artificial water containers suitable as breeding habitats (Espinosa et al., 2016; Kamgang et al., 2010; Lindsay et al., 2017; Tedjou et al., 2019; Wilson-Bahun et al., 2020). The indicators slope and water flow accumulation, defined by the Horton-Strahler number, were selected in consideration of their influence on water accumulation (Cornel et al., 2016). The water flow indicator was generated by applying a D8 approximation algorithm to 5 m elevation data provided upon request by the Urban Data Platform from PPGAU UFF (PPGAU UFF, 2023a). The indicator elevation level, including building heights, was additionally added as a covariate to account for *Ae. aegypti*’s sensitivity to altitude (Equihua et al., 2017; Liew and Curtis, 2004; Lozano-Fuentes et al., 2012; Moreno-Madriñán et al., 2014; Roslan et al., 2022;

Roslan et al., 2013).

The indicator local climate zones was selected to consider *Ae. aegypti* climate-sensitive reproduction and fertility rate (Azevedo et al., 2018; Jesús Crespo and Rogers, 2021). This indicator - based on urban climate estimates by Demuzere et al. (2021) - considers ten different urban built-up types (compact highrise, - midrise, - lowrise, open highrise, - midrise, - lowrise, lightweight lowrise, large lowrise, sparsely built, heavy industry) influencing shadow and heat accumulation together with seven land cover classes (dense, trees, scattered trees, bush and scrub, back rock or paved, bare soil or sand, water) in 30 m resolution. The monthly indicators of air temperature (Chang et al., 2007; Lambrechts et al., 2011; Misslin et al., 2018; Tsuda and Takagi, 2001), precipitation (Barrera et al., 2011; Li et al., 1985; Souza et al., 2010; Stewart Ibarra et al., 2013; Valdez et al., 2018; Vasconcelos et al., 2022; Vasconcelos et al., 2021), and relative humidity (Costa et al., 2010; Lega et al., 2017; Nasir et al., 2017; Reiskind and Lounibos, 2009) were derived from the Alerta system (Centro de Operacoes Rio, 2023). Therefore, we interpolated 15-minute interval measurements from 33 weather stations during the study period of 2019 to consider both climate and meteorological effects on *Ae. aegypti* populations. Additional urban heat island effects (Araujo et al., 2015; Oliveira Lemos et al., 2021; Wilk-da-Silva et al., 2018) were retrieved by Lucena et al. (2015) and Peres et al. (2018) and Miranda et al. (2022) including cloud masking techniques, atmospheric correction and surface emissivity.

As a further proxy to describe the habitat suitability of immature *Ae. aegypti* in urban landscape, we calculated the road network density from OSM to consider the barrier effects of roads on mosquito populations (Kaplan et al., 2010; Regilme et al., 2021). The distance to coastal water bodies was also generated utilizing OSM to account for additional wind exposure effects with a negative influence on mosquito activity (Wong and Jim, 2017). The distance and coverage of urban drain lines were derived from a hydrographic data set (Data.Rio, 2023b) as an additional urban-specific proxy for immature *Ae. aegypti* populations. Normalized difference vegetation index (NDVI) (Britos Molinas et al., 2022; Chaves et al., 2021; Estallo et al., 2018; Estallo et al., 2008; Martín et al., 2022a; Martín et al., 2022b) and normalized difference water index (NDWI) (Britos Molinas et al., 2022; Estallo et al., 2018; Estallo et al., 2012; German et al., 2018) were computed using Sentinel-2 satellite imagery from the European Space Agency to consider vegetation types and water availability influencing *Ae. aegypti* especially in non-built up areas. An algorithm for cloud masking was applied to calculate the mean of cloud-free pixels at a 30 m resolution from January 2019 until December 2019 using the Google Earth Engine. After band calculations, a threshold of  $\geq 0.2$  for the NDVI and  $\geq 0.3$  for the NDWI was applied to avoid false assumptions. On top of this, land cover maps were extracted from DataRioPortal (Data.Rio, 2023a) to incorporate land use classes (Albrieu-Llinás et al., 2018; Benitez et al., 2020; Egid et al., 2022a; Landau and van Leeuwen, 2012; Lorenz et al., 2020a; Montagner et al., 2018; Vanwambeke et al., 2007; Westby et al., 2021; Young et al., 2017; Zahouli et al., 2017) such as the location of favelas and to calculate the minimum distance from ovitraps to forest areas to consider forest-specific climate effects such as locally increased humidity (Costa et al., 2010; Rowley and Graham, 1968).

An urban morphological clustering was computed using the *momepy* python library (Fleischmann, 2019) and official building footprints provided upon request by the Urban Data Platform from PPGAU UFF (PPGAU UFF, 2023b). Most recent census statistics for 10,233 strata such as the amount of collected rubbish Bonnet et al., 2020; Chumsri et al., 2020; Manrique-Saide et al., 2008; Maquart et al., 2022; Stewart Ibarra et al., 2014; Whelan et al., 2020, statistics about waste water management (Burke et al., 2010; Chan et al., 1971; Martini et al., 2019; Novaes et al., 2022), sanitation (Gomes et al., 2023), and education level (Menchaca-Armenta et al., 2018; Stefopoulou et al., 2018) as well as socio-economic indices (Liu-Helmersson et al., 2019; Lorenz et al., 2020b; Moreno-Madriñán et al., 2014; Nagao et al., 2003; Vannavong et al., 2017) were obtained from the IBGE (Instituto Brasileiro de Geografia e Estatística), IPP (Instituto Pereira Passos, Prefeitura do Rio de Janeiro), IPEA (Instituto de Pesquisa Econômica Aplicada), and the DataRioPortal.

To quantitatively evaluate how well *Ae. aegypti* suitability indicators can capture the inner-urban distribution of immature *Ae. aegypti* abundance measured by entomological ovitrap data, we ran negative-binomial generalized linear regression models (NB-GLM) with a log-link functions (Hilbe, 2012). We ran the models for several estimated mosquito flight range scenarios to further test the robustness of our results against the assumed *Aedes* habitat size. NB-GLMs (cf. Equation 1 in publication II) were selected as they allow the model to account for the overdispersion present in the applied entomological count data (cf. Figure C. 1 in publication III).

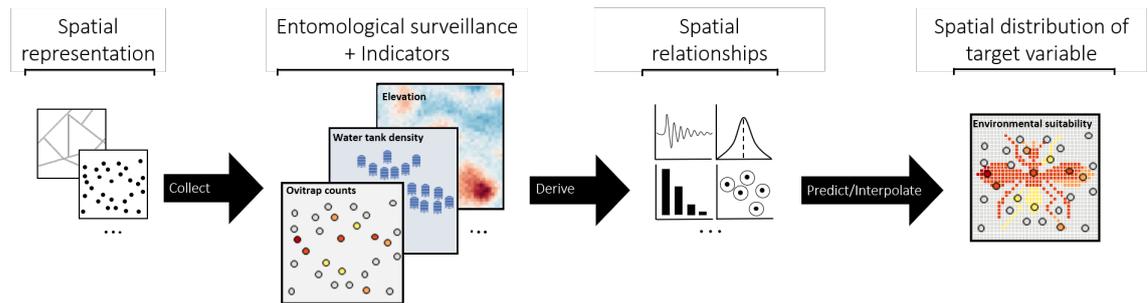
As our response variable  $Y_i$  we selected the seasonal mean eggs per trap (S-MET) and seasonal mean larvae per trap (S-MLT) rates for the year 2019. As explanatory variables, we used all main effects of our self-generated urban indicators for immature *Ae. aegypti* suitability as listed in Table A. 1 in publication III. All collected urban indicators were clipped to the geographical extent of the municipality of Rio de Janeiro. Using this stack of proxies, zonal statistics were run in different square buffers of flight range size around ovitrap locations to create five different feature vectors. Based on literature findings, we assumed that the flight range of *Ae. aegypti* is consistently lower than 1 000 m for the municipality of Rio de Janeiro (Honório et al., 2003; Moore and Brown, 2022). Therefore, we defined flight ranges of 50, 100, 200, 400 and 1 000 m as our flight range scenarios. The aerial coverage of ovitrap buffer regions in built-up areas showed notable variation: from 1.05% for a 50 m flight range buffer to 15.41% for a 200 m buffer, and up to 92.76% for a 1 000 m buffer. The lower percentages for smaller buffers, indicative of assumptions regarding lower mosquito flight ranges, underscore the difficulty in capturing the high spatial variability of urban immature *Ae. aegypti* abundance with sample-based entomological surveillance. To remove collinearity within our feature vectors we ran a PCA. All resulting 79 principal components (PC) were utilized to run the NB-GLMs for each feature vector. The combination of five different flight range buffers and the two response variables led to 10 different models, which were evaluated using Cohen's pseudo R-square (cf. Equation 2 in publication II).

### Indicator-driven Interpolation of Ovitrap Counts (M3)

The second step in *Ae. aegypti* suitability mapping involved the indicator-driven interpolation of ovitrap counts (cf. Figure 18). Here, we modeled ovitrap egg and larval counts, denoted as the response variable  $Y_{it}$ , observed at distinct spatial locations  $i = 1, \dots, 2698$  and time periods  $t = 1, 2, 3, 4$  using a Bayesian spatiotemporal model with a negative-binomial probability distribution. Given the observed overdispersion in the applied entomological count data (cf. Figure C. 1 in publication III), we employed a chi-squared test to evaluate the adequacy of the negative-binomial distribution as a likelihood function. The model assumes,

$$\begin{aligned}
 Y_{it} &\sim NB(\hat{\mu}_{it}, \hat{\theta}) \\
 \mathbb{E}(Y_{it}) &= \hat{\mu}_{it} * (1 - \hat{\theta}) / \hat{\theta} \\
 \text{Var}(Y_{it}) &= \hat{\mu}_{it} * (1 - \hat{\theta}) / \hat{\theta}^2 \\
 \log(\hat{\mu}_{it}) &= \hat{\beta}_0 + \sum_{j=1}^{79} \hat{\beta}_j * PC_{itj} + \xi(\mathbf{x}_i, t) \\
 \xi(\mathbf{x}_i, t) &= a\xi(\mathbf{x}_i, t - 1) + w(\mathbf{x}_i, t),
 \end{aligned} \tag{1}$$

and consists of an intercept  $\hat{\beta}_0$ , PCs of spatiotemporal suitability indicators  $PC_{itj}$ , and independent and identically distributed spatiotemporal random effects  $\xi(\mathbf{x}_i, t)$  that change in time with first order autoregressive dynamics ( $|a| < 1$ ) (Blangiardo and Cameletti, 2015; Lindgren et al., 2011; Zuur et al., 2017). The model's incorporation of covariates is assumed to enhance its predictive capability and facilitates a more holistic understanding of the actors influencing immature *Ae. aegypti* suitability. The spatial model component was modeled by Integrated nested Laplace approximation (INLA) using the Euclidean distances between ovitrap locations, a Matérn covariance function, and stochastic partial differential equations (SPDEs). Gaussian Markov random fields were built on triangle meshes considering boundary effects that could artificially inflate variance near the edges of the study area (cf. Figure E. 1 in publication III).



**Figure 18:** Schematic visualization of the process of indicator-driven interpolation of ovitrap egg and larval counts.

To evaluate indicator-driven interpolations of ovitrap counts we considered LIRAA indices as an evaluation set. Therefore, we organized monthly ovitrap data into quarters that corresponded to the four LIRAA seasons. This grouping was performed using the feature vector of the best-performing flight range buffer, identified through the NB-GLMs (cf. Section 3.2.1). Subsequently, we fitted a Bayesian spatiotemporal model with INLA (Rue and Lindgren, 2024), to generate

seasonal and spatially continuous urban suitability maps for immature *Ae. aegypti* covering the whole municipality of Rio de Janeiro. To map immature *Ae. aegypti* suitability in continuous space, the inverse distance weighting (IDW) algorithm was applied to interpolate point estimates from the mesh nodes to a uniformly distributed raster of 100 000 cells for each season of 2019, visualized using QGIS (QGIS Association, 2024). Here, point estimates represented the seasonal posterior means of the spatial random effects. Scatterplots were utilized to compare the predicted suitability values with the observed LIRAA indices. Additionally, we calculated the Pearson's correlation coefficient and applied locally weighted scatterplot smoothing (LOESS) across all seasons. To achieve this, zonal statistics were performed on LIRAA strata for each of the four seasons (Jan-Mar; Apr-Jun; Jul-Sep; Oct-Dec) and response variables, respectively. Before calculating zonal statistics, continuous egg and larva interpolations were clipped using the urbanization area to avoid false inference and high bias, as interpolations were created using egg and larval counts from urbanized area only. The results were then compared with mean values of ovitrap counts from the field.

### 3.2.2 Daytime *Aedes*-human Interactions (RO II)

Part two of the research framework focuses on modeling daytime *Aedes*-human interactions. This part consists of two sub-methods: (i) the retrieval of daytime human movement data, and (ii) the feature engineering of daytime *Aedes*-human interactions. To evaluate the results in this subsection, we acquired clinical health records of *Aedes*-borne arboviral infections in the municipality of Rio de Janeiro. The health records contained daily counts of DENV cases from 2015 to 2022 with geographical coordinates corresponding to residential addresses (cf. Figure 3 in publication V). In adherence to ethical considerations and following the approval granted by the Research Ethics Committee (CEP) under protocol number 66143722.0.00005243, this dataset underwent anonymization and was made accessible upon formal request by the municipal health ministry of Rio de Janeiro (Ministério da Saúde Brazil, 2013b).

### Modeling of Human Movement Patterns (M4)

To retrieve daytime human movements, we used anonymized mobile phone records provided by a large Brazilian telecommunications company. The dataset included individual antenna connections from approximately three million unique users. This is equal to an approximated penetration rate of around 45 percent for the population of the city of Rio de Janeiro. The temporal resolution of the raw data was five minutes. The data was provided at the level of the antennas (cf. Figure 4 in publication IV - top right). The mobile phone user is typically connected to the closest antenna, which is used as a proxy for the position of the user at this point in time. The number of antennas in our data set varied daily between 1200 and 1250 due to technical failures of some antennas. An antenna connection from an user was recorded when sending a text message, using mobile internet data, or making a call. We retrieved and processed the data of 164250 million mobile phone records via the distributed computing tool Apache Spark as well as the GPU-accelerated parallel computing framework

Dask using the mobilkit python library (Ubaldi et al., 2021). As a first cleaning step, we dropped connections with antennas outside the city boundaries. In order to derive human movement patterns, we generated a sequence of antenna connections for each user over the whole time period using a machine with 7 TB of local scratch. To increase the informative power of successive antenna connections for inferring human movement patterns, we introduced a lower bound (LB) and upper bound (UB) as filters for the inter event time (IET) between sequential antenna connections from a single user as proposed by Zhao et al. (2019). As a result, successive antenna connections between which less than 15 minutes (LB) or more than four hours (UB) elapsed were not counted as movements (cf. Figure 4 in publication IV - top left). The introduction of a LB was justified by the fact that antenna congestion can cause the user to jump back and forth between antennas without physical moving. A UB was introduced to avoid the counting of movements that are not necessarily made in a direct way. The choice of the lower threshold was selected based on Zhao et al. (2019) and Schlosser et al. (2020). The choice of the upper threshold was inspired by Barboza et al. (2021). Hourly origin-destination (OD) matrices were created based on IET-filtered daily user sequences, spanning from July 2021 to July 2022, encompassing a complete annual cycle of human mobility patterns.

### Feature-engineering of Daytime *Aedes*-Human Interactions (M5)

To model daytime *Aedes*-human interactions  $B_i$  we combined continuous estimates on *Aedes* suitability  $M_i$  from part 1 of the research framework with daytime human movement patterns  $\chi_{i,j}(h)$ , incorporating knowledge of *Aedes* biting behavior  $\omega(h)$ :

$$B_i = \left[ \sum_{h=1}^{24} \omega(h) \left( \sum_{j=1}^N \chi_{i,j}(h) M_j \right) \right] \quad (2)$$

Equation 2 aims to incorporating two key principles. First, due to human movement, individual hosts are exposed to different mosquito populations throughout the day. To capture this for each hour of the day  $h$ , we calculated a weighted sum approximating the contribution of mosquito populations  $M_j$  from all cells  $C_j$  to the biting risk of people resident in cell  $C_i$ . This sum reflects the extent to which the hourly mosquito biting risk originating from the mosquito population  $M_j$  in cell  $C_j$  affects individuals residing in cell  $C_i$ . To this end, we estimated  $\chi_{i,j}(h)$ , representing the fraction of people present in cellular tower tessellation cell  $C_j$  during hour  $h$ , relative to the total number of residents in cellular tower tessellation cell  $C_i$ . The calculation of  $\chi_{i,j}(h)$  utilized hourly OD matrices, indicating collective human mobility from cell  $C_i$  to cell  $C_j$ .

Second, considering the daytime variation in mosquito biting behavior, we introduced the hourly weighting function, denoted as  $w(h)$  in our model (cf. Equation 3). It is well-documented that *Ae. aegypti* and *Ae. albopictus* exhibit biting behavior primarily during daylight hours, with increased activity observed during twilight (Muhammad et al., 2020; Mutebi et al., 2022b; Zahid et al., 2023). Therefore, we assumed a decrease in mosquito biting activity during midday hours. However, we hypothesized that this behavior might persist in shaded areas with high humidity and other environmental

conditions favorable to mosquito activity (Baik et al., 2020; Egid et al., 2022b; Wei et al., 2023). Notably, mosquito biting activity during the night was excluded from our proposed model, as *Aedes* mosquitoes are typically inactive during nighttime, despite some studies reporting increased biting behavior under artificial light conditions (Rund et al., 2020).

$$\omega(h) = \begin{cases} 3, & \text{if } h \in \{6, 7, 8, 9, 15, 16, 17, 18\} \\ 2, & \text{if } h \in \{5, 10, 14, 19\} \\ 1, & \text{if } h \in \{4, 11, 12, 13, 20\} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The proposed model deviates from density-dependent epidemiological models, where transmission rates are directly proportional to population density and are typically modeled by multiplying the number of susceptible hosts by the vector population. Instead, for MBDs, a frequency-dependent model is more appropriate, as it accounts for the saturation effect, where each mosquito can feed on a limited number of hosts within a specific period. For simplification and consistency, we therefore modeled daytime *Aedes*-human interactions by propagating the environmental suitability for *Ae. aegypti* across space, utilizing human movement patterns weighted by hourly biting activity.

The proposed feature engineering underwent evaluation employing a quasi-Poisson generalized linear model (QP-GLM), wherein the target variable  $D_i$  was defined by overdispersed official dengue case counts aggregated on 1 359 cellular tower tessellations between the years 2015 and 2022 (cf. Equation 4). For evaluation, we calculated Cohen's pseudo- $R^2$  (cf. Equation 5). The explained deviance for this regression model was compared to the pseudo- $R^2$  of a base model that did not consider assumptions related to diurnal *Aedes* mosquito biting behavior and hourly human movement (cf. Figure 2). In contrast to the proposed model, the base model was implemented utilizing identity OD matrices for  $\chi_{i,j}(h)$ .

$$\begin{aligned} D_i &\sim \text{quasi-Poisson}(\hat{\mu}_i, \hat{\theta}) \\ \mathbb{E}(D_i) &= \hat{\mu}_i \\ \text{Var}(D_i) &= \hat{\mu}_i * \hat{\theta}, \text{ with } \hat{\theta} \neq 1 \\ \log(\hat{\mu}_i) &= \log(H_i) + \hat{\beta}_0 + \hat{\beta}_1 * B_i \end{aligned} \quad (4)$$

$$\text{Cohen's pseudo } R^2 = 1 - \frac{\text{model deviance}}{\text{null model deviance}} \quad (5)$$

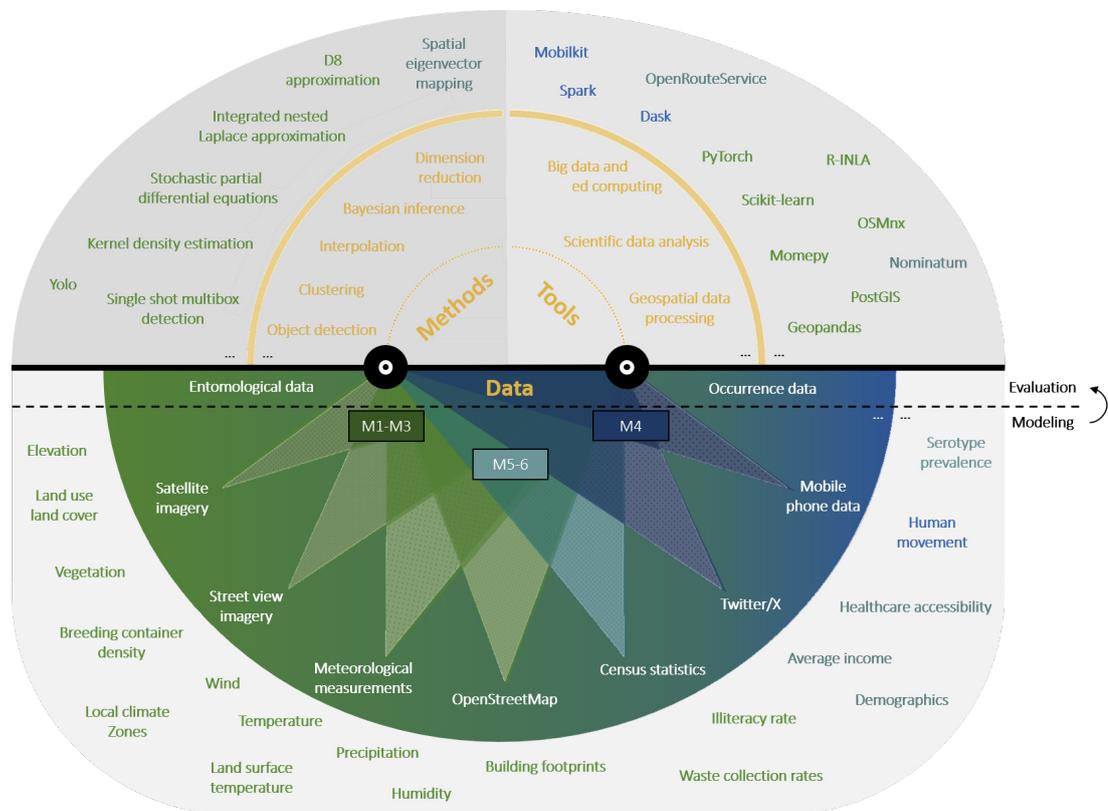
### Inference on Spatial Dengue Occurrence (M6)

After evaluating the feature engineering of daytime *Aedes*-human interactions, we expanded our QP-GLM in two aspects: (i) by incorporating vulnerability indicators to model the likelihood of an infected individual being registered in official health registries, geolocated by residency, and (ii) by integrating spatial eigenvectors to address spatial autocorrelation of residuals. These modifications were based on the hypothesis that including vulnerability indicators and spatial eigenvectors would further enhance the proposed QP-GLM, which considers *Aedes*-human interactions (cf. Equation 4) for predicting the spatial occurrence of dengue in the municipality of Rio de Janeiro. To mitigate multicollinearity among covariates, we selected vulnerability indicators with low intercorrelation ( $\leq 0.7$ ).

We additionally assessed the extent to which these additional variables improve the predictive capability and understanding of DENV transmission dynamics within the urban environment. Here, vulnerability features were defined as variables that

influence the appearance and collection process of DENV infections at the urban scale, but not the human-mosquito interactions itself. This strategic inclusion allows us to dissect the nuanced factors contributing to DENV occurrence, beyond solely focusing on the dynamics of human-mosquito interactions. In this study, these factors included the location of vulnerable age groups, accessibility to health care facilities, road network centrality, the socio-economic factor of average income, and estimates on immunization levels derived from past DENV infections. In contrast to the first model defined in Equation 3, the year 2022 was selected as the reference year for predictions, coinciding with the occurrence of the last major DENV outbreak in the municipality of Rio de Janeiro (cf. Figure B. 1). Consequently, immunization levels were estimated based on the spatial distribution of past infections recorded between 2015 and 2021.

The applied spatial eigenvector mapping, originally proposed by Griffith et al. (2019) (Griffith, 2019), involved the incorporation of additional covariates to absorb spatial autocorrelation. This ensures unbiased estimators for other predictors. These covariates, derived from the eigenfunction decomposition of the spatial weight matrix  $W$ , are called spatial eigenvectors. They represent orthogonal components that effectively separate and capture information on spatial autocorrelation, similar to PCA. In this dissertation, we employed daily aggregated OD matrices from July 2021 to July 2022 to illustrate human connectivity between antenna tessellations, serving as a spatial weight matrix (cf. Figure 5). This led to the generation of 1359 spatial eigenvectors, out of which the ME function from the 'spatialreg' R package facilitated the identification of a specific subset applying brute-force search (Bivand, 2023; Griffith, 2000) under consideration of an alpha threshold of 0.05 to mitigate residual autocorrelation. This selected subset of eigenvectors was integrated as additional covariates into the QP-GLM (cf. Equation 3).



**Figure 19:** Variety of applied methods (M), tools, and data sources in this dissertation.

### 3.3 Results and Discussion

This section presents the results and discussion related to the seven research questions, outlined in Figure 17 and grouped according to the two main research objectives. For clarity and ease of understanding, key findings are summarized in grey boxes.

#### 3.3.1 Environmental Suitability for *Ae. aegypti* (RO I)

##### Key Findings (Research Objective I)

The density of *Aedes* breeding containers, as estimated using object detection models trained on satellite and street view imagery (with F1-scores  $\geq 0.84$ ), has been identified as a significant predictor ( $p \leq 0.05$ ) for modeling *Ae. aegypti* egg and larval counts, as monitored using ovitraps in this case study. The statistical significance of this micro-habitat indicator varies depending on the type of breeding container, the selected imagery source, and the spatial scale of the *Aedes* habitat modeled around ovitrap locations. Future research should explore the potential of ultra-high-resolution satellite imagery, as the applied high-resolution satellite imagery has shown limitations in detecting smaller breeding sites, and drone imagery appears impractical for large-scale applications (cf. publication I). Street view imagery can assist in detecting smaller breeding sites but is subject to greater spatial bias than satellite or airborne imagery, as it is restricted by the road network and the availability of captured images (cf. publication II). Both satellite and street view imagery are unable to account for indoor breeding containers or those located beneath shelters or tree canopies in backyards and on rooftops. Given these inherent data limitations, the proposed concept is unable to estimate the absolute number of breeding containers, which aligns with carrying capacity - a key variable in many epidemiological models for VBDs. However, assuming uniform data biases, it can assess environmental suitability for *Ae. aegypti*, which reflects the likelihood of an area to sustain vector populations and is useful for guiding interventions. Additional limitations of the proposed concept include the temporal alignment of imagery with ovitrap records, as well as a lack of information on waste collection or breeding container removal initiatives. In this context, more frequent imagery updates, combined with citizen science initiatives such as “Mosquito Alert”, could play a valuable role in enhancing surveillance by capturing data on unmonitored factors, thereby addressing gaps in the proposed techniques. When combined with additional environmental suitability indicators for *Ae. aegypti* such as “water accumulation”, “urban morphology”, and “urban climate”, *Aedes* breeding container density can explain up to 75% of temporally aggregated ovitrap larval counts. The performance for egg counts was slightly lower, with a pseudo- $R^2$  value of 0.73 (cf. publication III). The predictive power of the density of single *Aedes* breeding container types was relatively low ( $\leq 2\%$ ) (cf. publication II). The predictive power, akin to the significance of individual indicators, exhibited a strong dependence on the timing of entomological surveillance and the spatial scale of the modeled *Aedes* habitat. The replicability of the results is influenced by data noise in entomological surveillance, including factors such as human error, observer variability, and undocumented micro-scale factors - such as the placement of ovitraps in shaded or unshaded areas - which impact the ovitrap counts used for evaluation. This variability underscores the rationale for temporally aggregating ovitrap counts to facilitate a more robust spatial analysis.

Broader applications should consider challenges related to non-uniform data availability and resolution, as well as environmental diversity that can affect the transferability of methods across diverse ecological contexts. These challenges may require adaptation of the selected hypotheses for indicator retrieval, along with models that explicitly account for identified data bias. Expanding the set of hypothesis-driven indicators could potentially further enhance predictive power, which is crucial for indicator-driven interpolation techniques of ovitrap counts, offering more precise guidance for local interventions than entomological surveillance alone (cf. publication III). Given that hypothesis-driven feature engineering has proven to be labor-intensive in this study, future research could additionally focus on developing multi-modal deep learning models that leverage diverse GBD sources across various ecological landscapes to enable more robust and scalable model inference for predicting the environmental suitability of *Ae. aegypti*.

**RQ1: To what potential extent can SSST outperform supervised learning when provided with an equivalent amount of manually labeled training data?**

The best semi-supervised self-trained water tank detection model achieved a precision score of 0.864, a recall of 0.823, and an F1 score of 0.843 on independent test datasets, outperforming the supervised base model, which yielded a precision score of 0.59, a recall of 0.85, and an F1 score of 0.69 on the same test dataset (cf. Table 2 in publication I). An increasing training time on the merged label set of manual and pseudo water tank labels continuously improved the F1-score of our object detection models (cf. Table 2 in publication I). The best water tank detection model was the model that used the additional pseudo labels for the longest SSST time (50% of the total 40,000 iteration, SSST-50) with an overall F1-score of 0.84 averaged over all test labels. This significant improvement of 22% compared to the supervised base model indicates a good balance of precision and recall. All SSST models showed a slightly decrease in recall compared to the supervised base model - i.e the proportion of correctly detected water tanks to the sum of all true water tanks decreased. However, this was countered by major improvements in precision, as the amount of correct water tank predictions on all predictions was higher for all four SSST models used.

For the best (SSST-50) model, the relative increase in F1-score (cf. Table 3 in publication I) was more obvious for urban structure types excluded in the manual label set (e.g. Commerce and Service, Education and Health, Industry) than for the urban structure types included in the manual label set (Favela, Residential). The F1-score improved, however, for all urban structure types. This makes the SSST-50 model more applicable for large-scale predictions than the supervised base model. These results were consistent with our expectations, namely that SSST models benefit from the additional knowledge collected by the machine itself, leading to more precise and robust water tank predictions across different urban structures relevant for large-scale predictions. The trained SSST-50 model is provided in the supplementary materials of this work.

However, we also identified several limitations in the results. First, not all urban structure types were used for model evaluation. We focused only on five of eleven land classes included in the LULC map where we expect human population and thus the largest risk for infections by *Ae. aegypti*. The second limitation results from the manual labeling process. We generated the test set on the basis of satellite imagery instead of a field study. Non-visible water tanks underneath shelters were thus not included in our test labels for model evaluation. However, we assume that a field study for

the labeling of water tanks would not mitigate the achieved performance improvement of the SSST approach. Much more likely, it would have an impact on the absolute performance metrics, but to the same extent on those of the supervised BM as on those of the SSST models. The third limitation of this dissertation is the low amount of manual training labels (4,000) compared to the amount of pseudo labels used for training (10,800). This implies a relatively high risk of an inappropriate training with potentially incorrect pseudo labels which can accumulate the error in the iterative self-training process. To reduce such a bias, one could either develop a debiased self-training algorithm similar to the one proposed by Chen et al. (2022a) or apply co-training of classifiers originally proposed by Blum and Mitchell (1998).

Further limitations of this dissertation become apparent when visually inspecting raw prediction images of the SSST-50 model (cf. Figure 7 in publication I). Common false negative predictions included water tanks in the shade or partial shade. To minimize the amount of these false negative predictions one could further fine-tune the SSST-50 model by feeding it with more shaded water tank labels. It is noteworthy that the number of objects in the study area which appear similar to water tanks was quite high resulting in high numbers of false positives. While similar objects such as blue cars and rooftop ventilators were rarely labeled as water tanks by our models, circular water pools or blue sunshades on beaches were frequently false positives. The false positive detection of water pools could be solved by applying a size filter. The detection of blue sunshades on beaches could be eliminated by applying an automatic land use map based filtering. However, these solution methods would only work to a limited extent. For very small water pools and blue sunshades not located on beaches this solution method becomes obsolete. Another solution would be the filtering of predictions by confidence score as applied during SSST training.

Further improvements of our models might be achieved by changing parameters of our SSST framework. This includes the size of the areas used for supervised model predictions to generate pseudo labels, the confidence threshold score applied for pseudo label filtering, the overall training time for object detection models, and the corresponding point for conducting the SSST loop. The training of a two-stage object detector like Faster-RCNNs as proposed by Cao et al. (2019) could also be an option for further investigations.

An alternative method for the reduction of manual labeling effort for object detection could be data augmentation. Data augmentation describes the technique of increasing the training set by creating slightly modified copies of provided training samples, for example by changing the rotation of the label (Shorten and Khoshgoftaar, 2019; van Dyk and Meng, 2001). It is a widely used method especially applied to avoid overfitting. However, for our use case of generating a robust model for large-scale predictions over various urban structure types, SSST seems to be more suitable. Instead of creating label copies, self-training can create completely new water tank labels (pseudo label) that can appear in different shape, color, and with varying shadow coverage. In addition, it allows to incorporate background features in the training process, like different rooftop types or water tank densities, not necessarily present in the limited manual label set used. All these additionally features gathered during pseudo label generation via self-training are extremely relevant, when trying to train robust object detector using Convolutional-Neural-Networks (CNNs). Especially for applying these models on over 10 million satellite image patches covering all types of urban structures. SSST can avoid overfitting similar to data augmentation (Nartey et al., 2020). Of course, do both methods, data augmentation and self-training, allow a cost-sensitive creation of additional labels, which is relevant for our use case to minimize the manual labeling effort and associated cost and time. However, the capability of learning

additionally background features, not present in the manual label set, is only possible through self-training in an automatic manner. Nevertheless, self-training requires a relatively high configuration effort to be successful compared to data augmentation techniques as described in the beginning of this result section.

In summary, the results for RQ1 show that SSST can outperform supervised learning by 22% in terms of performance (F1-score) and further improve generalization across diverse urban landscapes for water tank detection from satellite imagery when using an equivalent amount of manually labeled data. These findings motivate the use of SSST for the scalable generation of environmental suitability indicators for *Ae. aegypti*.

**RQ2: To what potential extent can the spatial density of an *Ae. aegypti* breeding container, derived from satellite and street view imagery, serve as a significant ( $p \leq 0.05$ ) indicator for modeling *Ae. aegypti* egg and larval counts?**

The results of the implemented NB-GLMs (cf. Table 4 in publication II) indicated that the density of all detected breeding containers, as identified through satellite and street view imagery, were highly significant ( $p \leq 0.05$ ) proxies for modeling urban *Ae. aegypti* egg and larval counts while considering limited mosquito flight range below 1,000 m. This was in line with our expectations and implies that breeding site density maps can be a useful indicator to enrich entomological surveillance data and thus support future local interventions by providing more continuous and high resolution insights for urban mosquito distributions.



**Figure 20:** Schematic visualization of AI-assisted mapping of water tanks, which serve as common *Ae. aegypti* breeding sites, using satellite and street-view imagery to model the spatial distribution of *Ae. aegypti* eggs and larvae, as monitored by ovitraps.

Water tanks, non-mounted car tires, storm drains, plastic bags, and potted plants consistently displayed positive coefficient estimates for both response variables, whereas the coefficient estimates for small and large trash bins, as well as dumpsters, consistently demonstrated negativity across both model variations. These findings aligned with the intuitive understanding that an increased presence of trash bins of any kind correlates with a reduced prevalence of uncontained refuse piles, thereby mitigating the potential for additional mosquito breeding sites. The correlation between the density of plastic bags and all three trash container classes was found to be negative, namely -0.1 for the dumpster class, -0.03 for large trash bins, and -0.3 for small trash bins. In addition, small and large trash bins, as well as dumpsters, are usually closed containers that rarely fill with water when it rains, which underlines their significance ( $p \leq 0.05$ ) and negative association with entomological data about *Ae. aegypti* immature

abundance. Furthermore, these containers are regularly emptied by refuse collection services, ensuring that they often remain dry and unsuitable for mosquito breeding, thus contributing to mosquito control efforts.

When analyzing the results independently from the response variable, it was observed that models using water tank density derived from satellite and street view imagery consistently led to the lowest Akaike information criterion (AIC), indicating a superior fit to the data across both immature abundance stages. Conversely, models employing the density of potted plants displayed the highest AIC values in relation to the mean eggs per trap (MET) rate, while models utilizing the density of large trash bins exhibited the highest AIC values in relation to the mean larvae per trap (MLT) rate. The extent of explained deviance in the regression models pertaining to the MLT rate generally exhibited higher values compared to those associated with the MET rate. Specifically, the MLT model, utilizing water tank density derived from satellite imagery, achieved the highest explained deviance at 0.05 as quantified by Cohen's pseudo- $R^2$  (Cohen, 2013) (cf. Equation 2 in publication II). This indicates that approximately 5% of the variance in the response variable is accounted for by the univariate model.

The deviance function of the NB-GLM captured the increasing variance with the mean that is typical for count data. The dispersion parameter captures how much the variance increases with the mean relative to a Poisson GLM, where the variance equals the mean. The theta values of all univariate regression models in this study indicated a substantial overdispersion. This overdispersion can be attributed to two primary factors. First, the dataset on entomological observations contained a substantial number of zero values, necessitating the adoption of a NB-GLM to account for excess variation. Second, the limited inclusion of predictors in modeling the urban distribution of *Ae. aegypti* also contributed to the observed low value of explained deviance. It is worth noting that certain potentially relevant predictors have been intentionally omitted from the model, further contributing to the constrained explanatory power. The incorporation of additional explanatory variables is planned for subsequent phases of this research.

The outcomes of the performed sensitivity analysis (cf. Table A. 2 in publication II), scrutinizing different assumed maximum flight ranges of *Ae. aegypti* (250 m, 500 m), confirmed the robustness of the results outlined in Table 4 in publication II. Similar to the results for a 1,000 m *Ae. aegypti* maximum flight range, at a maximum flight range of 500 m, all container types exhibited significant p-values ( $p \leq 0.05$ ) for both egg and larval counts. The same trend was observed for the assumed maximum *Ae. aegypti* flight range of 250 m, except for the container types dumpster, storm drain, and water tank detected from satellite imagery. Notably, the findings concerning water tanks from satellite imagery at 250 m scale show a slight contrast to our previous findings in Knoblauch et al. (2023), where a different time frame for entomological data was utilized; however, significance was detected at a flight range scale of 200 m. This divergence of these findings underscores the considerable influence of the selected time period of entomological surveillance on the validation of such results. The coefficients for small and large trash bins, as well as the dumpster category, remained negative also at lower estimated maximum *Ae. aegypti* flight ranges. Intriguingly, the coefficient for potted plants shifted from positive to negative when simulating a maximum flight range of 250 m for *Ae. aegypti*. Overall, there was an evident upward trend in significance (indicated by a downward trend in p-values) across all container classes, with larger buffer sizes, representing simulations of larger flight ranges, showing higher significance levels. Essentially, larger buffer areas augment the probability of encountering containers, consequently yielding more dependable statistical outcomes in our methodology for

modeling ovitrap count data with digital proxies. For a more nuanced understanding of the relationship between assumed maximum *Ae. aegypti* flight range and significance values, models implementing soft constraints could be considered, such as Bayesian models.

The collective findings presented here offer a comprehensive overview and extension of prior research about urban mosquito mapping (Haddawy et al., 2019; Su Yin et al., 2021). For the first time the results underscore the practical efficacy of integrating satellite and street view imagery for identifying mosquito breeding sites in urban areas, emphasizing the distinctive advantages of each method. A further alternative data source for mapping mosquito breeding containers in urban areas could be drone imagery, which offers both continuous spatial coverage and images in high resolution for small breeding container detection (Passos et al., 2022; Passos et al., 2023). However, it is essential to note that generating drone imagery incurs substantial costs and labor, thereby limiting the applicability in diverse global urban settings. A common limitation across all three data sources is their inability to detect breeding containers located inside buildings. Consequently, the digital strategies outlined in this study cannot fully replace on-site entomological surveillance. Instead, our approach aims to complement manual monitoring efforts by augmenting them with high-resolution digital information. Citizen Science offers a promising avenue to address this limitation, fostering public participation, including crowdsourced mapping, to enhance data collection and monitoring, particularly of indoor breeding sites. The primary challenge in utilizing digital data sources for mosquito mapping lies in achieving temporal alignment with entomological surveillance for modeling purposes.

Another challenge associated with digital data sources, such as satellite and street view imagery, pertains to the potential obsolescence of information and the insights derived from it. Street view images, in particular, are infrequently updated (Hou and Biljecki, 2022). It is also crucial to consider the transient nature and shifting locations of identified containers, especially for plastic bags, potted plants, non-mounted car tires, large trash bins, and dumpsters, which may have introduced a potential bias to the measured significance values of these container classes in our results. Conversely, water tanks, small city trash bins attached to streetlights, and storm drains are presumed to have relatively stable locations over time, leading to more reliable results. Furthermore, the calculated container densities in this study may be influenced by citywide solid waste collections or vector control campaigns, wherein breeding containers may have been removed before images were captured. In future studies, investigating the relationship between image timestamps and such interventions, as well as exploring alternative data sources (cf. Table 3 in publication II), could be beneficial. Crowdsourced platforms such as Mapillary (Mapillary, 2023b) and KartaView (KartaView, 2023) may particularly offer more continuous image updates (Biljecki et al., 2023).

In summary, these results demonstrated the enhanced efficiency in managing urban diseases such as dengue through the application of digital techniques. The increasing availability of spatial big data, such as satellite and street view imagery, presents a considerable opportunity for obtaining high-resolution indicators for mapping urban mosquito suitability beyond entomological sample points and allows interpolations without violating biological assumptions about limited mosquito flight ranges in the future. The proposed approach can be combined with further urban-specific mosquito proxies for enabling more targeted vector control. A task that is challenging with entomological surveillance alone. The proposed method can thus not only reduce surveillance costs but also facilitates the potential interruption of infection chains at earlier stages of an outbreak than with conventional methods.

**RQ3: What potential predictive power, in terms of explained deviance, can be achieved by harnessing GBD for estimating the environmental suitability for *Ae. aegypti*, as monitored through ovitrap egg and larval counts?**

The results presented in Table 1 in publication III underscore the degree to which hypothesis-driven urban indicators for *Ae. aegypti* suitability can capture entomological surveillance data on immature abundance collected via ovitraps in the municipality of Rio de Janeiro for the year 2019, given the constraints of a limited *Aedes* flight range. The Cohen’s explained deviance for NB-GLMs, using the seasonal mean eggs per trap (S-MET) rate as a response variable, reached up to 0.7253 and varied only marginally ( $\pm 0.003$ ) for different simulated flight range buffers. Increasing the flight range buffer from 50 m to 1 000 m exhibited similar patterns for models using the larval counts as a response variable. In this case, the predictive performance of the collected urban indicators was slightly higher, reaching a Cohen’s pseudo- $R^2$  of 0.7473 at a flight range of 200 m. This means that 74.73% of the deviance in the response is explained by hypothesis-driven urban indicators for immature *Ae. aegypti* suitability derived from openly available geospatial data. The deviance function of the NB-GLM captured the increasing variance with the mean that is typical for count data. Upon evaluating models with respect to both designated response variables, the best performance was observed in association with a flight range buffer characterized by a diameter of 200 m.

The results demonstrate the potential of hypothesis-driven urban indicators for predicting immature *Ae. aegypti* abundance at ovitrap locations, though several methodological considerations and limitations must be acknowledged. The use of digital indicators presents challenges related to (i) data availability, (ii) accuracy, and (iii) interpretation. (i) Data availability is a concern because the employed data sources are not universally accessible, limiting the generalizability of the framework to regions where similar data cannot be obtained. (ii) Accuracy is affected by the methods used to derive these indicators, which involve certain assumptions during processing. Additionally, not all data were available at a uniform spatial resolution, particularly one that aligns with the scale of *Aedes* habitats, potentially introducing inaccuracies. (iii) Interpretation of the results was complicated by high feature collinearity within the dataset, which necessitated the use of PCA. Although PCA effectively addressed this issue, it introduced a layer of abstraction that made it challenging to directly interpret the influence of individual indicators.

The selection of urban suitability indicators for immature *Ae. aegypti* in this study was hypothesis-driven. An alternative choice of indicators could have yielded different results. However, unlike much of the existing literature, the focus of this work was not on analyzing the importance of individual features, which can vary significantly across space and time and potentially offer limited practical utility, but rather on evaluating the potential of leveraging GBD to guide local interventions against *Aedes*-borne arboviral infections. Another key point in discussing our results is that our approach is inherently constrained by biases in the entomological data collection process and the influence of non-measurable micro-scale factors that affect the entomological count data used for validation. Consequently, the sensitivity of our results requires further investigation, particularly regarding different indicator selections, the inclusion of entomological surveillance data from multiple years, and the application of the framework to various case study regions. Such investigations could enhance the robustness of our findings. Follow-up research is planned to extend this framework to derive suitability indicators for the secondary vector of dengue, *Ae. albopictus*, which has been present in Brazil for nearly 30 years and is capable of transmitting YFV, DENV, ZIKV, and CHIKV. Future studies could also explore the feature importance of the proposed in-

dicators across various mosquito species, further enriching our understanding of urban mosquito suitability and its implications for vector control strategies.

**RQ4: To what potential degree can ovitrap count interpolations, refined by environmental suitability indicators, align with infestation assay indices?**

Figure 3 in publication III displays, as a highlight of this work, seasonal suitability maps for immature *Ae. aegypti* covering the municipality of Rio de Janeiro. The seasonal immature *Ae. aegypti* suitability maps accompany the spatiotemporal trend of entomological field measurements presented in Figure C. 1 in publication III. While the spatial variance of predicted immature *Ae. aegypti* suitability diverges significantly due to the small-scale heterogeneity of the urban landscape in the municipality of Rio de Janeiro, temporal effects are minimal owing to the year-long (sub)tropical climate conditions in southeast Brazil.

Figure 4 in publication III provides a more detailed insight into the results by illustrating how our best spatiotemporal model for *Ae. aegypti* larvae suitability (cf. Table 1 in publication III) performs in interpolating entomological field measurements from ovitrap locations. Specifically, it focuses on the exemplary regions of Jacarepaguá (RRJ) and Galeão (GIG) airports, chosen for their distinct spatial heterogeneity in the urban landscape, which enables a closer examination of the results at a finer spatial scale. While ovitrap larval counts and interpolated immature *Ae. aegypti* suitability remained predominantly low around both airport runways and buildings, abundance values were higher in nearby residential regions when examining measurements from the summer season of January to March 2019. The spatial heterogeneity observed in immature *Ae. aegypti* suitability at a small scale, as depicted in the map, highlights the impracticality of relying solely on the current state-of-the-art approach of coarse entomological surveillance at ovitrap sample locations or within large LIRAA strata (cf. Figure C. 1 in publication III) for targeted vector control interventions (Flores and O’Neill, 2018).

Figure 5 in publication III depicts the alignment between the generated suitability maps and spatiotemporal measurements obtained from entomological surveillance (ovitrap; LIRAA). Using locally estimated scatterplot smoothing (LOESS) analysis, we observed that the predicted suitability values did not consistently align with seasonal indices derived from LIRAA, regardless of the season (cf. Table 2 in publication III). Correlation analysis between predicted suitability and ovitrap field counts, averaged for each LIRAA zone over all seasons, reveals correlation coefficients of up to 0.76 for *Ae. aegypti* larval counts. However, the correlation between suitability predictions and LIRAA indices remains low, not exceeding a correlation coefficient of 0.08, as calculated between larva suitability and the LIRAA house index for the year 2019.

The substantial spatial disparity observed between ovitrap-based suitability predictions and LIRAA entomological indices is likely attributed to variations in collection methodologies. The BI and HI indices, which serve as block-level indicators derived from a manual sampling process, contrast with ovitrap counts obtained from fixed measurement stations purposely designed as breeding sites. These stations, characterized by water retention and a dark color to attract mosquitoes, differ significantly from manual sampling methods. Whereas LIRAA indices depend on the active search for breeding sites by health agents, the ovitrap is ‘sought’ by the mosquito, making ovitrap counts a more sensitive indicator. Furthermore, the LIRAA indices typically register low values, given the temporary nature of positive recipients and their limited persistence over time. Consequently, a comprehensive and representative sample necessitates the inclusion of numerous properties, considering that the majority may lack positive

recipients. The concentrated distribution of immature counts within LIRAA zones may have also skewed the comparative results. However, in operational terms, ovitraps could never replace LIRAA, as it provides additional indices like the “container type index”, monitoring infestation levels by container types not considered in this study.

In addition to the structural components mentioned, the reliability of LIRA indices relies on human components such as the dedication and expertise of field workers, encompassing their comprehension of vector biology and index calculation methods (Valle and Aguiar, 2023b). In a previous study, Ribeiro et al. documented a high level of coincidence between HI and BI derived from LIRAA in the municipality of Rio de Janeiro (Ribeiro et al., 2021). However, from a biological perspective, the BI is expected to be greater than the house index because the female *Aedes* spreads the eggs in close locations. Our findings on the misalignment with ovitrap counts corroborate this assessment of the fragility of the indicators generated by LIRAA.

Another factor contributing to spatial divergence is that ovitrap-based surveillance overlooks indoor breeding sites. Additionally, ovitrap surveillance concentrates on egg and larval counts, whereas LIRAA encompasses infestation by *Ae. aegypti* pupae, which exhibit distinct lifespans and lower mortality rates. Comparable findings have been reported by Nascimento et al. (2020), who additionally observed that ovitraps provide a more rapid information due to heightened sensitivity compared to LIRAA in detecting *Ae. aegypti*. Getis et al. (2003) indicated spatial divergence between immature and adult *Ae. aegypti* populations. As the life cycle of immature *Ae. aegypti*, from emergence as 1st instar (L1) larvae to adulthood, is estimated to be around 8-10 days, varying with humidity and temperature conditions (Center for Disease Control and Prevention, 2024; Hossain et al., 2022), comparing entomological data from different surveillance techniques on a seasonal level may be too coarse (Cromwell et al., 2017; Morrison et al., 2004).

In addition to the previously discussed arguments, the presented Bayesian modeling results are highly sensitive to the choice of priors, especially in constructing the Matérn covariance field. The most critical assumptions in our case study were made for the Penalized Complexity (PC) priors for the parameters’ range and marginal standard deviation of the Matérn field, modeling, among other factors, the extent of mosquito movement in space in our case study. Besides that, passive dispersal of mosquito eggs and adults, driven by transport and trade (Bennett et al., 2019; Díaz-Nieto et al., 2016; Eritja et al., 2017; Guagliardo et al., 2015) was completely neglected in this dissertation. Subsequent investigations may delve into methodologies for establishing these priors through more entomological surveillance and bio-ecological field studies. This would help eliminate potential bias in immature *Ae. aegypti* suitability maps and could also enable fine-tuning of the proposed framework for other mosquito species.

### 3.3.2 Daytime *Aedes*-human Interactions (RO II)

#### Key Findings (Research Objective II)

Integrating environmental suitability estimates for *Aedes* (cf. M1-M3) with human movement patterns (cf. M4) can enhance the precision of spatial predictions for dengue occurrence (cf. M5-M6). This finding was validated in a case study of the municipality of Rio de Janeiro, where a spatial model for predicting dengue occurrence (2015-2022) that incorporated a feature-engineered variable to capture hourly variations in *Aedes*-human interactions outperformed a baseline model that did not account for daytime variance in hazard exposure, resulting in a 13.5% improvement in predictive accuracy. These results highlight the need to incorporate *Aedes*-human interaction hotspots into local intervention strategies. The analysis revealed significant discrepancies between mosquito abundance, interaction points, and reported disease cases for the municipality of Rio de Janeiro. Therefore, a novel guidance map for local interventions was proposed, combining mosquito distribution, dengue occurrence, and human movement patterns, identifying potential transmission hotspots that may be underestimated by entomological surveillance or occurrence data alone (cf. Figure 6 in publication V).

While mobile phone data significantly improved predictive accuracy, potential biases may have been introduced by excluding individuals without mobile phones or those using alternative service providers. Another limitation is related to the spatial resolution of the analysis, as mobile phone records could only be georeferenced by antenna tessellation, which may have affected spatial precision. Future studies could address this by validating the findings using higher-resolution mobility data, such as GPS trajectories. Additional preprocessing challenges, such as accounting for offline movements and antenna congestion, further complicated the accurate extraction of human movement patterns. While upper and lower boundaries for inter-event times were implemented to mitigate data bias, future research could explore the impact of these parameter choices on the retrieval of human movements from mobile phone data. Incorporating cross-boundary human movements to out-of-city regions, which were not included in this dissertation, could additionally contribute to refining exposure estimates and enhancing overall predictive accuracy (cf. publication IV).

Given the restricted access to mobile phone data, this dissertation also evaluated the use of geotagged tweets as a more openly available data source for modeling human movement patterns. The corresponding findings revealed the need for caution when using Twitter/X data for short-term urban mobility modeling, as it is vulnerable to policy changes and fluctuations in the availability of publicly accessible geotagged tweets. However, the 27-month validation study demonstrated that combining multiple mobility metrics, analyzing both dynamic and static mobility changes, and employing robust preprocessing techniques - such as rolling window downsampling - can improve the inference capabilities of Twitter/X data. Nevertheless, despite the application of these advanced methods, Twitter/X data may not always perform as well as mobile phone records in capturing human movement patterns, as validated during the COVID-19 pandemic using the stringency index, which measures the strictness of government-imposed mobility restrictions (cf. publication IV).

Future research could build on this validation study by exploring the potential of additional openly available data sources for retrieving human movement patterns. These may include data from public transit systems, ride-sharing apps, delivery services, household surveys, wearables, smart city sensors, and volunteered geographic information from platforms like Strava and Waze. Such data sources may offer valuable alternatives to mobile phone data and could provide critical insights into movement dynamics, essential for accurately modeling the spread of infectious diseases.

Another key limitation of the results lies in the reliance on two relatively basic, temporally static statistical models for *Aedes*-human interactions: one assuming constant human exposure to mosquito bites throughout the day, and the other accounting for diurnal fluctuations. While these models offer valuable insights for comparing theoretical frameworks, they fall short of fully capturing the complexity of transmission dynamics. Future research could address this limitation by developing spatial process-based models that simulate transmission dynamics (Kache et al., 2022b; Wu et al., 2023). Though computationally demanding, such models would allow for the integration of mosquito behavior, ecological factors, and feedback mechanisms, including immunity dynamics and transmission cycles, to provide a more comprehensive understanding. In addition, incorporating the daytime variation in human host density across urban areas could enhance the models' capacity to reflect changes in local vectorial capacity, as fluctuations in mosquito biting behavior and mosquito-to-host ratios could significantly influence transmission risk. This approach would offer more refined insights into the efficacy of prevention and control strategies, presumably improving disease management in urban environments to a greater extent, as supported by the findings presented here. Besides that, future research could prioritize the development of spatiotemporal models that directly integrate high-resolution environmental suitability data and mobile phone data, rather than relying on spatiotemporal feature engineering. Adapting the proposed models to different mosquito species or regions will require accounting for variations in biting behavior influenced by factors such as day length. Seasonal shifts in sunrise and sunset times could further refine risk predictions. Moreover, incorporating pathogen penetration rates in both host and vector populations may enhance the predictive accuracy of dengue occurrence models. However, widespread testing is often constrained by high costs, logistical complexities, and limited availability of advanced laboratory infrastructure.

Overall, these findings underscore the importance of accounting for both vector ecology and human behavior in disease modeling. A major challenge persists due to the lack of data at the resolution required to accurately capture real-world phenomena, often limiting biologically sound interpretations of transmission dynamics derived from eco-epidemiological models.

**RQ5: To what potential degree can X data be leveraged to model long-term spatiotemporal mobility metrics derived from mobile phone data, and to what extent can rolling window downsampling enhance this modeling?**

Figure 7 in publication IV illustrates the results related to this research question. Examining the initial time period of analysis spanning from April 2020 to September 2020, all computed mobility metrics derived from Twitter exhibited discernible patterns that aligned with our expectations based on the implemented lockdown measures in the city of Rio de Janeiro. Notably, while the long-term trend of the graph modularity metrics and the percentage of activity in residential areas decreased, the long-term trends of average movement distance, overall movement volume, and the radius of gyration increased.

During the subsequent time period from September 2020 to May 2021, all mobility metrics derived from Twitter, except the percentage of activity in residential areas, displayed unexpected changes. They all showed a rapid shift starting in February 2021 dis-aligning our assumptions on more or less constant mobility behavior in that time period. Coinciding with this period, there was a sharp decline in the number of geolocated tweets collected via the public Twitter API (cf. Figure 2 in publication IV). We hypothesize that this decline was attributed to changes in the terms of use implemented by Twitter. However, official evidence of regulatory changes during that specific time period has not been found. Additional experiments using a constant amount of tweets per day, derived by the 98<sup>th</sup> percentile of tweet volume in the corresponding rolling window subset, showed a similar shift in mobility metrics (cf. Figure A. 3 in publication IV). This highlights the robustness of calculated mobility metrics in the face of daily fluctuations in the number of tweets.

For the analysis period subsequent to May 2021, the calculated mobility metrics once again aligned with our expectations and confirmed our knowledge of fewer mobility restrictions implemented in the city of Rio de Janeiro following the COVID-19 pandemic.

The results also demonstrate that, while a moving average can effectively eliminate weekly fluctuations and data noise, it does not suffice for generating accurate long-term trends for all considered mobility metrics in this analysis. However, when combined with the specifically designed rolling window downsampling (RWDS) approach, more precise long-term mobility trends can be derived. This effect becomes particularly evident when examining the calculated graph modularity metrics in our case study, as the modularity values between the one-day window size signals and the seven- or 27-day rolling window size signals exhibit larger differences. In contrast, for other calculated mobility metrics, the impact of RWDS appears to have relatively low significance and yields effects comparable to those obtained by calculating a one-day window trend signal. Supplementary materials provide corresponding results of daily mobility metrics calculated without applying a moving average (cf. Figure A. 3 in publication IV). The influence of different rolling window sizes is more extensively investigated in the subsequent section in conjunction with long-term trends derived from mobile phone data.

Long-term validations of urban mobility metrics derived from Twitter are infrequent, despite the well-established usage of Twitter applications in various research domains worldwide. However, the outcomes of our comprehensive long-term validation study emphasize the need for caution when utilizing Twitter data for urban studies within restricted time frames. Although urban mobility metrics derived from Twitter may exhibit high correlation values with mobility metrics computed from mobile phone data during short time periods, long-term validation with mobile phone data reveals

fluctuating deviations (cf. Figure 9 in publication IV). This phenomenon can potentially give rise to erroneous assumptions when relying solely on Twitter as a reliable source for modeling human movement patterns.

The results presented in section II in publication IV demonstrate that the RWDS method is a valuable tool for addressing the data scarcity challenge associated with urban Twitter data and deriving more precise long-term mobility trends. However, additional findings highlight the significant dependence of these findings on the chosen rolling window size (cf. Figure 8 in publication IV). In our experiments we observed the highest average correlation value between mobility metrics from Twitter and mobile data when using an 11-day rolling window size. Increasing the window size from one day to three days had the most pronounced effect on the calculated Pearson correlation values. For window sizes exceeding 11 days, the correlation values remained consistently high but showed a slight flattening. This can be attributed to the loss of high-resolution information resulting from the application of larger window sizes beyond 11 days. These findings align with our expectations regarding the functionality of the RWDS method described in section II in publication IV. The mean movement distance index yielded the highest average Pearson correlation coefficient among all considered mobility metrics, achieving its peak of 0.48 at the 11-day rolling window downsampling size (cf. Table A. 1 in publication IV).

During the dynamic analysis of the long-term trend of calculated mobility metrics using moving window synchrony, it becomes evident that the Pearson’s correlation coefficients exhibit substantial variations over time for all the calculated mobility signals (cf. Figure 9 in publication IV). We observed the occurrence of short time periods characterized by both extremely high and extremely low correlation values. These findings indicate that the informative capacity of mobility metrics derived from Twitter exhibits temporal variability and is strongly contingent upon the chosen time frame for analysis. During the initial phase of the study period, when the most stringent mobility restrictions were implemented (cf. Figure 6 in publication IV), we observed high positive correlation values across all metrics simultaneously. Conversely, we did not observe similar prolonged time periods characterized by a weak alignment, as indicated by low Pearson’s correlation coefficients around zero. Notably, higher moving window correlation values exhibited greater statistical significance than lower values.

To eliminate the possibility of spurious correlations, all time series were examined for unit roots using the appropriate version of the Dickey-Fuller test before calculating Pearson correlation coefficients. The test results indicated that seven out of ten time series were stationary, allowing for the calculation of Pearson correlation coefficients. However, the time series for “Number of movements”, “Graph modularity”, and “% activity in residential areas” measured based on Twitter data, remained non-stationary. Following the “Standard sequence of steps for dealing with non-stationary time series” as outlined by Studenmund, 2017, we tested the pairs of Twitter data and mobile phone time series for the metrics “Number of movements”, “Graph modularity”, “% activity in residential areas” for cointegration using the Engle-Granger test. The Engle-Granger test results indicated that the time series for the metrics “Number of movements” and “% activity in residential areas” were cointegrated at a confidence level of 95%, while the time series for the metric “Graph modularity” were cointegrated at a confidence level of 90%. According to Studenmund, 2017, if the variables have unit roots and are also cointegrated, this allows for the calculation of the Pearson correlation coefficient using the original units, thereby ruling out spurious correlations.

Additional findings from a static change detection analysis reinforce the results of our long-term trend analysis (cf. Figure 10 in publication IV). While it is evident that Twitter data does not always accurately capture long-term mobility trends, it does

have the potential to detect significant (cf. Table A. 2 in publication IV) inner-urban mobility changes measured by mobile phone data and indicate the correct direction of the shift. In our case study, this holds true for all the measured variables except for the percentage of activity in residential areas during the time period of the second onset. In summary, we conclude that both the Twitter and mobile phone datasets synchronously detected the shift in inner-urban human movement behavior between the years 2020, 2021, and 2022, attributable to COVID-19 lockdown policies. Static mobility changes between weekdays and weekends were not detected to be significant (cf. Table A. 2 in publication IV) when testing both datasets, leading to the conclusion that Twitter can be a useful substitute for mobile phone records when trying to derive the direction of static inner-urban mobility shifts.

We performed a sensitivity analysis of various window sizes for RWDS. Thereby, we employed a combination of different modeling techniques. This included a dynamic mobility trend analysis and a static mobility change detection. In addition, we considered a set of five distinct mobility metrics. However, our findings show certain limitations, primarily stemming from the choice of a 28-day moving average for trend calculation, a 60-day window synchrony time frame for analyzing dynamic alignment of trend signals, and the temporal selection of on- and offsets for static change detection analysis. Furthermore, our results may be subject to potential biases due to the uneven distribution of Twitter user groups within the overall population (Li et al., 2013; Malik et al., 2015). We did not account for the spatial distribution of inferential uncertainty in our analysis either, although districts with fewer geocoded tweets can be expected to exhibit a higher degree of uncertainty (Huang and Carley, 2019; Huang and Wong, 2015). This particularly affects the graph modularity metrics calculated based on daily OD matrices. The spatial distortion in the applied datasets is supported by the low correlation of non-zero OD matrix entries aggregated over the entire analysis period (cf. Figure 11 in publication IV). Additional results from spatial data exploration, which highlight these issues, are provided in the supplementary GitHub repository (Knoblauch and Gross, 2023).

To address these limitations, several approaches might be applicable: Recent studies on semantic analysis (Hu et al., 2023; Serere et al., 2023) demonstrate promising results in deriving geolocalized information from tweet texts of non-geolocated tweets, which could enhance the Twitter dataset with supplemental geoinformation. Another approach involves utilizing the locations provided in user profiles as a further source of geoinformation. However, it should be noted that these techniques have limited applicability in the context of inner-urban mobility studies (Nguyen et al., 2022).

Another aspect of discussion in our long-term validation study pertains to the disparate spatial and temporal resolutions of the employed datasets. Additionally, the raw Twitter data utilized represents less than one percent of the total mobile phone records used in this validation study, leading to a substantial imbalance with potential implications on our validation outcomes (Zhao et al., 2021). Furthermore, certain assumptions were made during the preprocessing stage to facilitate the generation of our validation signal. These assumptions include the selection of lower and upper bounds for IET filtering and the assumption of a uniform distribution of cellular activity in space when converting antenna-based OD matrices into neighborhood-based mobility flows. Additionally, we assumed that the sequential activities of individual users directly represent movements, disregarding the possibility of detours which may introduce a bias in our results. However, we believe that the overall impact of these constraints is relatively minor. We anticipate that conducting supplementary sensitivity analyses on the model parameters would not alter the main findings of this novel long-term validation study, primarily because all parameters and steps were carefully chosen and justified.

Overall, our findings indicate that rolling window downsampling is an effective strategy for mitigating the limited availability of geolocated tweets in urban areas (cf. Figure 7 in paper IV). Our results indicate that Twitter has the potential to capture short-term changes in mobility at an inner-urban scale (cf. Figure 10 in publication IV), although long-term disparities were observed when compared to mobility metrics derived from mobile phone data in our case study (cf. Figure 9 in publication IV). To enhance the reliability of short-term inference from Twitter data on inner-urban human movement patterns, we propose a combination of multiple analysis techniques, including dynamic and static mobility change detection, simultaneous consideration of various human movement metrics, and sensitivity analysis for modeling parameters. Implementing these approaches can significantly mitigate the risk of false inference in diverse application domains where Twitter is commonly utilized as an open-source proxy for deriving human movement patterns.

Considering the increasingly stringent open-access limitations to Twitter data, these results of the long-term validation study establishes a foundation for assessing the validity of also upcoming social media platforms. Voluntarily shared geosocial media data can be a powerful and promising tool, especially in locations where other mobility data sources are not openly-accessible or too costly to generate. Since the availability of data sources significantly impacts applications, future research should encompass not only data performance metrics for delineating mobility patterns but also sustainability in terms of long-lasting and openly accessible APIs. Another research option could involve the fusion of data from multiple sources such as Waze, GDELT, Facebook, Instagram, Reddit, Telegram, YouTube, or Weibo. The methods developed could then be transferred to other geosocial media platforms. Besides that developed methods and generated insights could always be applied with payment plans for API access offered by Twitter.

By conducting this study, our aim was not only to support researchers in effectively utilizing social media data for modeling human movement patterns but also to gain valuable insights into human mobility within the city of Rio de Janeiro, Brazil. These findings might open up new avenues for future research on unexplained mobility-driven phenomena in urban science, such as the location of informal economy (López-García, 2023), accessibility impacts of transport policy (Pereira, 2019), and inner-urban transmission processes of MBDs (Ramadona et al., 2019). Overall, the results for RQ5 led us to use to utilize mobile phone data for further analysis, even though it is not always openly available and does not offer a truly scalable approach.

**RQ6: To what potential extent can a model incorporating daytime *Aedes*-human interactions improve inference on spatial dengue occurrence, compared to a model that is identical except for neglecting human movement?**

The results in Table 1 in publication V demonstrate how considering mosquito biting hours and human movement corridors can enhance the accuracy of spatial estimates for urban DENV occurrence. The proposed feature engineering method outperforms the baseline model, which does not consider the daylight activity of *Aedes* mosquitoes, and demonstrates a 13.5% increase in the explained deviance within the response of the QP-GLM. Both models yielded positive and significant estimates for their hazard and exposure combined covariate of human-mosquito biting risk  $B_i$ . The computed global Moran's I value for the residuals was 0.59. Considering the aforementioned results, it implies that integrating knowledge of *Aedes* biting behavior with human movement patterns can also facilitate the inference of probable transmission sites for reported

dengue cases. If this holds true, increased mosquito control interventions in these locations would have the potential to combat *Aedes*-borne diseases more effectively.

Figure 6 in publication V presents, as a highlight of this work, the practical implications of these research findings for the municipality of Rio de Janeiro. A novel mapping approach for vector control intervention was developed, incorporating (i) the spatial distribution of mosquitoes, as indicated by temporally aggregated entomological surveillance data, (ii) the spatial dispersal of dengue occurrence, and (iii) the most likely transmission locations for reported dengue cases, taking into account daytime *Aedes* biting behavior. This target effectiveness map marks regions that were potentially underestimated for vector control planning using entomological datasets only, while at the same time emphasizing the enduring importance of areas with high mosquito abundance.

The results showed that the inferred degree of spatial variation in urban DENV occurrence was sensitive to assumptions about daytime mosquito activity. Spatial discrepancy existed between the dominant location of mosquitoes, the spatial patterns of human-mosquito interaction points, and disease occurrence collected by residency. Taking these findings into account, one can conclude that methodologies that presume consistent human exposure to mosquito bites throughout the day potentially yield exaggerated and biologically inadequate interpretations regarding the patterns of disease transmission. Additional knowledge about pathogen penetration rates in host and vector populations would potentially enhance prediction capabilities for urban DENV occurrence. However, the practical challenges associated with establishing such virus penetration measurements, e.g. within entomological surveillance systems, pose significant obstacles, especially due to the need for appropriate laboratory infrastructure. The utilization of mobile phone data as a proxy for human movement in the present study could have resulted in additional inherent constraints. Despite the high penetration rate of the mobile phone provider, mobility estimates may have been biased due to the exclusion of individuals without mobile phones or those using different services. To counteract this factor, an improvement strategy could involve integrating social media streams. Higher-order descriptions of movement, such as social network structure, have been shown to affect transmission dynamics in urban environments (Reiner et al., 2014; Stoddard et al., 2013; Vazquez-Prokopec et al., 2013). The consideration of the interplay among disease symptoms, infectiousness, and the mobility of individuals infected with DENV seems additionally promising in this context (Perkins et al., 2015; Perkins et al., 2016; Schaber et al., 2019). This complicates the assumption that the movement patterns of apparently healthy individuals can adequately represent the mobility patterns of those involved in transmission (Wesolowski et al., 2016).

Follow-up activities could connect by examining cross-boundary human movements to out-of-city regions and potentially model on individual human scale instead of aggregating risk value into areal units. The fixed scenario employed to model the daytime risk of human-mosquito biting could also be extended to consider seasonal fluctuations in sunset and sunrise times. We showed that approximating resident's exposure to different mosquito populations throughout the city, leveraging human movement flows derived from mobile phone data, results in higher predictive power than focal mosquito abundance alone. This highlights the importance of considering mechanisms driving human-mosquito interactions for understanding of mosquito-borne disease occurrence at urban scale. We employed a relatively basic and temporally static statistical model. Future studies could leverage similar data for building spatial process-based models of intra-city transmission dynamics (Kache et al., 2022b; Wu et al., 2023). Such models, although computationally and conceptually challenging, could effectively incorporate mosquito behavior as well as ecology and capture feedback processes, such as immunity

dynamics and transmission cycles, among others. Furthermore, daytime variation of human host density across the city could be incorporated, potentially impacting local vectorial capacity by modulating mosquito biting behavior and mosquito-to-host ratio. Implementing this modeling approach would yield additional insights on the efficacy of prevention and control strategies, thereby enhancing our understanding and management of MBDs in urban environments.

**RQ7: What potential predictive power, in terms of explained deviance, can be achieved for inference on spatial dengue occurrence by considering daytime *Aedes*-human interactions, vulnerability indicators, and spatial autocorrelation?**

We hypothesized that incorporating vulnerability indicators and spatial eigenvectors would further enhance the proposed QP-GLM, which considers *Aedes*-human interactions (cf. Equation 4) for predicting the spatial occurrence of dengue in the municipality of Rio de Janeiro. The Cohen’s pseudo-R<sup>2</sup> of the more extensive QP-GLM considering hourly human-mosquito biting risk was determined to be 0.77, indicating that the extended model was capable of explaining up to 77 percent of the deviance in dengue occurrence on the sub-neighborhood level for the municipality of Rio de Janeiro in the year 2022. The computed global Moran’s I value for the residuals was 0.07, indicating low spatial autocorrelation. A QP-GLM with human-mosquito biting risk and vulnerability indicators but without spatial eigenvectors was not considered, as it yielded a higher overdispersion value of 26.85 and a higher global Moran’s I of 0.2, despite having a Cohen’s pseudo-R<sup>2</sup> of 0.83. This underscores the importance of vulnerability indicators and spatial eigenvector mapping in improving spatial predictions of sub-neighborhood dengue occurrences, which are georeferenced based on residency. Additional result on the applied vulnerability indicators and spatial eigenvectors are listed in Table B. 1 and Figure B. 2 in publication V.

Among the applied vulnerability indicators, the socio-economic variable of average income emerged as the most influential predictor, demonstrating a negative association (Table B. 1 in publication V). This suggests that higher average income levels in the municipality of Rio de Janeiro are associated with a reduced risk of dengue infections. Conversely, the indicators of the hypothesized vulnerability categories accessibility and centrality did not exhibit significant predictive power, contradicting our initial assumptions. The same applied to the density of older individuals. However, the density of individuals under five years emerged as a significant predictor in our model, exhibiting a negative estimate, which suggests that a higher concentration of children was linked to fewer dengue cases. This could potentially be explained by the fact that first dengue infection per individual have a higher probability of being clinically mild (Guzman et al., 2016). Additionally, behavioral factors could play a role, as households with young children may be more vigilant in implementing mosquito control measures, thereby reducing dengue transmission.

In the assessment of the vulnerability class of immunization, our analysis indicated that the calculated significance values for each year of past infections depend on the magnitude of the outbreak pattern (cf. Figure B. 1 in publication V). The years with minor outbreaks and less spatial variance in DENV occurrence (2017, 2018, 2020) exhibited either no significant association or marginally statistically significant association, whereas the major outbreak years with larger spatial variance in DENV occurrence (2016, 2019) showed lower p-values. The most recent year in our analysis, 2021, yielded the highest p-value among the hypothesized immunological vulnerability indicators, despite the occurrence of lower dengue incidence. Surprisingly, most esti-

mates of this vulnerability category were positive, contrary to our initial hypothesis about past infections conferring population immunity. We hypothesized that this is related to the fact that environmental factors facilitating transmission are overruling marginal gains in population immunity (under the assumption that cross-immunity between subsequent serotypes or genotypes is relevant). The complex immunological interactions between infections with the four dengue serotypes over time are not further discussed in this context (Simmons et al., 2012). In brief, past infections with a heterologous serotype confer short-term cross-immunity, while past infections with a homotypic serotype confer long-term immunity to the same serotype. The duration and effect size of the heterologous cross-immunity and potentially enhancement is dependent on the time interval between the infections as well as on the specific sequence of the serotypes and their genotypic similarity (Guzman et al., 2016; Katzelnick et al., 2015). These complex immunological interactions between dengue serotypes make it challenging to utilize spatial distribution patterns of dengue cases from previous years to model immunity levels. Here we can only show the possible existence of a confounding factor not accounted for in the model but influential in driving the spatial distribution of dengue cases at the sub-neighborhood scale.

It is important to note that the presented results are dependent on the selection and calculation methods for vulnerability indicators and spatial eigenvectors. Not all additional variables showed significance in our model, underscoring the nuanced impact and selective relevance of certain variables within such broader predictive frameworks. The significance of the proposed human-mosquito biting risk indicator did also diminish in a more extended spatial model. The spatial eigenvectors (cf. Figure B. 2 in publication V) effectively absorbed a significant portion of the spatial autocorrelation present in the residuals of the proposed QP-GLM. They can be instrumental in formulating additional hypotheses regarding potential missing covariates or confounding factors for integration into future models. Furthermore, they can serve as a tool for testing spatially varying regression coefficients.

## 3.4 Research in Context

### 3.4.1 Environmental Suitability for *Ae. aegypti* (ROI)

#### Evidence Before this Dissertation

Few studies have looked at the urban suitability for immature *Ae. aegypti*. In February 2024, we searched articles indexed on PubMed using the syntax “(Urban AND *Aedes aegypti* AND Suitability)” and obtained 20 search results. While the diverse urban environment and the restricted flight capacity of *Ae. aegypti* can result in significant spatial variation in abundance, only three studies have incorporated this bio-ecological understanding into their models of *Ae. aegypti* suitability. None of them generated continuous suitability maps covering a whole municipality, which is particularly relevant for guiding local vector control. Modeling the suitability of immature *Ae. aegypti* for large municipalities considering the limited flight range of *Aedes* remains challenging. Given the scarcity of significant suitability indicators, particularly at *Aedes* habitat sizes.

#### Added Value of This Dissertation

The added value of this dissertation lies in the development of a comprehensive set of hypothesis-driven urban landscape indicators and geospatial methods to model the spatiotemporal likelihood of hosting *Ae. aegypti* populations. Scientific advancements were achieved, particularly in the realm of spatial resolution, aligning with an *Ae. aegypti* habitat size of 200 meters. We demonstrated for the *Aedes*-endemic municipality of Rio de Janeiro how the proposed suitability indicators, derived from openly available GBD, can explain the distribution of *Ae. aegypti* egg counts measured with ovitraps by up to 73% and the dispersion of larval counts by up to 75%. By enriching sample-based entomological field measurements from ovitraps with the proposed indicators, we were able to create the first continuous *Ae. aegypti* suitability maps covering a whole municipality on a seasonal basis while considering *Ae. aegypti* habitat size.

#### Implications of All The Available Evidence

Mapping the high spatial variability of *Ae. aegypti*, which can occur under suitable weather conditions due to heterogeneous landscape and limited *Ae. aegypti* flight range, is relevant for advancing vector control. The proposed indicators have substantial value to: (i) inform and optimize targeted vector control interventions such as *Wolbachia*; (ii) allow cost savings in entomological surveillance; (iii) reduce environmental pollution, including mosquito insecticide resistance; and most importantly, (iv) provide more efficient overall disease control of *Aedes*-borne diseases such as yellow fever, dengue fever, Zika, and chikungunya.

### 3.4.2 Daytime *Aedes*-human Interactions (RO II)

#### Evidence Before this Dissertation

Few studies have looked at the occurrence of dengue on a city scale, incorporating human movement patterns. In May 2024, we searched articles indexed on PubMed using the syntax "(Urban AND Dengue AND Human Mobility)" and obtained 42 search results. Although mosquito flight range is estimated to be below 1 000 m without the assistance of wind, only 2 studies analyzed the urban spread of dengue below the neighborhood scale. None of these studies additionally considered the diurnality of *Ae. aegypti* and *Ae. albopictus* biting behavior, nor the corresponding fluctuation in human exposure. Modeling dengue occurrence at a high spatial resolution while considering the characteristics of *Aedes* mosquitoes is challenging due to the scarcity of (i) high-resolution data on urban mosquito abundance, (ii) daytime human movements, and (iii) sub-neighborhood knowledge about dengue infections.

#### Added Value of This Dissertation

To my knowledge, this dissertation is the first study to analyze the spatial distribution of urban dengue occurrence at a sub-neighborhood scale, considering ecological and behavioral vector characteristics, such as diurnal biting behavior, and corresponding daytime variations in human exposure to mosquito bites. The findings show that modeling human movement patterns on an hourly basis can bring significant benefit to the explanation of urban dengue outbreaks in the municipality of Rio de Janeiro. With these insights, we were able to showcase the potential of integrating knowledge about human mobility for the design of vector control interventions. This dissertation suggests that the spatial allocation of resources for mosquito control should not only consider mosquito abundance but also prioritize locations with high connectivity during mosquito bite hours to areas with a high dengue reporting rate.

#### Implications of All The Available Evidence

The evidence presented underscores the critical role of high-resolution data in eco-epidemiological models, which aim to accurately simulate transmission dynamics by integrating both vector ecology and human behavior at fine spatiotemporal scales. However, implementing such models is challenging due to the widespread lack of high-resolution data, including crucial information on human movement patterns, pathogen penetration rates, serotype prevalence, and health occurrences. While the abstract models presented in this study validate theoretical concepts, fostering an open-sharing culture of anonymized, high-resolution data relevant to disease transmission would facilitate the development of more advanced models that more accurately capture the complexities of transmission dynamics. Building on these efforts, further advancements in the integration of human and vector dynamics have the potential to significantly enhance our understanding of vector-borne diseases and improve the effectiveness of local intervention strategies for diseases such as dengue.

### 3.5 Future Directions Towards Practical Application

Building on the reported research findings and key contributions, this section outlines future directions for enhancing the real-world applicability of the proposed hybrid local intervention guidance for *Aedes*-borne arboviral infections. It critically examines key aspects such as scalability, transferability, and replicability of the proposed concepts, while also identifying remaining research gaps and highlighting emerging technologies that have the potential to facilitate reliable implementation in the near future.

#### Towards scalability

To successfully deploy the proposed hybrid local intervention guidance in real-world settings, which integrates entomological surveillance with digital monitoring, scalable workflows are essential. To promote such workflows and provide cost-sensitive solutions, this research aimed to (i) develop scalable methods and (ii) utilize openly available datasets where available. Building on these two foundational objectives, one key contribution of this work is the fine-tuning and publishing of computer vision models to detect common *Aedes* breeding containers using openly available satellite and street view imagery.

The designed semi-supervised water tank detection model, trained on openly available satellite imagery (cf. publication I), minimizes the need for manual labeling - an often resource-intensive requirement for training object detection models. By employing a pseudo-labeling technique, the model generates high-confidence predictions at an intermediate stage of training, which are then incorporated for further fine-tuning. This algorithmic design enhances scalability and improves model performance and generalization while reducing costs, especially important for regions with limited public health budgets. Once trained, the model's computational complexity scaled linearly with increasing data size during prediction, ensuring efficient deployment across large application areas. However, despite its scalability, the approach was limited to detecting larger *Aedes* breeding containers, such as water tanks, when using openly available satellite imagery. To overcome this limitation, our research subsequently focused on expanding detection capabilities to smaller breeding habitats by leveraging alternative GBD sources, specifically openly available street view imagery.

While the multi-object detection network applied to street view imagery demonstrated the capability to detect smaller breeding containers (cf. publication II), it faced more pronounced limitations in scalability compared to satellite-based methods. These limitations were primarily due to inconsistencies in data availability, rather than model performance, which remained comparable to that of satellite imagery-based approaches. Street view imagery is inherently confined to areas accessible by vehicle, introducing spatial bias as coverage is restricted to road networks where images have been captured. This bias is particularly evident in regions with heterogeneous road infrastructure and urban design, such as our case study area, where a marked disparity in street view imagery availability was observed between favelas - characterized by narrow, non-car-accessible streets - and other parts of the municipality with wider, car-accessible roads. Furthermore, areas with higher crime rates, despite being car-accessible, often lacked coverage in street view imagery, revealing additional spatial biases shaped by socio-economic factors and safety concerns, rather than solely by urban infrastructure.

While Google Street View (GSV) provided the most extensive and uniform spatial coverage across the municipality of Rio de Janeiro, along with more frequent temporal updates compared to crowdsourced platforms like Mapillary and KartaView, the scalability of GSV-based applications may be limited when extending beyond municipal

boundaries. Specifically, in rural areas, GSV coverage is often sparse or entirely absent, posing challenges for broader geographic applications. Similar limitations arise when scaling these approaches to other *Aedes*-endemic countries, such as Cuba, Papua New Guinea, or China, where GSV is banned.

Given these insights on provider-specific local data availability, this thesis suggests that future research aiming to scale the detection of *Aedes* breeding containers using street view imagery should begin with a comprehensive evaluation of all available street view data sources in the study area. This initial assessment is critical for identifying data availability and addressing spatiotemporal biases to ensure the reliable scalability of the proposed methods. Moreover, while crowdsourced platforms may offer alternatives to proprietary datasets, they present inherent challenges, including non-standardized image formats and inconsistent 360-degree coverage. These factors complicate data processing and the integration of multiple street view imagery sources into a single application.

Airborne imagery, such as orthophotos from drones, could provide an additional GBD source for detecting *Aedes* breeding containers, particularly in small-scale applications. This imagery offers the potential for (i) spatially continuous detection of breeding containers, including those located on rooftops and in backyards, similar to satellite imagery, and (ii) the ability to detect smaller container types similar to street view imagery. As a result, airborne imagery holds the potential to combine the strengths of both satellite and street view data. However, its scalability is constrained by high acquisition costs and the need for regulatory permissions, which limit its feasibility for large-scale applications, especially in resource-constrained settings.

Given these limitations of openly available GBD for detecting common *Aedes* breeding containers, this study recommends that future research prioritize the use of ultra-high-resolution satellite imagery - specifically, imagery with a resolution exceeding 0.03 meters per pixel, as utilized in this research (cf. publication I). At the time of this study, ultra-high-resolution satellite imagery was only available for purchase from providers such as DigitalGlobe, Planet Labs, or Skybox, making it incompatible with the cost-sensitive focus on scalability and broad applicability. However, free-to-use access to such ultra-high-resolution satellite imagery may change in the near future, similar to scientific opportunities for collaboration with commercial providers, particularly those promoting initiatives for social good. Although not tested in this study, ultra-high-resolution satellite imagery could enhance detection capabilities by (i) capturing a broader range of *Aedes* breeding containers, and (ii) reducing spatiotemporal biases that arised during this study from combining coarse satellite imagery with street view data collected over multiple timestamps. Despite these potential advantages, inherent challenges of top-down satellite imagery and bottom street view imagery will remain, such as the inability to detect indoor breeding sites. This remaining limitation underscores the importance of future research focusing on complementary strategies, including citizen science initiatives, such as crowdsourced mapping platforms (e.g., the Mosquito Alert platform (Bartumeus et al., 2018; Capineri et al., 2016; ICREA, 2024; Jennex et al., 2017; Južnič-Zonta et al., 2022; Pataki et al., 2021)), which can help identify indoor breeding habitats.

Building on the discussion of scalability and spatial biases in GBD, it is equally important to address the temporal aspects of data richness, which should not be assumed constant over time. This became evident in the long-term validation study on the use of X data for tracking urban human movements (cf. publication III), which revealed that the accuracy in detecting mobility pattern changes fluctuated due to temporal variations in data access, particularly when compared to proprietary mobile phone records. Short-term validation studies may have overlooked these limitations,

underscoring the need for comprehensive evaluations of GBD-based applications across both spatial and temporal dimensions before scaling the proposed methods. While the increasing availability of openly available GBD (cf. Section 2.1) generally supports the scalability of the proposed workflows over time, unforeseen changes in data formats or sudden restrictions - such as those observed for X data (cf. publication IV) - pose significant challenges to temporal scalability. Addressing these issues will be essential for future research and applications. Potential solutions could include: (i) fostering collaborations with GBD providers to ensure stable and continuous access to data, (ii) advancing data fusion techniques to integrate multiple data sources, thereby reducing reliance on any single dataset to improve the adaptability and robustness of the proposed workflows.

In our foundational objectives, scalability was defined as the method's capacity to handle large volumes of geospatial big data (GBD) and to be applied across the municipality of Rio de Janeiro, covering 1,221 km<sup>2</sup>. This involved the use of GBD, including (i) 10 million satellite images (cf. publication I), (ii) approximately 500,000 street view images (cf. publication II), (iii) around 700,000 tweets (cf. publication IV), and (iv)  $1.6 \times 10^{11}$  mobile phone records (cf. publication IV), alongside many smaller geospatial datasets, which might have varying structures and availability in other case study regions. Insights from this thesis indicate that the primary limitation to scalability for both research objectives was not the performance of individual models, but rather the quality of the GBD, particularly in terms of (i) availability, (ii) resolution, and (iii) consistency in structure. To enhance scalability and enable real-world applications, future research could focus on developing data standards and quality indicators for each GBD source, which could be used to correct spatiotemporal biases and create more robust workflows. Additionally, increasing the degree of automation in data acquisition, processing, and integration would further strengthen the proposed concepts. Until such workflows are established, this research remains a proof of concept, harnessing GBD to demonstrate the potential of geoinformatics applications in the public health sector.

### Towards transferability

Transferability refers to a method's ability to maintain its performance when applied across different contexts. In this study, transferability is defined as the applicability of the developed methods across various ecological contexts, shaped by factors such as climate, culture, and socio-economic conditions. These factors can influence, for example, (i) the availability, type, and physical characteristics of mosquito breeding containers, including variations in size, material, form, and color, or (ii) the availability and use of mobile phones, contributions to crowdsourced platforms, or social media data, which may vary across locations.

One example is the water tanks discussed in publication I, which serve as common breeding sites for *Aedes* mosquitoes in several Latin American cities where piped water systems are often unreliable or unavailable. However, in other *Aedes*-endemic regions, such water tanks may differ in appearance or be entirely lacking, regardless of the piped water systems structure. This variation poses challenges to the transferability of the trained water tank detection models for suitability mapping of *Aedes* mosquitoes. Socio-economic factors can also affect ecological contexts. For instance, higher education levels may promote more proactive efforts to eliminate common breeding sites, potentially reducing *Aedes* habitat availability in certain areas. Such differences were also observed within the case study region of Rio de Janeiro (cf. publication I). Additionally, cultural practices contribute to variations in ecological contexts, such as using plastic bottles as fishing buoys, repurposing discarded tires as planters, or decorating gardens with colorful glass bottles, all of which can create unique, localized conditions.

To reliably transfer the proposed methods across such socio-economic and cultural gradients, thorough evaluation of these concepts is essential before drawing inferences. In the long term, ecological differences and speciation barriers may also drive evolutionary processes, leading to ecological or even genetic adaptations in local mosquito populations.

To enhance the transferability of the proposed methods across diverse ecological contexts, future research should focus on two key aspects: (i) acquiring detailed knowledge of local environmental and socio-cultural conditions, and (ii) adapting and recalibrating suitability indicators to account for these variations. This approach would help ensure that the models remain accurate and relevant in diverse contexts. Transferability to other mosquito species would require a more extensive recalibration of the models, such as developing new hypotheses to derive species-specific suitability indicators. Geospatial big data (GBD) has already been widely used to study other mosquito species, such as the malaria-transmitting *Anopheles* mosquito. Therefore, this research can be seen as an extension of transferability studies in the field, recalibrating existing geoinformatics methods to better account for the specific ecological characteristics of *Aedes aegypti*, such as its limited flight range, breeding habitats, and daytime biting behavior.

The overall thesis objective of harnessing GBD to guide local interventions for *Aedes*-borne arboviral infections can be seen as a transferable concept, as it can be adapted to various ecological contexts and is not dependent on specific data sources. The PCA applied in this study was chosen due to the high collinearity between the derived suitability indicators, highlighting that many of the GBD sources can represent similar environmental features, albeit at different scales due to spatiotemporal biases, as discussed in the previous subsection on scalability. Future research could further assess the robustness of the proposed concepts, particularly in their ability to explain ovitrap counts across diverse ecological contexts and GBD sources. Another potential task could be the recalibration of the models to fit the different stages of the *Aedes* life cycle or other entomological indicators relevant for local intervention guidance.

The concept of using mobile phone and social media data to model *Aedes*-human interactions is partially transferable. In our increasingly globalized world, mobile phone and internet usage patterns are becoming more consistent across regions, creating comparable contexts. However, local variations in data providers and spatiotemporal resolution persist, which affect the scalability of the proposed methods more than their transferability, as discussed in the subsection on scalability.

## **Towards reproducibility and replicability**

Following the guidelines established by the Committee on Reproducibility and Replicability in Science (National Academy of Sciences, 2019), we conclude that the workflows presented in this thesis demonstrate experimental reproducibility (R1), meaning that an independent research team can achieve the same results using the same implementation on the same data. However, achieving method reproducibility (R2), where an independent research team can replicate the results using different methods on the same data, and replicability (R3), where consistent results are obtained by applying similar methods to new data to address the same research question, will require further investigation.

While access to the original data is essential for ensuring experimental reproducibility (R1) (Wilkinson et al., 2016), it was not always feasible to share some of the original datasets outside the research team. The health records, entomological surveillance data, and the mobile phone records we used are restricted and cannot be made publicly available, while the mobile phone dataset was additionally too large for inclusion in a data

repository. In these cases, we have provided details on the data sources, including the entities from which the data were obtained, version numbers or access dates, and other relevant information to aid experimental reproducibility (R1), recognizing that many datasets are subject to continuous updates. Restrictions on data sharing exist primarily due to ethical considerations related to data privacy, confidentiality, and security. Sensitive information, such as residential addresses, could potentially identify individuals, even in anonymized data files, when combined with other datasets (Crigger and Khoury, 2019).

Regarding method reproducibility (R2), future research should investigate how different definitions of spatiotemporal representations or dependencies may influence the results. In this thesis, we have provided a rationale for our methodological choices in conceptualizing spatiotemporal space, including the use of train-test splits, the selection of specific sizes for simulated *Aedes* flight ranges, the preference for field-based rather than object-based representations, the adoption of raster-based models over kernel-based encoders, and the decision to model urban dengue occurrence at an areal level despite the Modifiable Areal Unit Problem (MAUP) instead of using point-based methods. However, these decisions are founded on assumptions that may not fully align with biological realities. Thus, future research should (i) explore the sensitivity of the findings to these assumptions, as demonstrated in the analysis in Table A. 2 of publication II, and (ii) examine how varying spatiotemporal representations may affect the conclusions drawn from this thesis. Assessing the replicability (R3) of the results across different spatial and temporal contexts remains a further important area for future research, especially given the potential influence of sampling bias. Future studies should investigate the robustness of the findings presented in this thesis across various spatiotemporal scenarios to ensure their validity and reliability in diverse settings. Additionally, spatiotemporal relationships may vary significantly across locations and over time. For example, the elevation threshold above which *Ae. aegypti* is no longer present can differ markedly between regions, as can the relationship between breeding container densities and entomological measurements in non-endemic regions. Consequently, applying the methods developed in this thesis to new contexts, or to different mosquito species, without first demonstrating replicability across space, time, and potentially species may not be advisable.

### Promising Technologies

Promising technologies to further advance the field of hybrid intervention guidance beyond the proposed techniques include, but are not limited to (i) Geospatial artificial intelligence (GeoAI) methods, (ii) molecular and biochemical techniques for identifying pathogen presence in vector populations, and (iii) smart traps equipped with bioacoustics sensors.

(i) GeoAI methods represent an interesting advancement in leveraging GBD. Unlike traditional approaches, GeoAI combines geographical knowledge with artificial intelligence to analyze complex spatiotemporal data patterns. The methodological advantage of GeoAI lies in its ability to process and learn from large volumes of geospatial data in near-real time, using ML algorithms to identify subtle spatial and temporal trends that are often overlooked by conventional analysis. As GBD becomes more widely available, GeoAI's adaptive learning capabilities allow for the continuous refinement of predictive models, enhancing both their reliability and applicability. The emergence of foundation models - large-scale models pretrained on vast and diverse datasets - holds particular promise in this context by enabling transfer learning and reducing the need for large amounts of domain-specific training data. They can provide a robust starting point for various GeoAI tasks in our research field, such as *Ae. aegypti* breeding container

detection, thereby accelerating the development of sophisticated models for guiding local interventions. While GeoAI facilitates large-scale learning and feature extraction, it should not replace mechanistic models and hypothesis-driven approaches. Although data-driven learning can generate a richer set of features, it may offer less explainability and could be influenced by data-intrinsic spatiotemporal biases, potentially leading to spurious correlations - especially problematic for inferential analysis. Therefore, the development of explainable GeoAI methods tailored to guide local interventions for *Aedes*-borne arboviral infections could be a key focus for future research. This would help balance the strengths of data-driven approaches with the need for transparency in public health decision-making.

(ii) Molecular and biochemical techniques offer another promising technology for enhancing intervention guidance. These techniques allow for the precise identification of pathogen presence within vector populations, providing critical insights into the dynamics of *Aedes*-borne arboviral disease transmission. Understanding factors such as pathogen penetration rates within mosquito and human populations can help in assessing the risk of disease spread and identifying hotspots for targeted interventions. Although these advanced epidemiological studies can yield valuable data, they are often costly and challenging to conduct regularly at high spatial resolution. Thus, while molecular and biochemical approaches can complement geospatial analyses by providing a deeper understanding of disease dynamics, their application may be best suited for high-priority areas where resources are concentrated.

(iii) Smart traps and bioacoustics represent cutting-edge technologies for in-field data collection, offering near real-time monitoring of mosquito populations. These tools improve the effectiveness of hybrid intervention approaches by enhancing the temporal resolution of entomological surveillance and reducing the potential for human error in entomological surveillance. While still in the early stages of development, smart traps can address some of the limitations of existing GBD sources, such as openly available satellite imagery, which may lack the temporal precision needed for dynamic intervention guidance due to infrequent data updates. Additionally, integrating indoor entomological surveillance, whether through smart traps or traditional methods such as ovitraps, can provide more comprehensive data, further refining mapping estimates and intervention strategies. However, the high costs associated with deploying smart traps highlight the need for targeted research on more cost-effective solutions.

For these promising technologies to be implemented in real-world settings, it is essential to (i) foster acceptance of digital innovations among public health stakeholders through scientific validation, such as that provided in this thesis, and (ii) secure adequate funding to scale these digital tools into production-ready decision support systems. An illustrative example in this context is the is GPS technology. Although GPS has the potential to model human movement flows with high precision, the limited availability of GPS data for scientific research significantly restricts its utility. This example underscores that the mere existence of advanced technologies is not enough to drive progress; successful application of these technologies for intervention guidance also requires broad access to the data they generate, along with targeted funding and active engagement from stakeholders.

Addressing these future directions will be crucial for enhancing the scalability, transferability, and replicability of hybrid intervention guidance, motivating the harnessing of emerging technologies and GBD sources to mitigate the spread of *Aedes*-borne arboviral infections in the future. Integrating future research with the proposed concepts can lead to the development of a more robust, reliable, and adaptable framework for guiding local interventions across diverse contexts, with potential applicability to other vector-borne diseases.

### Reflection on Limitations and Action Points

1. **Data Limitations:** The effectiveness of hybrid intervention guidance is contingent upon the availability and quality of GBD. Inconsistent data collection methods and gaps in spatial or temporal coverage limit the accuracy of the proposed models and intervention guidance.
  - Invest in standardized, high-resolution geospatial and temporal data infrastructure to enhance data collection methods and reduce gaps.
  - Foster partnerships between government agencies, research institutions, and local organizations to improve data accessibility and sharing.
  - Utilize remote sensing technology and engage communities in data collection through citizen science initiatives to expand coverage in underserved areas.
2. **Ecological Variability:** Complex interactions between ecological factors and vector populations vary significantly across regions, impacting the generalizability of proposed strategies and necessitating localized adaptations.
  - Develop region-specific adaptation models to fine-tune vector control strategies based on local ecological conditions.
  - Initiate pilot studies in different ecological regions to evaluate the effectiveness of proposed strategies and make necessary adjustments.
  - Promote collaboration between regions with similar ecological conditions to share best practices and enhance adaptability.
3. **Implementation Challenges:** Successful implementation may be hindered by socio-economic factors, including resource availability, political will, and community acceptance.
  - Involve local stakeholders, including community leaders and health practitioners, in project design to ensure alignment with community needs.
  - Provide training, financial resources, and tools to local agencies for sustainable implementation of GBD-informed interventions.
  - Adapt vector control measures to reflect local socio-economic realities, ensuring interventions are feasible and acceptable to communities.
4. **Dynamic Environmental Factors:** Climate change and other environmental changes can rapidly alter vector population dynamics and disease transmission.
  - Create adaptive management protocols for regular updates to intervention strategies based on changing environmental data.
  - Integrate climate projection models with GBD to anticipate shifts in vector populations and proactively adapt control measures.
  - Implement real-time monitoring systems to track environmental changes and their impacts on vector populations, enabling prompt responses to emerging threats.

## 3.6 Conclusion

### Potential of GBD in Vector Surveillance and Disease Control

This dissertation underscores the potential of GBD to advance intervention strategies for controlling *Aedes*-borne arboviral infections. By integrating sample-based entomological surveillance with digital indicators that capture *Aedes*-suitable environments and modeling daytime *Aedes*-human interactions, GBD-based approaches can generate continuous *Aedes* environmental suitability maps and identify potential pathogen transmission hotspots. These advancements can improve the guidance for interventions, with particular relevance to those that are resource-intensive or subject to environmental regulations. Beyond this, the proposed concepts can optimize ovitrap placement, refine vector carrying capacity estimates to improve eco-epidemiological models (Wu et al., 2022), or empower communities through educational programs that utilize GBD to locate local vector breeding sites for elimination (Knoblauch, 2024). Although this research focuses on *Aedes*-borne arboviral infections, GBD's potential extends to the surveillance and control of other infectious diseases. Harnessing GBD can provide a flexible framework adaptable to diverse epidemiological contexts, serving as a valuable tool for a range of public health applications, such as (i) enabling early detection of outbreaks, (ii) mapping high-risk areas, (iii) optimizing resource allocation, and (iv) monitoring the effectiveness of interventions across various regions.

### Challenges and Innovative GBD Application for *Aedes* Control

Major challenges persist, especially in terms of (i) data availability at relevant spatiotemporal scales and (ii) the need for advanced data fusion techniques to address spatiotemporal biases in GBD due to inconsistent data quality. Emerging technologies related to GeoAI and smart mosquito traps offer promising solutions for controlling *Aedes*-borne arboviral infections by enhancing the temporal resolution of entomological surveillance and enabling the integration of hypothesis-driven feature engineering with data-driven learning approaches. Particularly promising in this context could be Theory-Trained Neural Networks, which integrate domain knowledge directly into the learning process. When combined with multi-modal foundation models, these networks could be leveraged to develop disease control systems that account for ecological conditions and integrate diverse GBD sources, presumably enhancing the precision and scope of spatiotemporal risk analysis for *Aedes*-borne arboviral infections in the future. Broader application of the proposed hybrid intervention guidance requires fine-tuning on GBD that spans diverse ecological and seasonal contexts to ensure robust and scalable outcomes. Additionally, given the high computational demands associated with training such data-driven systems, responsible and sustainable AI development - focused on reducing energy consumption - should be prioritized. However, the effectiveness of such models will remain limited as long as paired datasets, which are necessary for capturing the complex relationships between ecological and epidemiological variables, continue to be insufficient.

### Towards Sustainable *Aedes* Control

Beyond supporting public health authorities with intervention guidance, GBD has the potential to inform urban planning by guiding infrastructure design to mitigate disease transmission in high-risk areas. Traditionally, many vector control strategies, such as the use of larvicides like BTI and temephos in water tanks, focus on directly targeting the vector but do not address the underlying structural causes of the problem. Rethinking these approaches through interdisciplinary and innovative solutions,

similar to the proposed concepts in this research, can lead to more sustainable outcomes. Implementing water pipe systems in underserved regions can eliminate the need for water tanks, addressing a key driver of *Aedes* breeding in many areas. Additional GBD-informed urban design strategies include (i) engineering trash bins with fine grid bottoms to prevent water accumulation, (ii) enhancing solid waste management in critical areas, and (iii) enforcing regulations to prohibit water-holding containers in cemeteries. Regular educational outreach to local stakeholders, such as tire repair shops, can further reinforce preventive efforts by promoting best practices for vector control. Nature-based solutions also offer sustainable alternatives to conventional vector control. These strategies include planting mosquito-repellent plants (e.g., citronella, lemongrass), introducing carnivorous plants (e.g., pitcher plants) in specific zones, and fostering natural mosquito predators (e.g., dragonflies, birds) through habitat creation. While these methods reduce reliance on chemical controls, careful management is essential to maintain ecosystem balance and avoid unintended impacts on human and environmental health. In conjunction with the proposed concept of hybrid intervention guidance, these strategies represent a reimagined, holistic approach to *Aedes*-borne arboviral control, emphasizing long-term, structural solutions over short-term fixes. This is particularly crucial in developing regions, where healthcare systems are often overstretched. By integrating proactive infrastructure improvements with community-driven empowerment, these concepts can presumably improve public health outcomes, enhance societal resilience to epidemics, and reduce the long-term social and economic burdens associated with vector-borne diseases.

### **GBD in Action: From Vector Control to Environmental Resilience**

On a global scale, the use of GBD in controlling *Aedes*-borne diseases aligns with broader global health and environmental challenges, particularly in adapting to climate change. As rising temperatures and shifting weather patterns alter the geographical distribution of mosquito populations, GBD can support the monitoring of vector population redistributions and predictive modeling, enabling countries to proactively adjust their disease control strategies to respond to these evolving ecological dynamics. Extending the proposed concepts from *Aedes*-endemic regions to newly emerging *Aedes* suitability areas can foster global collaboration, strengthening transnational efforts to mitigate the spread of diseases like dengue, Zika, and chikungunya, which pose escalating public health threats. Beyond vector surveillance, this dissertation underscores GBD's potential as a transformative data source for a wider range of public health, environmental, and urban planning challenges. As the volume of collected GBD is projected to increase, coupled with advancements in analytical tools and interdisciplinary collaboration, GBD is poised to enhance our understanding of complex interactions between built-up environments, ecosystems, and human health. This integration offers a powerful resource for addressing multifaceted spatial and environmental challenges, from urban resilience to ecosystem sustainability.

### Take home messages

1. Harnessing GBD for integration with entomological surveillance allows a broader and more precise identification of mosquito breeding hotspots and *Aedes*-human interactions, aiding in strategically placing entomological surveillance and enhancing guidance for local interventions for *Aedes*-borne arboviral infections. High-resolution indicators play a critical role in improving the accuracy of these models.
2. Approximating residents' exposure to *Aedes* populations by integrating daytime human movement and with knowledge on diurnal *Aedes* biting behavior can enhance predictive accuracy of models estimating the spatial occurrence of dengue cases, compared to those that omit these factors.
3. Scalability and transferability remain critical challenges to the broader applicability of GBD-enriched intervention guidance. Effectively addressing these challenges requires the implementation of adaptive approaches that account for variations in data quality and ecological contexts.
4. The potential of harnessing GBD extends beyond public health to urban planning, biodiversity monitoring, and assessing ecological impacts of urbanization, offering valuable insights for enhancing social well-being, reducing economic burdens, and supporting sustainable development.

### Actionable Recommendations

1. **Enhance Data Collection Efforts:** Public health authorities should invest in robust data collection and sharing systems to ensure high-quality, paired datasets. This includes deploying remote sensing technologies and involving local communities in citizen science initiatives to monitor mosquito populations and breeding sites.
2. **Strengthen Community Engagement:** Implement educational programs that empower local communities to participate in vector control efforts. Engaging stakeholders such as schools and local organizations can foster awareness and responsibility regarding mosquito breeding sites and preventive measures.
3. **Promote Integrated Urban Planning:** Urban planners should incorporate GBD insights into infrastructure design to optimize urban environments for mosquito control. This includes improving water management, enhancing waste practices, and enforcing regulations to reduce standing water.
4. **Invest in Nature-Based Solutions:** Encourage the adoption of nature-based strategies, such as planting mosquito-repellent plants and creating habitats for natural predators. These approaches provide sustainable alternatives to chemical controls and promote ecosystem health.
5. **Support Research and Development:** Allocate funding for research into advanced data-driven models and innovative vector control technologies, including Theory-Trained Neural Networks and other machine learning techniques to enhance predictive capabilities in disease transmission.



# Bibliography

- Aguiar, M., Stollenwerk, N., and Halstead, S. B. (2016). “The risks behind Dengvaxia recommendation”. *The Lancet. Infectious diseases* 16.8, pp. 882–883. DOI: 10.1016/S1473-3099(16)30168-2.
- Albrieu-Llinás, G., Espinosa, M. O., Quaglia, A., Abril, M., and Scavuzzo, C. M. (2018). “Urban environmental clustering to assess the spatial dynamics of *Aedes aegypti* breeding sites”. *Geospatial health* 13.1, p. 654. DOI: 10.4081/gh.2018.654.
- Alphabet Inc. (2024). *TensorFlow*.
- Althouse, B. M., Scarpino, S. V., Meyers, L. A., Ayers, J. W., Bargsten, M., Baumbach, J., Brownstein, J. S., Castro, L., Clapham, H., Cummings, D. A., Del Valle, S., Eubank, S., Fairchild, G., Finelli, L., Generous, N., George, D., Harper, D. R., Hébert-Dufresne, L., Johansson, M. A., Konty, K., Lipsitch, M., Milinovich, G., Miller, J. D., Nsoesie, E. O., Olson, D. R., Paul, M., Polgreen, P. M., Priedhorsky, R., Read, J. M., Rodríguez-Barraquer, I., Smith, D. J., Stefansen, C., Swerdlow, D. L., Thompson, D., Vespignani, A., and Wesolowski, A. (2015). “Enhancing disease surveillance with novel data streams: challenges and opportunities”. *EPJ data science* 4.1. DOI: 10.1140/epjds/s13688-015-0054-0.
- Araujo, R. V., Albertini, M. R., Costa-da-Silva, A. L., Suesdek, L., Franceschi, N. C. S., Bastos, N. M., Katz, G., Cardoso, V. A., Castro, B. C., Capurro, M. L., and Allegro, V. L. A. C. (2015). “São Paulo urban heat islands have a higher incidence of dengue than other urban areas”. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases* 19.2, pp. 146–155. DOI: 10.1016/j.bjid.2014.10.004.
- Azevedo, T. S. de, Bourke, B. P., Piovezan, R., and Sallum, M. A. M. (2018). “The influence of urban heat islands and socioeconomic factors on the spatial distribution of *Aedes aegypti* larval habitats”. *Geospatial health* 13.1, p. 623. DOI: 10.4081/gh.2018.623.
- Baik, L. S., Nave, C., Au, D. D., Guda, T., Chevez, J. A., Ray, A., and Holmes, T. C. (2020). “Circadian Regulation of Light-Evoked Attraction and Avoidance Behaviors in Daytime-versus Nighttime-Biting Mosquitoes”. *Current biology: CB* 30.16, 3252–3259.e3. DOI: 10.1016/j.cub.2020.06.010.
- Barboza, M. H. C., Alencar, R. d. S., Chaves, J. C., Silva, M. A. H. B., Orrico, R. D., and Evsukoff, A. G. (2021). “Identifying Human Mobility Patterns in the Rio de Janeiro Metropolitan Area using Call Detail Records”. *Transportation Research Record: Journal of the Transportation Research Board* 2675.4, pp. 213–221. DOI: 10.1177/0361198120977655.
- Barrera, R., Amador, M., and MacKay, A. J. (2011). “Population dynamics of *Aedes aegypti* and dengue as influenced by weather and human behavior in San Juan, Puerto Rico”. *PLoS neglected tropical diseases* 5.12, e1378. DOI: 10.1371/journal.pntd.0001378.

- Bartumeus, F., Oltra, A., and Palmer, J. R. B. (2018). “Citizen Science: A Gateway for Innovation in Disease-Carrying Mosquito Management?” *Trends in parasitology* 34.9, pp. 727–729. DOI: 10.1016/j.pt.2018.04.010.
- Benitez, E. M., Ludueña-Almeida, F., Frías-Céspedes, M., Almirón, W. R., and Estallo, E. L. (2020). “Could land cover influence *Aedes aegypti* mosquito populations?” *Medical and veterinary entomology* 34.2, pp. 138–144. DOI: 10.1111/mve.12422.
- Bennett, K. L., Gómez Martínez, C., Almanza, A., Rovira, J. R., McMillan, W. O., Enriquez, V., Barraza, E., Diaz, M., Sanchez-Galan, J. E., Whiteman, A., Gittens, R. A., and Loaiza, J. R. (2019). “High infestation of invasive *Aedes* mosquitoes in used tires along the local transport network of Panama”. *Parasites & vectors* 12.1, p. 264. DOI: 10.1186/s13071-019-3522-8.
- Biljecki, F., Zhao, T., Liang, X., and Hou, Y. (2023). “Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives”. *International Journal of Applied Earth Observation and Geoinformation* 122, p. 103385. DOI: 10.1016/j.jag.2023.103385.
- Bivand, R. (2023). *Spatialreg R package documentation*.
- Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley. DOI: 10.1002/9781118950203.
- Blum, A. and Mitchell, T. (1998). “Combining labeled and unlabeled data with co-training”. *Proceedings of the eleventh annual conference on Computational learning theory - COLT' 98*. Ed. by P. Bartlett and Y. Mansour. New York, New York, USA: ACM Press, pp. 92–100. DOI: 10.1145/279943.279962.
- Boeing, G. (2021). “Spatial information and the legibility of urban form: Big data in urban morphology”. *International Journal of Information Management* 56, p. 102013. DOI: 10.1016/j.ijinfomgt.2019.09.009.
- Bonnet, E., Fournet, F., Benmarhnia, T., Ouedraogo, S., Dabiré, R., and Ridde, V. (2020). “Impact of a community-based intervention on *Aedes aegypti* and its spatial distribution in Ouagadougou, Burkina Faso”. *Infectious diseases of poverty* 9.1, p. 61. DOI: 10.1186/s40249-020-00675-6.
- Britos Molinas, M. B., Melgarejo, E. G., and Arias, A. R. de (2022). “Phenotypic Variations of *Aedes aegypti* Populations and Egg Abundance According to Environmental Parameters in Two Dengue-Endemic Ecoregions in Paraguay”. *The American journal of tropical medicine and hygiene* 107.2, pp. 300–307. DOI: 10.4269/ajtmh.21-1184.
- Buhler, C., Winkler, V., Runge-Ranzinger, S., Boyce, R., and Horstick, O. (2019). “Environmental methods for dengue vector control - A systematic review and meta-analysis”. *PLoS neglected tropical diseases* 13.7, e0007420. DOI: 10.1371/journal.pntd.0007420.
- Burke, R., Barrera, R., Lewis, M., Kluchinsky, T., and Claborn, D. (2010). “Septic tanks as larval habitats for the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus* in Playa-Playita, Puerto Rico”. *Medical and veterinary entomology* 24.2, pp. 117–123. DOI: 10.1111/j.1365-2915.2010.00864.x.
- Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y., and Wu, Z. (2019). “An Improved Faster R-CNN for Small Object Detection”. *IEEE Access* 7, pp. 106838–106846. DOI: 10.1109/ACCESS.2019.2932731.
- Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F., and Purves, R. (2016). *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press. DOI: 10.5334/bax.
- Center for Disease Control and Prevention (2024). *Factsheet Aedes aegypti life cycle*.
- Centro de Operacoes Rio (2023). *Sistema Alerta Rio da Prefeitura do Rio de Janeiro*.

- Chan, K. L., Ho, B. C., and Chan, Y. C. (1971). “Aedes aegypti (L.) and Aedes albopictus (Skuse) in Singapore City: 2. Larval habitats\*”. *Bulletin of the World Health Organization* 44.5, pp. 629–633.
- Chang, C., Ortiz, K., Ansari, A., and Gershwin, M. E. (2016). “The Zika outbreak of the 21st century”. *Journal of autoimmunity* 68, pp. 1–13. DOI: 10.1016/j.jaut.2016.02.006.
- Chang, L.-H., Hsu, E.-L., Teng, H.-J., and Ho, C.-M. (2007). “Differential Survival of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) Larvae Exposed to Low Temperatures in Taiwan”. *Journal of medical entomology* 44.2, pp. 205–210. DOI: 10.1093/jmedent/44.2.205.
- Chaves, L. F., Valerín Cordero, J. A., Delgado, G., Aguilar-Avenidaño, C., Maynes, E., Gutiérrez Alvarado, J. M., Ramírez Rojas, M., Romero, L. M., and Marín Rodríguez, R. (2021). “Modeling the association between Aedes aegypti ovitrap egg counts, multi-scale remotely sensed environmental data and arboviral cases at Puntarenas, Costa Rica (2017-2018)”. *Current research in parasitology & vector-borne diseases* 1, p. 100014. DOI: 10.1016/j.crpvbd.2021.100014.
- Chen, B., Jiang, J., Wang, X., Wan, P., Wang, J., and Long, M. (2022a). “Debiased Self-Training for Semi-Supervised Learning”. *arXiv*. DOI: 10.48550/arXiv.2202.07136.
- Chen, S., Liu, L., Chen, C., and Haase, D. (2022b). “The interaction between human demand and urban greenspace supply for promoting positive emotions with sentiment analysis from twitter”. *Urban Forestry & Urban Greening* 78, p. 127763. DOI: 10.1016/j.ufug.2022.127763.
- Chumsri, A., Jaroensutasinee, M., and Jaroensutasinee, K. (2020). “Spatial and temporal variations on the coexistence of Aedes and Culex larvae in Southern Thailand”. *Journal of Animal Behaviour and Biometeorology* 8.4, pp. 250–256. DOI: 10.31893/jabb.20033.
- CNN World (2024). *Rio declares dengue emergency as Brazil gears up for Carnival*.
- Cohen (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. DOI: 10.4324/9780203774441.
- Cornel, A. J., Holeman, J., Nieman, C. C., Lee, Y., Smith, C., Amorino, M., Brisco, K. K., Barrera, R., Lanzaro, G. C., and Mulligan Iii, F. S. (2016). “Surveillance, insecticide resistance and control of an invasive Aedes aegypti (Diptera: Culicidae) population in California”, p. 194. DOI: 10.12688/f1000research.8107.3.
- Costa, E. A. P. d. A., Santos, E. M. d. M., Correia, J. C., and Albuquerque, C. M. R. de (2010). “Impact of small variations in temperature and humidity on the reproductive activity and survival of Aedes aegypti (Diptera, Culicidae)”. *Revista Brasileira de Entomologia* 54.3, pp. 488–493. DOI: 10.1590/S0085-56262010000300021.
- Crigger, E. and Khoury, C. (2019). “Making Policy on Augmented Intelligence in Health Care”. *AMA journal of ethics* 21.2, E188–191. DOI: 10.1001/amajethics.2019.188.
- Cromwell, E. A., Stoddard, S. T., Barker, C. M., van Rie, A., Messer, W. B., Meshnick, S. R., Morrison, A. C., and Scott, T. W. (2017). “The relationship between entomological indicators of Aedes aegypti abundance and dengue virus infection”. *PLoS neglected tropical diseases* 11.3, e0005429. DOI: 10.1371/journal.pntd.0005429.
- Data.Rio (2023a). *Hydrography lines*.
- Data.Rio (2023b). *Land use land cover*.
- Demuzere, M., Kittner, J., and Bechtel, B. (2021). “LCZ Generator: A Web Application to Create Local Climate Zone Maps”. *Frontiers in Environmental Science* 9. DOI: 10.3389/fenvs.2021.637455.
- Díaz-Nieto, L. M., Chiappero, M. B., Díaz de Astarloa, C., Maciá, A., Gardenal, C. N., and Berón, C. M. (2016). “Genetic Evidence of Expansion by Passive Transport of

- Aedes (Stegomyia) aegypti in Eastern Argentina”. *PLoS neglected tropical diseases* 10.9, e0004839. DOI: 10.1371/journal.pntd.0004839.
- Dolley, S. (2018). “Big Data’s Role in Precision Public Health”. *Frontiers in public health* 6, p. 68. DOI: 10.3389/fpubh.2018.00068.
- Egid, B. R., Coulibaly, M., Dadzie, S. K., Kamgang, B., McCall, P. J., Sedda, L., Toe, K. H., and Wilson, A. L. (2022a). “Review of the ecology and behaviour of Aedes aegypti and Aedes albopictus in Western Africa and implications for vector control”. *Current research in parasitology & vector-borne diseases* 2, p. 100074. DOI: 10.1016/j.crpvbd.2021.100074.
- Egid, B. R., Coulibaly, M., Dadzie, S. K., Kamgang, B., McCall, P. J., Sedda, L., Toe, K. H., and Wilson, A. L. (2022b). “Review of the ecology and behaviour of Aedes aegypti and Aedes albopictus in Western Africa and implications for vector control”. *Current research in parasitology & vector-borne diseases* 2, p. 100074. DOI: 10.1016/j.crpvbd.2021.100074.
- Equihua, M., Ibáñez-Bernal, S., Benítez, G., Estrada-Contreras, I., Sandoval-Ruiz, C. A., and Mendoza-Palmero, F. S. (2017). “Establishment of Aedes aegypti (L.) in mountainous regions in Mexico: Increasing number of population at risk of mosquito-borne disease and future climate conditions”. *Acta tropica* 166, pp. 316–327. DOI: 10.1016/j.actatropica.2016.11.014.
- Eritja, R., Palmer, J. R. B., Roiz, D., Sanpera-Calbet, I., and Bartumeus, F. (2017). “Direct Evidence of Adult Aedes albopictus Dispersal by Car”. *Scientific reports* 7.1, p. 14399. DOI: 10.1038/s41598-017-12652-5.
- Espinosa, M. O., Polop, F., Rotela, C. H., Abril, M., and Scavuzzo, C. M. (2016). “Spatial pattern evolution of Aedes aegypti breeding sites in an Argentinean city without a dengue vector control programme”. *Geospatial health* 11.3, p. 471. DOI: 10.4081/gh.2016.471.
- Estallo, E. L., Sangermano, F., Grech, M., Ludueña-Almeida, F., Frías-Cespedes, M., Ainete, M., Almirón, W., and Livdahl, T. (2018). “Modelling the distribution of the vector Aedes aegypti in a central Argentine city”. *Medical and veterinary entomology* 32.4, pp. 451–461. DOI: 10.1111/mve.12323.
- Estallo, E. L., Lamfri, M. A., Scavuzzo, C. M., Almeida, F. F. L., Introini, M. V., Zaidenberg, M., and Almirón, W. R. (2008). “Models for predicting Aedes aegypti larval indices based on satellite images and climatic variables”. *Journal of the American Mosquito Control Association* 24.3, pp. 368–376. DOI: 10.2987/5705.1.
- Estallo, E. L., Ludueña-Almeida, F. F., Visintin, A. M., Scavuzzo, C. M., Lamfri, M. A., Introini, M. V., Zaidenberg, M., and Almirón, W. R. (2012). “Effectiveness of normalized difference water index in modelling Aedes aegypti house index”. *International Journal of Remote Sensing* 33.13, pp. 4254–4265. DOI: 10.1080/01431161.2011.640962.
- Finger, F., Genolet, T., Mari, L., Magny, G. C. de, Manga, N. M., Rinaldo, A., and Bertuzzo, E. (2016). “Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks”. *Proceedings of the National Academy of Sciences of the United States of America* 113.23, pp. 6421–6426. DOI: 10.1073/pnas.1522305113.
- Fleischmann, M. (2019). “momepy: Urban Morphology Measuring Toolkit”. *Journal of Open Source Software* 4.43, p. 1807. DOI: 10.21105/joss.01807.
- Flores, H. A. and O’Neill, S. L. (2018). “Controlling vector-borne diseases by releasing modified mosquitoes”. *Nature reviews. Microbiology* 16.8, pp. 508–518. DOI: 10.1038/s41579-018-0025-0.
- Gao, S., Hu, Y., and Li, W., eds. (2024). *Handbook of geospatial artificial intelligence*. First edition. Boca Raton, FL: CRC Press.

- German, A., Espinosa, M. O., Abril, M., and Scavuzzo, C. M. (2018). “Exploring satellite based temporal forecast modelling of *Aedes aegypti* oviposition from an operational perspective”. *Remote Sensing Applications: Society and Environment* 11, pp. 231–240. DOI: 10.1016/j.rsase.2018.07.006.
- Getis, A., Morrison, A. C., Gray, K., and Scott, T. W. (2003). “Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru”. *The American journal of tropical medicine and hygiene* 69.5, pp. 494–505.
- Gomes, H., Jesus, A. G. de, and Quaresma, J. A. S. (2023). “Identification of risk areas for arboviruses transmitted by *Aedes aegypti* in northern Brazil: A One Health analysis”. *One health (Amsterdam, Netherlands)* 16, p. 100499. DOI: 10.1016/j.onehlt.2023.100499.
- Google LLC (2023). *Street View Static API*.
- Griffith, D. A. (2000). “Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses”. *Linear Algebra and its Applications* 321.1-3, pp. 95–112. DOI: 10.1016/S0024-3795(00)00031-8.
- Griffith, D. A. (2019). *Spatial regression analysis using Eigenvector spatial filtering*. S.I.: Academic Press.
- Guagliardo, S. A., Morrison, A. C., Barboza, J. L., Requena, E., Astete, H., Vazquez-Prokopec, G., and Kitron, U. (2015). “River boats contribute to the regional spread of the dengue vector *Aedes aegypti* in the Peruvian Amazon”. *PLoS neglected tropical diseases* 9.4, e0003648. DOI: 10.1371/journal.pntd.0003648.
- Guo, C., Zhou, Z., Wen, Z., Liu, Y., Zeng, C., Di Xiao, Ou, M., Han, Y., Huang, S., Liu, D., Ye, X., Zou, X., Wu, J., Wang, H., Zeng, E. Y., Jing, C., and Yang, G. (2017). “Global Epidemiology of Dengue Outbreaks in 1990-2015: A Systematic Review and Meta-Analysis”. *Frontiers in cellular and infection microbiology* 7, p. 317. DOI: 10.3389/fcimb.2017.00317.
- Guzman, M. G., Gubler, D. J., Izquierdo, A., Martinez, E., and Halstead, S. B. (2016). “Dengue infection”. *Nature reviews. Disease primers* 2, p. 16055. DOI: 10.1038/nrdp.2016.55.
- Guzman, M. G. and Harris, E. (2015). “Dengue”. *Lancet (London, England)* 385.9966, pp. 453–465. DOI: 10.1016/S0140-6736(14)60572-9.
- Haddawy, P., Wettayakorn, P., Nonthaleerak, B., Su Yin, M., Wiratsudakul, A., Schöning, J., Laosiritaworn, Y., Balla, K., Euaungkanakul, S., Quengdaeng, P., Choknitipakin, K., Traivijitkhun, S., Erawan, B., and Kraisang, T. (2019). “Large scale detailed mapping of dengue vector breeding sites using street view images”. *PLoS neglected tropical diseases* 13.7, e0007555. DOI: 10.1371/journal.pntd.0007555.
- Halstead, S. B. (2007). “Dengue”. *Lancet (London, England)* 370.9599, pp. 1644–1652. DOI: 10.1016/S0140-6736(07)61687-0.
- Harbach, R. (2007). “The Culicidae (Diptera): a review of taxonomy, classification and phylogeny\*”. *Zootaxa* 1668.1. DOI: 10.11646/zootaxa.1668.1.28.
- Hilbe, J. M. (2012). *Negative Binomial Regression*. Cambridge University Press. DOI: 10.1017/CB09780511973420.
- Honório, N. A., Da Silva, W. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003). “Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil”. *Memorias do Instituto Oswaldo Cruz* 98.2, pp. 191–198. DOI: 10.1590/s0074-02762003000200005.
- Horstick, O., Runge-Ranzinger, S., Nathan, M. B., and Kroeger, A. (2010). “Dengue vector-control services: how do they work? A systematic literature review and country case studies”. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 104.6, pp. 379–386. DOI: 10.1016/j.trstmh.2009.07.027.

- Hossain, M. S., Raihan, M. E., Hossain, M. S., Syeed, M. M. M., Rashid, H., and Reza, M. S. (2022). “Aedes Larva Detection Using Ensemble Learning to Prevent Dengue Endemic”. *BioMedInformatics* 2.3, pp. 405–423. DOI: 10.3390/biomedinformatics2030026.
- Hou, Y. and Biljecki, F. (2022). “A comprehensive framework for evaluating the quality of street view imagery”. *International Journal of Applied Earth Observation and Geoinformation* 115, p. 103094. DOI: 10.1016/j.jag.2022.103094.
- Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., Wang, Z., Zhao, T., and Biljecki, F. (2024). “Global Streetscapes — A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics”. *ISPRS Journal of Photogrammetry and Remote Sensing* 215, pp. 216–238. DOI: 10.1016/j.isprsjprs.2024.06.023.
- Hu, X., Hu, Y., Resch, B., and Kersten, J. (2023). “Geographic Information Extraction from Texts (GeoExT)”. *Advances in Information Retrieval*. Ed. by J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo. Vol. 13982. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 398–404. DOI: 10.1007/978-3-031-28241-6{\textunderscore}44.
- Huang, B. and Carley, K. M. (2019). “A large-scale empirical study of geotagging behavior on Twitter”. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Ed. by F. Spezzano, W. Chen, and X. Xiao. New York, NY, USA: ACM, pp. 365–373. DOI: 10.1145/3341161.3342870.
- Huang, Q. and Wong, D. W. S. (2015). “Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data”. *Annals of the Association of American Geographers* 105.6, pp. 1179–1197. DOI: 10.1080/00045608.2015.1081120.
- ICREA (2024). *Mosquito Alert*.
- Iggidr, A., Koiller, J., Penna, M., Sallet, G., Silva, M. A., and Souza, M. O. (2017). “Vector borne diseases on an urban environment: The effects of heterogeneity and human circulation”. *Ecological Complexity* 30, pp. 76–90. DOI: 10.1016/j.ecocom.2016.12.006.
- Instituto Brasileiro de Geografia e Estatística (2024). *Censo Demográfico 2022*.
- Ito, K., Kang, Y., Zhang, Y., Zhang, F., and Biljecki, F. (2024). “Understanding urban perception with visual data: A systematic review”. *Cities* 152, p. 105169. DOI: 10.1016/j.cities.2024.105169.
- Jennex, M., Ceccaroni, L., and Piera, J. (2017). *Analyzing the Role of Citizen Science in Modern Research*. IGI Global. DOI: 10.4018/978-1-5225-0962-2.
- Jesús Crespo, R. de and Rogers, R. E. (2021). “Habitat Segregation Patterns of Container Breeding Mosquitos: The Role of Urban Heat Islands, Vegetation Cover, and Income Disparity in Cemeteries of New Orleans”. *International journal of environmental research and public health* 19.1. DOI: 10.3390/ijerph19010245.
- Južnič-Zonta, Ž., Sanpera-Calbet, I., Eritja, R., Palmer, J. R. B., Escobar, A., Garriga, J., Oltra, A., Richter-Boix, A., Schaffner, F., Della Torre, A., Miranda, M. Á., Koopmans, M., Barzon, L., and Bartumeus Ferre, F. (2022). “Mosquito alert: leveraging citizen science to create a GBIF mosquito occurrence dataset”. *GigaByte (Hong Kong, China)* 2022, gigabyte54. DOI: 10.46471/gigabyte.54.
- Kache, P. A., Santos-Vega, M., Stewart-Ibarra, A. M., Cook, E. M., Seto, K. C., and Diuk-Wasser, M. A. (2022a). “Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases”. *Nature ecology & evolution* 6.11, pp. 1601–1616. DOI: 10.1038/s41559-022-01876-y.

- Kache, P. A., Santos-Vega, M., Stewart-Ibarra, A. M., Cook, E. M., Seto, K. C., and Diuk-Wasser, M. A. (2022b). “Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases”. *Nature ecology & evolution* 6.11, pp. 1601–1616. DOI: 10.1038/s41559-022-01876-y.
- Kamgang, B., Happi, J. Y., Boisier, P., Njiokou, F., Hervé, J.-P., Simard, F., and Paupy, C. (2010). “Geographic and ecological distribution of the dengue and chikungunya virus vectors *Aedes aegypti* and *Aedes albopictus* in three major Cameroonian towns”. *Medical and veterinary entomology* 24.2, pp. 132–141. DOI: 10.1111/j.1365-2915.2010.00869.x.
- Kaplan, L., Kendell, D., Robertson, D., Livdahl, T., and Khatchikian, C. (2010). “*Aedes aegypti* and *Aedes albopictus* in Bermuda: extinction, invasion, invasion and extinction”. *Biological invasions* 12.9, pp. 3277–3288. DOI: 10.1007/s10530-010-9721-z.
- KartaView (2023). *OpenStreetCam API Documentation*.
- Katzelnick, L. C., Fonville, J. M., Gromowski, G. D., Bustos Arriaga, J., Green, A., James, S. L., Lau, L., Montoya, M., Wang, C., VanBlargan, L. A., Russell, C. A., Thu, H. M., Pierson, T. C., Buchy, P., Aaskov, J. G., Muñoz-Jordán, J. L., Vasilakis, N., Gibbons, R. V., Tesh, R. B., Osterhaus, A. D. M. E., Fouchier, R. A. M., Durbin, A., Simmons, C. P., Holmes, E. C., Harris, E., Whitehead, S. S., and Smith, D. J. (2015). “Dengue viruses cluster antigenically but not as discrete serotypes”. *Science (New York, N.Y.)* 349.6254, pp. 1338–1343. DOI: 10.1126/science.aac5017.
- Killeen, G. F., Chaki, P. P., Reed, T. E., Moyes, C. L., and Govella, N. J. (2018). “Entomological surveillance as a cornerstone of malaria elimination: a critical appraisal”. *Towards malaria elimination—a leap forward*. Manguin S, Vas D, Eds. *IntechOpen*, pp. 403–429.
- Knoblauch, S. and Gross, S. (2023). *GitHub repository for this manuscript called Long-term validation of inner-urban mobility metrics derived from Twitter*. DOI: 10.5281/zenodo.8304597.
- Knoblauch, S. (2024). *Bekämpfung von Mücken mittels Satellitenbildern und Street View-Daten*. URL: <https://www.ireso.org/bekaempfung-von-muecken-mittels-satellitenbilder-und-street-view-daten/>.
- Knoblauch, S., Li, H., Lautenbach, S., Elshiaty, Y., Rocha, A. A. d. A., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2023). “Semi-supervised water tank detection to support vector control of emerging infectious diseases transmitted by *Aedes Aegypti*”. *International Journal of Applied Earth Observation and Geoinformation* 119, p. 103304. DOI: 10.1016/j.jag.2023.103304.
- Knoblauch, S., Su Yin, M., Chatrinan, K., Rocha, A. A. d. A., Haddawy, P., Biljecki, F., Lautenbach, S., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2024). “High-resolution mapping of urban *Aedes aegypti* immature abundance through breeding site detection based on satellite and street view imagery (under review)”. *Scientific Reports: Collection on Urbanization and communicable diseases*. DOI: 10.1038/s41598-024-67914-w.
- Kogan, N. E., Clemente, L., Liautaud, P., Kaashoek, J., Link, N. B., Nguyen, A. T., Lu, F. S., Huybers, P., Resch, B., Havas, C., Petutschnig, A., Davis, J., Chinazzi, M., Mustafa, B., Hanage, W. P., Vespignani, A., and Santillana, M. (2021). “An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time”. *Science advances* 7.10. DOI: 10.1126/sciadv.abd6989.
- Kraemer, M. U. G., Bisanzio, D., Reiner, R. C., Zakar, R., Hawkins, J. B., Freifeld, C. C., Smith, D. L., Hay, S. I., Brownstein, J. S., and Perkins, T. A. (2018). “Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from

- Lahore, Pakistan". *EPJ data science* 7.1, p. 16. DOI: 10.1140/epjds/s13688-018-0144-x.
- Kraemer, M. U. G., Hay, S. I., Pigott, D. M., Smith, D. L., Wint, G. R. W., and Golding, N. (2016). "Progress and Challenges in Infectious Disease Cartography". *Trends in parasitology* 32.1, pp. 19–29. DOI: 10.1016/j.pt.2015.09.006.
- Kraemer, M. U. G., Reiner, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D. M., Yi, D., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Bisanzio, D., Perkins, T. A., Lai, S., Lu, X., Jones, P., Coelho, G. E., Carvalho, R. G., van Bortel, W., Marsboom, C., Hendrickx, G., Schaffner, F., Moore, C. G., Nax, H. H., Bengtsson, L., Wetter, E., Tatem, A. J., Brownstein, J. S., Smith, D. L., Lambrechts, L., Cauchemez, S., Linard, C., Faria, N. R., Pybus, O. G., Scott, T. W., Liu, Q., Yu, H., Wint, G. R. W., Hay, S. I., and Golding, N. (2019). "Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*". *Nature microbiology* 4.5, pp. 854–863. DOI: 10.1038/s41564-019-0376-y.
- Kwan, M.-p. and Schwanen, T., eds. (2017). *Geographies of mobility: Recent advances in theory and method*. London and New York, New York: Routledge.
- Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., and Scott, T. W. (2011). "Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*". *Proceedings of the National Academy of Sciences of the United States of America* 108.18, pp. 7460–7465. DOI: 10.1073/pnas.1101377108.
- Landau, K. I. and van Leeuwen, W. J. D. (2012). "Fine scale spatial urban land cover factors associated with adult mosquito abundance and risk in Tucson, Arizona". *Journal of vector ecology : journal of the Society for Vector Ecology* 37.2, pp. 407–418. DOI: 10.1111/j.1948-7134.2012.00245.x.
- Lega, J., Brown, H. E., and Barrera, R. (2017). "Aedes aegypti (Diptera: Culicidae) Abundance Model Improved With Relative Humidity and Precipitation-Driven Egg Hatching". *Journal of medical entomology* 54.5, pp. 1375–1384. DOI: 10.1093/jme/tjx077.
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frías-Martínez, E., and Ramasco, J. J. (2014). "Cross-checking different sources of mobility information". *PloS one* 9.8, e105184. DOI: 10.1371/journal.pone.0105184.
- Leta, S., Beyene, T. J., Clercq, E. M. de, Amenu, K., Kraemer, M. U. G., and Revie, C. W. (2018). "Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*". *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 67, pp. 25–35. DOI: 10.1016/j.ijid.2017.11.026.
- Levin, S. A., ed. (2013). *Encyclopedia of biodiversity*. 2nd ed. Amsterdam: Academic Press.
- Li, C. F., Lim, T. W., Han, L. L., and Fang, R. (1985). "Rainfall, abundance of *Aedes aegypti* and dengue infection in Selangor, Malaysia". *The Southeast Asian journal of tropical medicine and public health* 16.4, pp. 560–568.
- Li, L., Goodchild, M. F., and Xu, B. (2013). "Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr". *Cartography and geographic information science* 40.2, pp. 61–77. DOI: 10.1080/15230406.2013.777139.
- Li, Z. and Dong, J. (2022). "Big Geospatial Data and Data-Driven Methods for Urban Dengue Risk Forecasting: A Review". *Remote Sensing* 14.19, p. 5052. DOI: 10.3390/rs14195052.

- Liew, C. and Curtis, C. F. (2004). “Horizontal and vertical dispersal of dengue vector mosquitoes, *Aedes aegypti* and *Aedes albopictus*, in Singapore”. *Medical and veterinary entomology* 18.4, pp. 351–360. DOI: 10.1111/j.0269-283X.2004.00517.x.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). “Microsoft COCO: Common Objects in Context”. DOI: 10.48550/arXiv.1405.0312.
- Lindgren, F., Rue, H., and Lindström, J. (2011). “An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73.4, pp. 423–498. DOI: 10.1111/j.1467-9868.2011.00777.x.
- Lindsay, S. W., Wilson, A., Golding, N., Scott, T. W., and Takken, W. (2017). “Improving the built environment in urban areas to control *Aedes aegypti*-borne diseases”. *Bulletin of the World Health Organization* 95.8, pp. 607–608. DOI: 10.2471/BLT.16.189688.
- Liu, W.-L., Wang, Y., Chen, Y.-X., Chen, B.-Y., Lin, A. Y.-C., Dai, S.-T., Chen, C.-H., and Liao, L.-D. (2023). “An IoT-based smart mosquito trap system embedded with real-time mosquito image processing by neural networks for mosquito surveillance”. *Frontiers in bioengineering and biotechnology* 11, p. 1100968. DOI: 10.3389/fbioe.2023.1100968.
- Liu-Helmersson, J., Brännström, Å., Sewe, M. O., Semenza, J. C., and Rocklöv, J. (2019). “Estimating Past, Present, and Future Trends in the Global Distribution and Abundance of the Arbovirus Vector *Aedes aegypti* Under Climate Change Scenarios”. *Frontiers in public health* 7, p. 148. DOI: 10.3389/fpubh.2019.00148.
- López-García, D. (2023). *Worker mobility and urban policy in latin america: Policy interactions and urban Outcomes in mexico city*. 1 Edition. Routledge advances in regional economics, science and policy. New York, NY: Routledge.
- Lorenz, C., Castro, M. C., Trindade, P. M. P., Nogueira, M. L., Oliveira Lage, M. de, Quintanilha, J. A., Parra, M. C., Dibo, M. R., Fávares, E. A., Guirado, M. M., and Chiaravalloti-Neto, F. (2020a). “Predicting *Aedes aegypti* infestation using landscape and thermal features”. *Scientific reports* 10.1, p. 21688. DOI: 10.1038/s41598-020-78755-8.
- Lorenz, C., Chiaravalloti-Neto, F., Oliveira Lage, M. de, Quintanilha, J. A., Parra, M. C., Dibo, M. R., Fávares, E. A., Guirado, M. M., and Nogueira, M. L. (2020b). “Remote sensing for risk mapping of *Aedes aegypti* infestations: Is this a practical task?” *Acta tropica* 205, p. 105398. DOI: 10.1016/j.actatropica.2020.105398.
- Louis, V. R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y., and Dambach, P. (2014). “Modeling tools for dengue risk mapping - a systematic review”. *International journal of health geographics* 13, p. 50. DOI: 10.1186/1476-072X-13-50.
- Lozano-Fuentes, S., Hayden, M. H., Welsh-Rodriguez, C., Ochoa-Martinez, C., Tapia-Santos, B., Kobylinski, K. C., Uejio, C. K., Zielinski-Gutierrez, E., Monache, L. D., Monaghan, A. J., Steinhoff, D. F., and Eisen, L. (2012). “The dengue virus mosquito vector *Aedes aegypti* at high elevation in Mexico”. *The American journal of tropical medicine and hygiene* 87.5, pp. 902–909. DOI: 10.4269/ajtmh.2012.12-0244.
- Lucena, A. J. de, Faria Peres, L. de, Correa Rotunno Filho, O., and Almeida Franca, J. R. de (2015). “Estimation of the urban heat island in the Metropolitan Area of Rio de Janeiro - Brazil”. *2015 Joint Urban Remote Sensing Event (JURSE)*. IEEE, pp. 1–4. DOI: 10.1109/JURSE.2015.7120540.
- Luz, P. M., Vanni, T., Medlock, J., Paltiel, A. D., and Galvani, A. P. (2011). “Dengue vector control strategies in an urban setting: an economic modelling assessment”.

- Lancet (London, England)* 377.9778, pp. 1673–1680. DOI: 10.1016/S0140-6736(11)60246-8.
- Maddison, R. and Ni Mhurchu, C. (2009). “Global positioning system: a new opportunity in physical activity measurement”. *The international journal of behavioral nutrition and physical activity* 6, p. 73. DOI: 10.1186/1479-5868-6-73.
- Makepeace, B. L. and Gill, A. C. (2016). “Wolbachia”. *Rickettsiales*. Ed. by S. Thomas. Cham: Springer International Publishing, pp. 465–512. DOI: 10.1007/978-3-319-46859-4{\textunderscore}21.
- Malik, M., Lamba, H., Nakos, C., and Pfeffer, J. (2015). “Population Bias in Geotagged Tweets”. *Proceedings of the International AAAI Conference on Web and Social Media* 9.4, pp. 18–27. DOI: 10.1609/icwsm.v9i4.14688.
- Manrique-Saide, P., Davies, C. R., Coleman, P. G., Rebollar-Tellez, E., Che-Medoza, A., Dzul-Manzanilla, F., and Zapata-Peniche, A. (2008). “Pupal surveys for *Aedes aegypti* surveillance and potential targeted control in residential areas of Mérida, México”. *Journal of the American Mosquito Control Association* 24.2, pp. 289–298. DOI: 10.2987/5578.1.
- Mapillary (2023a). *Mapillary API Documentation*.
- Mapillary (2023b). *Mapillary API Documentation*.
- Maquart, P.-O., Froehlich, Y., and Boyer, S. (2022). “Plastic pollution and infectious diseases”. *The Lancet. Planetary health* 6.10, e842–e845. DOI: 10.1016/S2542-5196(22)00198-X.
- Martín, M. E., Alonso, A. C., Faraone, J., Stein, M., and Estallo, E. L. (2022a). *Aedes aegypti and Aedes albopictus abundance, landscape coverage and spectral indices effects in a subtropical city of Argentina*. DOI: 10.1101/2022.01.11.475665.
- Martín, M. E., Alonso, A. C., Faraone, J., Stein, M., and Estallo, E. L. (2022b). “Satellite observation to assess dengue risk due to *Aedes aegypti* and *Aedes albopictus* in a subtropical city of Argentina”. *Medical and veterinary entomology*. DOI: 10.1111/mve.12604.
- Martinez, J., Showering, A., Oke, C., Jones, R. T., and Logan, J. G. (2021). “Differential attraction in mosquito-human interactions and implications for disease control”. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 376.1818, p. 20190811. DOI: 10.1098/rstb.2019.0811.
- Martini, M., Triasputri, Y., Hestningsih, R., Yuliawati, S., and Purwantisasi, S. (2019). “Longevity and development of *Aedes aegypti* larvae to imago in domestic sewage water”. *Journal of the Medical Sciences (Berkala Ilmu Kedokteran)* 51.04. DOI: 10.19106/JMedSci005104201906.
- McCall, P. J. and Lenhart, A. (2008). “Dengue control”. *The Lancet. Infectious diseases* 8.1, pp. 7–9. DOI: 10.1016/S1473-3099(07)70298-0.
- Menchaca-Armenta, I., Ocampo-Torres, M., Hernández-Gómez, A., and Zamora-Cerritos, K. (2018). “Risk perception and level of knowledge of diseases transmitted by *Aedes aegypti*”. *Revista do Instituto de Medicina Tropical de Sao Paulo* 60, e10. DOI: 10.1590/S1678-9946201860010.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. (2019). “The current and future global distribution and population at risk of dengue”. *Nature microbiology* 4.9, pp. 1508–1515. DOI: 10.1038/s41564-019-0476-8.
- Ministério da Saúde Brazil (2013a). *Levantamento Rápido de Índices para Aedes aegypti*.
- Ministério da Saúde Brazil (2013b). *Levantamento rápido de índices para aedes aegypti*.

- Mir, S. A., Bhat, M. S., Rather, G. M., and Mattoo, D. (2022). “Role of big geospatial data in the COVID-19 crisis”. *Data Science for COVID-19*. Elsevier, pp. 589–609. DOI: 10.1016/B978-0-323-90769-9.00031-1.
- Miranda, V. F. V. V. de, Faria Peres, L. de, José de Lucena, A., França, J. R. d. A., and Libonati, R. (2022). “Urbanization-induced impacts on heat-energy fluxes in tropical South America from 1984 to 2020: The Metropolitan Area of Rio de Janeiro/Brazil”. *Building and Environment* 216, p. 109008. DOI: 10.1016/j.buildenv.2022.109008.
- Misslin, R., Vaguet, Y., Vaguet, A., and Daudé, É. (2018). “Estimating air temperature using MODIS surface temperature images for assessing *Aedes aegypti* thermal niche in Bangkok, Thailand”. *Environmental monitoring and assessment* 190.9, p. 537. DOI: 10.1007/s10661-018-6875-0.
- Monath, T. P. and Vasconcelos, P. F. C. (2015). “Yellow fever”. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 64, pp. 160–173. DOI: 10.1016/j.jcv.2014.08.030.
- Montagner, F. R. G., Silva, O. S., and Jahnke, S. M. (2018). “Mosquito species occurrence in association with landscape composition in green urban areas”. *Brazilian journal of biology = Revista brasleira de biologia* 78.2, pp. 233–239. DOI: 10.1590/1519-6984.04416.
- Moore, T. C. and Brown, H. E. (2022). “Estimating *Aedes aegypti* (Diptera: Culicidae) Flight Distance: Meta-Data Analysis”. *Journal of medical entomology* 59.4, pp. 1164–1170. DOI: 10.1093/jme/tjac070.
- Moraga, P. (2024). *Spatial statistics for data science: Theory and practice with R*. First edition. Chapman & Hall/CRC data science series. Boca Raton, FL: CRC Press.
- Moreno-Madriñán, M., Crosson, W., Eisen, L., Estes, S., Estes Jr., M., Hayden, M., Hemmings, S., Irwin, D., Lozano-Fuentes, S., Monaghan, A., Quattrochi, D., Welsh-Rodriguez, C., and Zielinski-Gutierrez, E. (2014). “Correlating Remote Sensing Data with the Abundance of Pupae of the Dengue Virus Mosquito Vector, *Aedes aegypti*, in Central Mexico”. *ISPRS International Journal of Geo-Information* 3.2, pp. 732–749. DOI: 10.3390/ijgi3020732.
- Morrison, A. C., Astete, H., Chapilliquen, F., Ramirez-Prada, C., Diaz, G., Getis, A., Gray, K., and Scott, T. W. (2004). “Evaluation of a sampling methodology for rapid assessment of *Aedes aegypti* infestation levels in Iquitos, Peru”. *Journal of medical entomology* 41.3, pp. 502–510. DOI: 10.1603/0022-2585-41.3.502.
- Muhammad, N. A. F., Abu Kassim, N. F., Ab Majid, A. H., Abd Rahman, A., Dieng, H., and Avicor, S. W. (2020). “Biting rhythm and demographic attributes of *Aedes albopictus* (Skuse) females from different urbanized settings in Penang Island, Malaysia under uncontrolled laboratory conditions”. *PloS one* 15.11, e0241688. DOI: 10.1371/journal.pone.0241688.
- Muñoz, Á. G., Chourio, X., Rivière-Cinamond, A., Diuk-Wasser, M. A., Kache, P. A., Mordecai, E. A., Harrington, L., and Thomson, M. C. (2020). “AeDES: a next-generation monitoring and forecasting system for environmental suitability of *Aedes*-borne disease transmission”. *Scientific reports* 10.1, p. 12640. DOI: 10.1038/s41598-020-69625-4.
- Murray, N. E. A., Quam, M. B., and Wilder-Smith, A. (2013). “Epidemiology of dengue: past, present and future prospects”. *Clinical epidemiology* 5, pp. 299–309. DOI: 10.2147/CLEP.S34440.
- Musso, D. and Gubler, D. J. (2016). “Zika Virus”. *Clinical microbiology reviews* 29.3, pp. 487–524. DOI: 10.1128/cmr.00072-15.
- Mutebi, J.-P., Wilke, A. B. B., Ostrum, E., Vasquez, C., Cardenas, G., Carvajal, A., Moreno, M., Petrie, W. D., Rodriguez, A., Presas, H., Rodriguez, J., Barnes, F.,

- Hamer, G. L., Juarez, J. G., Carbajal, E., Vitek, C. J., Estrada, X., Rios, T., Marshall, J., and Beier, J. C. (2022a). “Diel activity patterns of two distinct populations of *Aedes aegypti* in Miami, FL and Brownsville, TX”. *Scientific reports* 12.1, p. 5315. DOI: 10.1038/s41598-022-06586-w.
- Mutebi, J.-P., Wilke, A. B. B., Ostrum, E., Vasquez, C., Cardenas, G., Carvajal, A., Moreno, M., Petrie, W. D., Rodriguez, A., Presas, H., Rodriguez, J., Barnes, F., Hamer, G. L., Juarez, J. G., Carbajal, E., Vitek, C. J., Estrada, X., Rios, T., Marshall, J., and Beier, J. C. (2022b). “Diel activity patterns of two distinct populations of *Aedes aegypti* in Miami, FL and Brownsville, TX”. *Scientific reports* 12.1, p. 5315. DOI: 10.1038/s41598-022-06586-w.
- Nagao, Y., Thavara, U., Chitnumsup, P., Tawatsin, A., Chansang, C., and Campbell-Lendrum, D. (2003). “Climatic and social risk factors for *Aedes* infestation in rural Thailand”. *Tropical medicine & international health : TM & IH* 8.7, pp. 650–659. DOI: 10.1046/j.1365-3156.2003.01075.x.
- Nartey, O. T., Yang, G., Wu, J., and Asare, S. K. (2020). “Semi-Supervised Learning for Fine-Grained Classification With Self-Training”. *IEEE Access* 8, pp. 2109–2121. DOI: 10.1109/ACCESS.2019.2962258.
- Nascimento, K. L. C., Da Silva, J. F. M., Zequi, J. A. C., and Lopes, J. (2020). “Comparison Between Larval Survey Index and Positive Ovitrap Index in the Evaluation of Populations of *Aedes* (*Stegomyia*) *aegypti* (Linnaeus, 1762) North of Paraná, Brazil”. *Environmental health insights* 14, p. 1178630219886570. DOI: 10.1177/1178630219886570.
- Nasir, S., Jabeen, F., Abbas, S., Nasir, I., and Debboun, M. (2017). “Effect of Climatic Conditions and Water Bodies on Population Dynamics of the Dengue Vector, *Aedes aegypti* (Diptera: Culicidae)”. *Journal of Arthropod-Borne Diseases* 11.1, pp. 50–59.
- National Academy of Sciences (2019). *Reproducibility and Replicability in Science, Engineering, and Medicine*. Washington (DC). DOI: 10.17226/25303.
- Neiva, H., Da Silva, M., and Cardoso, C. (2017). “Analysis of Climate Behavior and Land Use in the City of Rio de Janeiro, RJ, Brazil”. *Climate* 5.3, p. 52. DOI: 10.3390/cli5030052.
- Nguyen, H. L., Tsolak, D., Karmann, A., Knauff, S., and Kühne, S. (2022). “Efficient and Reliable Geocoding of German Twitter Data to Enable Spatial Data Linkage to Official Statistics and Other Data Sources”. *Frontiers in sociology* 7, p. 910111. DOI: 10.3389/fsoc.2022.910111.
- Novaes, C., Silva Pinto, F., and Marques, R. C. (2022). “*Aedes Aegypti*-Insights on the Impact of Water Services”. *GeoHealth* 6.11, e2022GH000653. DOI: 10.1029/2022GH000653.
- O’Hare, G. and Barke, M. (2002). *GeoJournal* 56.3, pp. 225–240. DOI: 10.1023/A:1025134625932.
- Oliveira Lemos, L. de, Oscar Júnior, A. C., and Assis Mendonça, F. de (2021). “Urban climate maps as a public health tool for urban planning: The case of dengue fever in Rio De Janeiro/Brazil”. *Urban Climate* 35, p. 100749. DOI: 10.1016/j.uclim.2020.100749.
- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., Nadai, M. de, Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., Cordes, N. de, Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., Saramäki, J., Scarpino, S. V., Tizzoni, M., Verhulst, S., and Vinck, P. (2020). “Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle”. *Science advances* 6.23, eabc0764. DOI: 10.1126/sciadv.abc0764.

- Panigrahi, S. K., Priyadarshini, O., Jena, A., Jhankar, A., and Mishra, S. (2024). “Review of feeding, biting and resting behaviour of *Aedes aegypti* and *Aedes albopictus*”. *International Journal of Mosquito Research* 11.3, pp. 01–11. DOI: 10.22271/23487941.2024.v11.i3a.771.
- Panigutti, C., Tizzoni, M., Bajardi, P., Smoreda, Z., and Colizza, V. (2017). “Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models”. *Royal Society open science* 4.5, p. 160950. DOI: 10.1098/rsos.160950.
- Passos, W. L., Araujo, G. M., Lima, A. A. de, Netto, S. L., and Da Silva, E. A. (2022). “Automatic detection of *Aedes aegypti* breeding grounds based on deep networks with spatio-temporal consistency”. *Computers, Environment and Urban Systems* 93, p. 101754. DOI: 10.1016/j.compenvurbsys.2021.101754.
- Passos, W. L., Da Barreto, C. S., Araujo, G. M., Haque, U., Netto, S. L., and Da Silva, E. A. (2023). “Toward improved surveillance of *Aedes aegypti* breeding grounds through artificially augmented data”. *Engineering Applications of Artificial Intelligence* 123, p. 106488. DOI: 10.1016/j.engappai.2023.106488.
- Pataki, B. A., Garriga, J., Eritja, R., Palmer, J. R. B., Bartumeus, F., and Csabai, I. (2021). “Deep learning identification for citizen science surveillance of tiger mosquitoes”. *Scientific reports* 11.1, p. 4718. DOI: 10.1038/s41598-021-83657-4.
- Pereira, R. H. (2019). “Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro”. *Journal of Transport Geography* 74, pp. 321–332. DOI: 10.1016/j.jtrangeo.2018.12.005.
- Peres, L. d. F., Lucena, A. J. de, Rotunno Filho, O. C., and França, J. R. d. A. (2018). “The urban heat island in Rio de Janeiro, Brazil, in the last 30 years using remote sensing data”. *International Journal of Applied Earth Observation and Geoinformation* 64, pp. 104–116. DOI: 10.1016/j.jag.2017.08.012.
- Perkins, T. A., Metcalf, C. J. E., Grenfell, B. T., and Tatem, A. J. (2015). “Estimating Drivers of Autochthonous Transmission of Chikungunya Virus in its Invasion of the Americas”. *PLoS Currents* 7. DOI: 10.1371/currents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc.
- Perkins, T. A., Paz-Soldan, V. A., Stoddard, S. T., Morrison, A. C., Forshey, B. M., Long, K. C., Halsey, E. S., Kochel, T. J., Elder, J. P., Kitron, U., Scott, T. W., and Vazquez-Prokopec, G. M. (2016). “Calling in sick: impacts of fever on intra-urban human mobility”. *Proceedings. Biological sciences* 283.1834. DOI: 10.1098/rspb.2016.0390.
- Petersen, L. R., Jamieson, D. J., Powers, A. M., and Honein, M. A. (2016). “Zika Virus”. *The New England journal of medicine* 374.16, pp. 1552–1563. DOI: 10.1056/NEJMr1602113.
- Pialoux, G., Gaüzère, B.-A., Jauréguiberry, S., and Strobel, M. (2007). “Chikungunya, an epidemic arbovirolosis”. *The Lancet. Infectious diseases* 7.5, pp. 319–327. DOI: 10.1016/S1473-3099(07)70107-X.
- PPGAU UFF (2023a). *Portal de dados urbanos: Building footprints*.
- PPGAU UFF (2023b). *Portal de dados urbanos: High resolution elevation data set for the city of Rio de Janeiro*.
- Puntasecca, C. J., King, C. H., and LaBeaud, A. D. (2021). “Measuring the global burden of chikungunya and Zika viruses: A systematic review”. *PLoS neglected tropical diseases* 15.3, e0009055. DOI: 10.1371/journal.pntd.0009055.
- QGIS Association (2024). *QGIS Geographic Information System*.
- Ramadona, A. L., Tozan, Y., Lazuardi, L., and Rocklöv, J. (2019). “A combination of incidence data and mobility proxies from social media predicts the intra-urban

- spread of dengue in Yogyakarta, Indonesia”. *PLoS neglected tropical diseases* 13.4, e0007298. DOI: 10.1371/journal.pntd.0007298.
- Regilme, M. A. F., Carvajal, T. M., Honnen, A.-C., Amalin, D. M., and Watanabe, K. (2021). “The influence of roads on the fine-scale population genetic structure of the dengue vector *Aedes aegypti* (Linnaeus)”. *PLoS neglected tropical diseases* 15.2, e0009139. DOI: 10.1371/journal.pntd.0009139.
- Reiner, R. C., Stoddard, S. T., and Scott, T. W. (2014). “Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal”. *Epidemics* 6, pp. 30–36. DOI: 10.1016/j.epidem.2013.12.003.
- Reiskind, M. H. and Lounibos, L. P. (2009). “Effects of intraspecific larval competition on adult longevity in the mosquitoes *Aedes aegypti* and *Aedes albopictus*”. *Medical and veterinary entomology* 23.1, pp. 62–68. DOI: 10.1111/j.1365-2915.2008.00782.x.
- Ribeiro, M. S., Ferreira, D. F., Azevedo, R. C., Santos, G. B. G. D., and Medronho, R. d. A. (2021). “Índices larvais de *Aedes aegypti* e incidência de dengue: um estudo ecológico no Estado do Rio de Janeiro, Brasil”. *Cadernos de saude publica* 37.7, e00263320. DOI: 10.1590/0102-311X00263320.
- Roiz, D., Pontifes, P. A., Jourdain, F., Diagne, C., Leroy, B., Vaissière, A.-C., Tolsá-García, M. J., Salles, J.-M., Simard, F., and Courchamp, F. (2024). “The rising global economic costs of invasive *Aedes* mosquitoes and *Aedes*-borne diseases”. *The Science of the total environment* 933, p. 173054. DOI: 10.1016/j.scitotenv.2024.173054.
- Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., and Hsiang, S. (2021). “A generalizable and accessible approach to machine learning with global satellite imagery”. *Nature communications* 12.1, p. 4392. DOI: 10.1038/s41467-021-24638-z.
- Roslan, M. A., Ngui, R., Vythilingam, I., Chan, K. F., Ong, P. S., Low, C. K., Muhammed, N. H., and Wan Sulaiman, W. Y. (2022). “Surveillance of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in high-rise apartment buildings in Selangor, Malaysia”. *International Journal of Tropical Insect Science* 42.2, pp. 1959–1969. DOI: 10.1007/s42690-021-00725-y.
- Roslan, M. A., Shafie, A., Ngui, R., Lim, Y. A. L., and Sulaiman, W. Y. W. (2013). “Vertical infestation of the dengue vectors *Aedes aegypti* and *Aedes albopictus* in apartments in Kuala Lumpur, Malaysia”. *Journal of the American Mosquito Control Association* 29.4, pp. 328–336. DOI: 10.2987/13-6363.1.
- Rowley, W. A. and Graham, C. L. (1968). “The effect of temperature and relative humidity on the flight performance of female *Aedes aegypti*”. *Journal of insect physiology* 14.9, pp. 1251–1257. DOI: 10.1016/0022-1910(68)90018-8.
- Rue, H. and Lindgren, F. (2024). *R-INLA package*.
- Rund, S. S. C., Labb, L. F., Benefiel, O. M., and Duffield, G. E. (2020). “Artificial Light at Night Increases *Aedes aegypti* Mosquito Biting Behavior with Implications for Arboviral Disease Transmission”. *The American journal of tropical medicine and hygiene* 103.6, pp. 2450–2452. DOI: 10.4269/ajtmh.20-0885.
- Sallam, M. F., Fizer, C., Pilant, A. N., and Whung, P.-Y. (2017). “Systematic Review: Land Cover, Meteorological, and Socioeconomic Determinants of *Aedes* Mosquito Habitat for Risk Mapping”. *International journal of environmental research and public health* 14.10. DOI: 10.3390/ijerph14101230.
- Sattenspiel, Lisa and Lloyd, Alun (2009). *The geographic spread of infectious diseases: models and applications*. 5th ed. Princeton University Press.
- Schaber, K. L., Paz-Soldan, V. A., Morrison, A. C., Elson, W. H. D., Rothman, A. L., Mores, C. N., Astete-Vega, H., Scott, T. W., Waller, L. A., Kitron, U., Elder, J. P.,

- Barker, C. M., Perkins, T. A., and Vazquez-Prokopec, G. M. (2019). “Dengue illness impacts daily human mobility patterns in Iquitos, Peru”. *PLoS neglected tropical diseases* 13.9, e0007756. DOI: 10.1371/journal.pntd.0007756.
- Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., and Brockmann, D. (2020). “COVID-19 lockdown induces disease-mitigating structural changes in mobility networks”. *Proceedings of the National Academy of Sciences of the United States of America* 117.52, pp. 32883–32890. DOI: 10.1073/pnas.2012326117.
- Schwartz, L. M., Halloran, M. E., Durbin, A. P., and Longini, I. M. (2015). “The dengue vaccine pipeline: Implications for the future of dengue control”. *Vaccine* 33.29, pp. 3293–3298. DOI: 10.1016/j.vaccine.2015.05.010.
- Secretario Municipal de Saúde Rio de Janeiro (2024). *Dengue: dados epidemiológicos*.
- Semenza, J. C., Sudre, B., Miniota, J., Rossi, M., Hu, W., Kossowsky, D., Suk, J. E., van Bortel, W., and Khan, K. (2014). “International dispersal of dengue through air travel: importation risk for Europe”. *PLoS neglected tropical diseases* 8.12, e3278. DOI: 10.1371/journal.pntd.0003278.
- Serere, H. N., Resch, B., and Havas, C. R. (2023). “Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection”. *PloS one* 18.3, e0282942. DOI: 10.1371/journal.pone.0282942.
- Shah, P. and Patel, C. R. (2020). “Prevention is Better than Cure: An Application of Big Data and Geospatial Technology in Mitigating Pandemic”. *Transactions of the Indian National Academy of Engineering : an international journal of engineering and technology* 5.2, pp. 187–192. DOI: 10.1007/s41403-020-00120-y.
- Shepard, D. S., Undurraga, E. A., Halasa, Y. A., and Stanaway, J. D. (2016). “The global economic burden of dengue: a systematic analysis”. *The Lancet. Infectious diseases* 16.8, pp. 935–941. DOI: 10.1016/S1473-3099(16)00146-8.
- Shorten, C. and Khoshgoftaar, T. M. (2019). “A survey on Image Data Augmentation for Deep Learning”. *Journal of Big Data* 6.1. DOI: 10.1186/s40537-019-0197-0.
- Simmons, C. P., Farrar, J. J., van Nguyen, V. C., and Wills, B. (2012). “Dengue”. *The New England journal of medicine* 366.15, pp. 1423–1432. DOI: 10.1056/NEJMra1110265.
- Souza, S. S. de, Da Silva, I. G., and Da Silva, H. H. G. (2010). “Associação entre incidência de dengue, pluviosidade e densidade larvária de *Aedes aegypti*, no Estado de Goiás”. *Revista da Sociedade Brasileira de Medicina Tropical* 43.2, pp. 152–155. DOI: 10.1590/S0037-86822010000200009.
- Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., Hay, S. I., Bedi, N., Bensenor, I. M., Castañeda-Orjuela, C. A., Chuang, T.-W., Gibney, K. B., Memish, Z. A., Rafay, A., Ukwaja, K. N., Yonemoto, N., and Murray, C. J. L. (2016). “The global burden of dengue: an analysis from the Global Burden of Disease Study 2013”. *The Lancet. Infectious diseases* 16.6, pp. 712–723. DOI: 10.1016/S1473-3099(16)00026-8.
- Stefopoulou, A., Balatsos, G., Petraki, A., LaDeau, S. L., Papachristos, D., and Michalakis, A. (2018). “Reducing *Aedes albopictus* breeding sites through education: A study in urban area”. *PloS one* 13.11, e0202451. DOI: 10.1371/journal.pone.0202451.
- Stewart Ibarra, A. M., Luzadis, V. A., Borbor Cordova, M. J., Silva, M., Ordoñez, T., Beltrán Ayala, E., and Ryan, S. J. (2014). “A social-ecological analysis of community perceptions of dengue fever and *Aedes aegypti* in Machala, Ecuador”. *BMC public health* 14, p. 1135. DOI: 10.1186/1471-2458-14-1135.
- Stewart Ibarra, A. M., Ryan, S. J., Beltrán, E., Mejía, R., Silva, M., and Muñoz, A. (2013). “Dengue vector dynamics (*Aedes aegypti*) influenced by climate and social

- factors in Ecuador: implications for targeted control”. *PloS one* 8.11, e78263. DOI: 10.1371/journal.pone.0078263.
- Stoddard, S. T., Forshey, B. M., Morrison, A. C., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., Astete, H., Reiner, R. C., Vilcarrromero, S., Elder, J. P., Halsey, E. S., Kochel, T. J., Kitron, U., and Scott, T. W. (2013). “House-to-house human movement drives dengue virus transmission”. *Proceedings of the National Academy of Sciences of the United States of America* 110.3, pp. 994–999. DOI: 10.1073/pnas.1213349110.
- Stoddard, S. T., Morrison, A. C., Vazquez-Prokopec, G. M., Paz Soldan, V., Kochel, T. J., Kitron, U., Elder, J. P., and Scott, T. W. (2009). “The role of human movement in the transmission of vector-borne pathogens”. *PLoS neglected tropical diseases* 3.7, e481. DOI: 10.1371/journal.pntd.0000481.
- Studenmund, A. H. (2017). *A practical guide to using econometrics*. Seventh edition, global edition. Harlow, England: Pearson.
- Su Yin, M., Bicout, D. J., Haddawy, P., Schöning, J., Laosiritaworn, Y., and Sa-Angchai, P. (2021). “Added-value of mosquito vector breeding sites from street view images in the risk mapping of dengue incidence in Thailand”. *PLoS neglected tropical diseases* 15.3, e0009122. DOI: 10.1371/journal.pntd.0009122.
- Sudeep, A. B. and Parashar, D. (2008). “Chikungunya: an overview”. *Journal of bio-sciences* 33.4, pp. 443–449. DOI: 10.1007/s12038-008-0063-2.
- Sussman, A. and Hollander, J. B., eds. (2021). *Urban experience and design: Contemporary perspectives on improving the public realm*. New York, NY: Routledge.
- Swan, T., Russell, T. L., Staunton, K. M., Field, M. A., Ritchie, S. A., and Burkot, T. R. (2022). “A literature review of dispersal pathways of *Aedes albopictus* across different spatial scales: implications for vector surveillance”. *Parasites & vectors* 15.1, p. 303. DOI: 10.1186/s13071-022-05413-5.
- Tedjou, A. N., Kamgang, B., Yougang, A. P., Njiokou, F., and Wondji, C. S. (2019). “Update on the geographical distribution and prevalence of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae), two major arbovirus vectors in Cameroon”. *PLoS neglected tropical diseases* 13.3, e0007137. DOI: 10.1371/journal.pntd.0007137.
- Terroso-Saenz, F., Muñoz, A., Arcas, F., and Curado, M. (2022). “Can Twitter be a Reliable Proxy to Characterize Nation-wide Human Mobility? A Case Study of Spain”. *Social Science Computer Review*, p. 089443932110710. DOI: 10.1177/08944393211071071.
- Thomas, S. J. and Yoon, I.-K. (2019). “A review of Dengvaxia®: development to deployment”. *Human vaccines & immunotherapeutics* 15.10, pp. 2295–2314. DOI: 10.1080/21645515.2019.1658503.
- Tsuda, Y. and Takagi, M. (2001). “Survival and Development of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) Larvae Under a Seasonally Changing Environment in Nagasaki, Japan”. *Environmental Entomology* 30.5, pp. 855–860. DOI: 10.1603/0046-225X-30.5.855.
- Tully, D. and Griffiths, C. L. (2021). “Dengvaxia: the world’s first vaccine for prevention of secondary dengue”. *Therapeutic advances in vaccines and immunotherapy* 9, p. 25151355211015839. DOI: 10.1177/25151355211015839.
- TuzuTa Lin (2023). *labelImg: Graphical image annotation tool and label object bounding boxes in images*.
- Ubaldi, E., Yabe, T., Jones, N. K. W., Khan, M. F., Ukkusuri, S. V., Di Clemente, R., and Strano, E. (2021). “Mobilkit: A Python Toolkit for Urban Resilience and Disaster Risk Management Analytics using High Frequency Human Mobility Data”. DOI: 10.48550/arXiv.2107.14297.

- Valdez, L. D., Sibona, G. J., and Condat, C. A. (2018). “Impact of rainfall on *Aedes aegypti* populations”. *Ecological Modelling* 385, pp. 96–105. DOI: 10.1016/j.ecolmodel.2018.07.003.
- Valle, D. and Aguiar, R. (2023a). “Is the mosquito the problem? *Aedes aegypti* and urban arboviruses - contradictions and reflections”. *Epidemiologia e servicios de saude : revista do Sistema Unico de Saude do Brasil* 32.2, e2023538. DOI: 10.1590/S2237-96222023000200024.
- Valle, D. and Aguiar, R. (2023b). “Is the mosquito the problem? *Aedes aegypti* and urban arboviruses - contradictions and reflections”. *Epidemiologia e servicios de saude : revista do Sistema Unico de Saude do Brasil* 32.2, e2023538. DOI: 10.1590/S2237-96222023000200024.
- Valle, D., Bellinato, D. F., Viana-Medeiros, P. F., Lima, J. B. P., and Martins Junior, A. d. J. (2019). “Resistance to temephos and deltamethrin in *Aedes aegypti* from Brazil between 1985 and 2017”. *Memorias do Instituto Oswaldo Cruz* 114, e180544. DOI: 10.1590/0074-02760180544.
- van Dyk, D. A. and Meng, X.-L. (2001). “The Art of Data Augmentation”. *Journal of Computational and Graphical Statistics* 10.1, pp. 1–50. DOI: 10.1198/10618600152418584.
- Vannavong, N., Seidu, R., Stenström, T.-A., Dada, N., and Overgaard, H. J. (2017). “Effects of socio-demographic characteristics and household water management on *Aedes aegypti* production in suburban and rural villages in Laos and Thailand”. *Parasites & vectors* 10.1, p. 170. DOI: 10.1186/s13071-017-2107-7.
- Vanwambeke, S. O., Somboon, P., Harbach, R. E., Isenstadt, M., Lambin, E. F., Walton, C., and Butlin, R. K. (2007). “Landscape and Land Cover Factors Influence the Presence of *Aedes* and *Anopheles* Larvae”. *Journal of medical entomology* 44.1, pp. 133–144. DOI: 10.1093/jmedent/41.5.133.
- Vasconcelos, A. S. V., Lima, J. S., Cardoso, R. T. N., Acebal, J. L., and Loaiza, A. M. (2022). “Optimal control of *Aedes aegypti* using rainfall and temperature data”. *Computational and Applied Mathematics* 41.3. DOI: 10.1007/s40314-022-01804-7.
- Vasconcelos, D., Yin, M. S., Wetjen, F., Herbst, A., Ziemer, T., Förster, A., Barkowsky, T., Nunes, N., and Haddawy, P. (2021). “Counting Mosquitoes in the Wild”. *Proceedings of the Conference on Information Technology for Social Good*. New York, NY, USA: ACM, pp. 43–48. DOI: 10.1145/3462203.3475914.
- Vasconcelos, P. F. C. (2017). “Dengue”. *Arthropod Borne Diseases*. Ed. by C. B. Marcondes. Cham: Springer International Publishing, pp. 89–99. DOI: 10.1007/978-3-319-13884-8\_textunderscore7.
- Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W., and Kitron, U. (2013). “Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment”. *PloS one* 8.4, e58802. DOI: 10.1371/journal.pone.0058802.
- Vinhal Frutuoso, L. C. and Barbosa Duraes, S. M. (2023). *Information Note No. 37/2023 - CGARB/ DEDT/ SVSA/ MS*. Ed. by Ministry of Health - Department of Communicable Diseases - General Coordination of Arbovirus Surveillance. Unpublished confidential document.
- Vu, D. M., Jungkind, D., and Angelle, D. L. (2017). “Chikungunya Virus”. *Clinics in laboratory medicine* 37.2, pp. 371–382. DOI: 10.1016/j.cll.2017.01.008.
- Wei, L., Fernández-Santos, N. A., Hamer, G. L., Lara-Ramírez, E. E., and Rodríguez-Pérez, M. A. (2023). “Daytime Resting Activity of *Aedes Aegypti* and *Culex Quin-*

- quefasciatus Populations in Northern Mexico”. *Journal of the American Mosquito Control Association* 39.3, pp. 157–167. DOI: 10.2987/23-7122.
- Wen, T.-H., Lin, M.-H., Teng, H.-J., and Chang, N.-T. (2015). “Incorporating the human-Aedes mosquito interactions into measuring the spatial risk of urban dengue fever”. *Applied Geography* 62, pp. 256–266. DOI: 10.1016/j.apgeog.2015.05.003.
- Werner, M. (2021). *Handbook of Big Geospatial Data*. Cham: Springer International Publishing AG.
- Werren, J. H., Baldo, L., and Clark, M. E. (2008). “Wolbachia: master manipulators of invertebrate biology”. *Nature reviews. Microbiology* 6.10, pp. 741–751. DOI: 10.1038/nrmicro1969.
- Wesolowski, A., Buckee, C. O., Engø-Monsen, K., and Metcalf, C. J. E. (2016). “Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data”. *The Journal of infectious diseases* 214.suppl\_4, S414–S420. DOI: 10.1093/infdis/jiw273.
- Westby, K. M., Adalsteinsson, S. A., Biro, E. G., Beckermann, A. J., and Medley, K. A. (2021). “Aedes albopictus Populations and Larval Habitat Characteristics across the Landscape: Significant Differences Exist between Urban and Rural Land Use Types”. *Insects* 12.3. DOI: 10.3390/insects12030196.
- Whelan, P. I., Kurucz, N., Pettit, W. J., and Krause, V. (2020). “Elimination of Aedes aegypti in northern Australia, 2004-2006”. *Journal of vector ecology : journal of the Society for Vector Ecology* 45.1, pp. 118–126. DOI: 10.1111/jvec.12379.
- WHO (2017). “Global Vector Control Response 2017-2030”.
- Wilder-Smith, A., Ooi, E.-E., Horstick, O., and Wills, B. (2019). “Dengue”. *Lancet (London, England)* 393.10169, pp. 350–363. DOI: 10.1016/S0140-6736(18)32560-1.
- Wilk-da-Silva, R., Souza Leal Diniz, M. M. C. de, Marrelli, M. T., and Wilke, A. B. B. (2018). “Wing morphometric variability in Aedes aegypti (Diptera: Culicidae) from different urban built environments”. *Parasites & vectors* 11.1, p. 561. DOI: 10.1186/s13071-018-3154-4.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). “The FAIR Guiding Principles for scientific data management and stewardship”. *Scientific data* 3, p. 160018. DOI: 10.1038/sdata.2016.18.
- Willoughby, J. R., McKenzie, B. A., Ahn, J., Steury, T. D., Lepczyk, C. A., and Zohdy, S. (2024). “Assessing and managing the risk of Aedes mosquito introductions via the global maritime trade network”. *PLoS neglected tropical diseases* 18.4, e0012110. DOI: 10.1371/journal.pntd.0012110.
- Wilson-Bahun, T. A., Kamgang, B., Lenga, A., and Wondji, C. S. (2020). “Larval ecology and infestation indices of two major arbovirus vectors, Aedes aegypti and Aedes albopictus (Diptera: Culicidae), in Brazzaville, the capital city of the Republic of the Congo”. *Parasites & vectors* 13.1, p. 492. DOI: 10.1186/s13071-020-04374-x.

- Wong, G. K. and Jim, C. Y. (2017). “Urban-microclimate effect on vector mosquito abundance of tropical green roofs”. *Building and Environment* 112, pp. 63–76. DOI: 10.1016/j.buildenv.2016.11.028.
- World Health Organization et al. (2016). *Entomological surveillance for Aedes spp. in the context of Zika virus: interim guidance for entomologists*.
- Wu, S. L., Henry, J. M., Citron, D. T., Mbabazi Ssebuliba, D., Nakakawa Nsumba, J., Sánchez C, H. M., Brady, O. J., Guerra, C. A., García, G. A., Carter, A. R., Ferguson, H. M., Afolabi, B. E., Hay, S. I., Reiner, R. C., Kiware, S., and Smith, D. L. (2023). “Spatial dynamics of malaria transmission”. *PLoS computational biology* 19.6, e1010684. DOI: 10.1371/journal.pcbi.1010684.
- Wu, S. L., Henry, J. M., Citron, D. T., Ssebuliba, D. M., Nsumba, J. N., Sánchez C., H. M., Brady, O. J., Guerra, C. A., García, G. A., Carter, A. R., Ferguson, H. M., Afolabi, B. E., Hay, S. I., Reiner, R. C., Kiware, S., and Smith, D. L. (2022). *Spatial Dynamics of Malaria Transmission*. DOI: 10.1101/2022.11.07.22282044.
- Young, K. I., Mundis, S., Widen, S. G., Wood, T. G., Tesh, R. B., Cardosa, J., Vasilakis, N., Perera, D., and Hanley, K. A. (2017). “Abundance and distribution of sylvatic dengue virus vectors in three different land cover types in Sarawak, Malaysian Borneo”. *Parasites & vectors* 10.1, p. 406. DOI: 10.1186/s13071-017-2341-z.
- Zahid, M. H., van Wyk, H., Morrison, A. C., Coloma, J., Lee, G. O., Cevallos, V., Ponce, P., and Eisenberg, J. N. S. (2023). “The biting rate of *Aedes aegypti* and its variability: A systematic review (1970-2022)”. *PLoS neglected tropical diseases* 17.8, e0010831. DOI: 10.1371/journal.pntd.0010831.
- Zahouli, J. B. Z., Koudou, B. G., Müller, P., Malone, D., Tano, Y., and Utzinger, J. (2017). “Effect of land-use changes on the abundance, distribution, and host-seeking behavior of *Aedes* arbovirus vectors in oil palm-dominated landscapes, southeastern Côte d’Ivoire”. *PLoS one* 12.12, e0189082. DOI: 10.1371/journal.pone.0189082.
- Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., and Ratti, C. (2018). “Measuring human perceptions of a large-scale urban region using machine learning”. *Landscape and Urban Planning* 180, pp. 148–160. DOI: 10.1016/j.landurbplan.2018.08.020.
- Zhao, Y., Yin, P., Li, Y., He, X., Du, J., Tao, C., Guo, Y., Prosperi, M., Veltri, P., Yang, X., Wu, Y., and Bian, J. (2021). “Data and Model Biases in Social Media Analyses: A Case Study of COVID-19 Tweets”. *AMIA ... Annual Symposium proceedings. AMIA Symposium 2021*, pp. 1264–1273.
- Zhao, Z., Shaw, S.-L., Yin, L., Fang, Z., Yang, X., Zhang, F., and Wu, S. (2019). “The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data”. *International Journal of Geographical Information Science* 33.7, pp. 1471–1495. DOI: 10.1080/13658816.2019.1584805.
- Zuur, A. F., Ieno, E. N., and Saveliev, A. A. (2017). “Beginner’s Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA”.



**Part II**  
**Publications**



## Publication I:

# Semi-supervised Water Tank Detection to Support Vector Control of Emerging Infectious Diseases Transmitted by *Aedes aegypti*

**Authors.** Steffen Knoblauch, Hao Li, Sven Lautenbach, Yara Elshiaty, Antonio Rocha, Bernd Resch, Dorian Arifi, Thomas Jänisch, Ivonne Morales, Alexander Zipf

**Outlet.** *International Journal of Applied Earth Observation and Geoinformation (Special Issue: Harnessing Geospatial Big Data for Infectious Diseases)*

**Status.** Published on 2021-06-28 under CC-BY 4.0

### Abstract.

The disease transmitting mosquito *Aedes Aegypti* is an increasing global threat. It breeds in small artificial containers such as rainwater tanks and can be characterized by a short flight range. The resulting high spatial variability of abundance is challenging to model. Therefore, we tested an approach to map water tank density as a spatial proxy for urban *Aedes Aegypti* habitat suitability. Water tank density mapping was performed by a semi-supervised self-training approach based on open accessible satellite imagery for the city of Rio de Janeiro. We ran a negative binomial generalized linear regression model to evaluate the statistical significance of water tank density for modeling inner-urban *Aedes Aegypti* distribution measured by an entomological surveillance system between January 2019 and December 2021. Our proposed semi-supervised model outperformed a supervised model for water tank detection with respect to the F1-score by 22%. Water tank density was a significant predictor for the mean eggs per trap rate of *Aedes Aegypti*. This shows the potential of the proposed indicator to enrich urban entomological surveillance systems to plan more targeted vector control interventions, presumably leading to less infectious rates of dengue, Zika, and chikungunya in the future.

**Keywords.** *Aedes Aegypti* · Eco-epidemiology · GeoAI · Object detection · Ovitraps · Rio de Janeiro · Semi-supervised self-training · Urban health · Vector control

## 1 Introduction

The recurring worldwide outbreaks of the severe acute respiratory syndrome (SARS)-associated coronavirus in 2003 and 2020 demonstrate how frequently and rapidly infectious diseases can spread in a globalized world. However, it is not globalization alone that is a driving factor for increased occurrences of infectious diseases but also climate change (Semenza et al., 2022). This is particularly true for mosquito-borne diseases, as rising global temperatures lengthens annual transmission seasons and leads to larger suitability areas for mosquitoes (Colón-González et al., 2021; Rocklöv and Dubrow, 2020). Considering all pathogen transmitting mosquitoes worldwide, *Aedes Aegypti* is the most prevalent one (European Centre for Disease Prevention and Control, 2016; Wilke et al., 2020). It is the primary vector for Zika, chikungunya, yellow fever, and dengue with a 30-fold increase in incidences over the last 50 years (Ebi and Nealon, 2016). The WHO is

estimating that by 2080 over 60 percent of the world's population will live under direct risk of *Aedes Aegypti* (Messina et al., 2019; WHO, 2017). This turns this disease vector into an emerging global threat (Ebi and Nealon, 2016).

As of now there is no effective vaccine for dengue (Amorim and Birbrair, 2022; Schwartz et al., 2015; WHO, 2022), Zika, or chikungunya (Kantor, 2018; Schrauf et al., 2020). Accordingly, vector control, involving the process of eliminating vector breeding habitats and the application of insecticides to maintain mosquito populations at acceptable level, remains the most effective countermeasure for these diseases (Hladish et al., 2020; Lobo et al., 2018; Wilson et al., 2020). Vector control, however, is very costly as it requires a massive workforce and is often limited by regulative constraints on the use of insecticides, especially in urban areas where most infections by *Aedes Aegypti* occur (Knerer et al., 2020; Tabora et al., 2022). Therefore, mapping of *Aedes Aegypti* on the urban scale is of particular interest in order to implement local vector control measures in a more targeted manner and, above all, at lower costs (Boser et al., 2021; Da Queiroz and Medronho, 2022; Limkittikul et al., 2014; Runge-Ranzinger et al., 2014). This is especially important for the Global South where public health budgets for disease prevention are often scarce despite financial support from the WHO (Yukich et al., 2008).

The mapping of *Aedes Aegypti*, however, is not trivial. *Aedes Aegypti* is an urban favoring mosquito with a short flight range of around 200 meters and thus limited habitat size (Bomfim et al., 2020; Harrington et al., 2005; Honório et al., 2003). It lives in close vicinity to its breeding sites. These can be characterized as small artificial water containers, such as discarded tires (Getachew et al., 2015), buckets, barrels, pet dishes, construction blocks (Morrison et al., 2004; WHO, 2012), storm drains (Paploski et al., 2016), trash (Banerjee et al., 2015), flower pots (Vezzani, 2007), or water tanks (Trewin et al., 2021). Many of these containers occur with great spatial variance due to social urban structures (David et al., 2009). This, in combination with the small size and the limited flight range of mosquitoes, leads to a high spatial variability of *Aedes Aegypti* abundance. It differentiates *Aedes Aegypti* strongly from other mosquito species such as the malaria transmitting *Anopheles* mosquito, which tends to breed in large natural water bodies and thus occurs in higher concentrations (Chavasse, 2002; Gwitira et al., 2018; Youssefi et al., 2022). Consequently, the task of spatial mapping of mosquito distribution is more challenging for *Aedes Aegypti* than for other mosquito species (Boser et al., 2021).

Currently, there are two *Aedes Aegypti* mapping approaches in use to implement vector control in a more efficient and cost-saving manner. One of them are sample-based entomological surveillance systems including the positioning of mosquito traps and the conduction of household surveys (Bowman et al., 2014; Pan American Health Organization, 2019). These monitoring systems require a large amount of manual work but provide valid information on mosquito abundance, such as precise counts of mosquito eggs, larvae, and pupa. Nevertheless, they are hard to scale due to their labor-exhaustive nature, often cannot cover larger areas in high resolution and need trained personnel, which all together limits its practical scope (Vasconcelos et al., 2021). The alternative approach is based on the modeling of mosquito abundance via spatial proxies using modern computing techniques and harnessing big spatial data such as satellite imagery.

These methods are less precise, because proxies by definition only provide indirect evidence of a phenomena. However, they require less manual work and consequently offer the possibility for a much broader spatial coverage as well as continuous mapping to capture the high spatial variability of *Aedes Aegypti* abundance in urban areas (Boser et al., 2021; Louis et al., 2014).

There are several approaches in literature showing the benefits of *Aedes Aegypti* mapping with spatial proxies. Some studies retrieve proxies from citizen science (Agarwal et al., 2014; Caputo et al., 2020; Cho et al., 2021; Low et al., 2021; Low et al., 2022; Muñoz et al., 2020), others from remote sensing (Chang et al., 2009; Cunha et al., 2021; Dandabathula, 2019; Fernandes et al., 2020; Lorenz et al., 2020; Machault et al., 2014; McFeeters, 2013; Uusitalo et al., 2019), street view (Andersson et al., 2018; Haddawy et al., 2019; Su Yin et al., 2021), or drone imagery (Dias et al., 2018; Haas-Stapleton et al., 2019; Mehra et al., 2016; Passos et al., 2020; Schenkel et al., 2020). However, these approaches have shortcomings. Remote sensing studies, based on not very-high-resolution (VHR) satellite imagery, are presumably too coarse to detect small scale features that provide breeding habitats. Approaches that rely on citizen science, drone imagery, or street view are hard to transfer to other case studies since the data used is only available for selected sites or expensive to get. Put differently, most of these approaches neither derive high resolution proxies with open accessible data, nor use scalable workflows to map *Aedes Aegypti* below 200 meters to consider limited mosquito flight ranges (Louis et al., 2014; Sallam et al., 2017). The application of object detection models for urban *Aedes Aegypti* breeding site mapping based on open accessible satellite imagery is rare. The same applies for the combination of entomological surveillance data with automatic mapping workflows in the field, although there is a WHO pillar of action called “scale up and integrate tools and approaches for global vector control” (WHO, 2017). This paper addresses the following research gaps:

- **Research Gap 1:** There is a **need for high resolution proxies** to enrich entomological surveillance data of *Aedes Aegypti* to conduct vector control with more focus and lower costs.
- **Research Gap 2:** There is a **need for scalable approaches based on open accessible data** to retrieve high resolution proxies for the mapping of inner-urban *Aedes Aegypti* distributions.

We analyze how far recent advances in deep learning can help to close these research gaps. We envision that these can be applied to create both, more scalable and also more precise methods for the mapping of urban *Aedes Aegypti* abundance to support vector control. From our perspective advances that allow to capture small *Aedes Aegypti* breeding containers with low manual labeling effort are particularly promising. Especially the semi-supervised self-training (SSST) of object detection networks, as SSST can reach similar prediction performance as supervised methods but with less manual labels (Chuck Rosenberg et al., 2005).

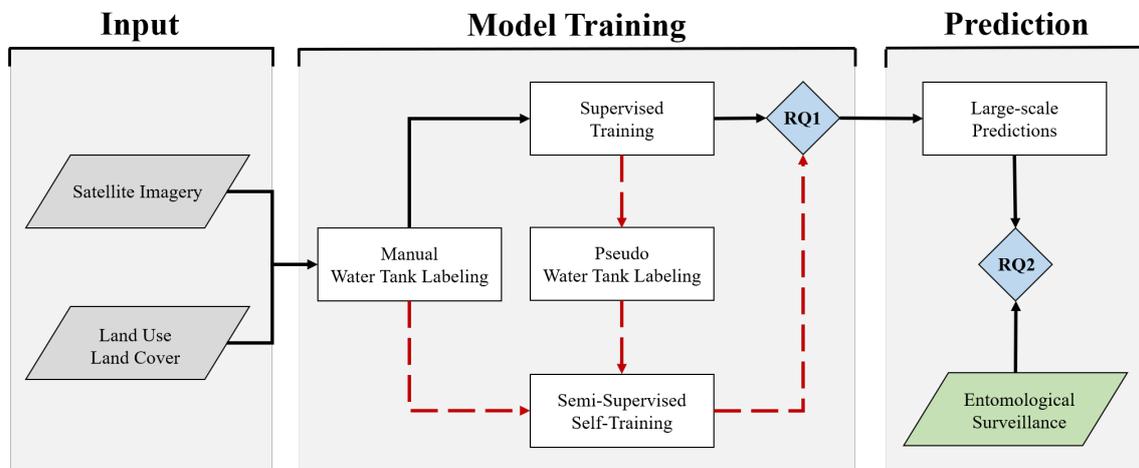
In this paper, we demonstrate the potential of SSST for the mapping of urban *Aedes Aegypti* distributions. We apply deep learning based object detection models on VHR satellite imagery and design a SSST-based fine-tuning algorithm to map the density of rainwater tanks, a typical *Aedes Aegypti* breeding spot in urban areas of the Global South. We evaluate water tank density as a high-resolution

proxy to model entomological surveillance data originating from mosquito traps. For our approach, we exclusively use open accessible data. This increases the applicability of our workflow for real-world scenarios. More specifically, we address the two following research questions:

- **RQ1:** To what extent can semi-supervised self-training outperform supervised learning methods with equal labeling effort for water tank detection?
- **RQ2:** How well does water tank density capture the observed inner-urban distribution of *Aedes Aegypti* in the case study?

## 2 Experimental Design

In order to answer the derived research questions, we propose a novel framework for the semi-automatic mapping of water tanks (cf. Figure 1). Our concept consists of mainly three parts: first open accessible input data to increase the transferability of our experiment, second the supervised and semi-supervised self-training of water tank detection models to evaluate the usefulness of SSST over supervised learning to support labor-intensive public health practices like entomological surveillance, and third large-scale water tank predictions to evaluate the significance of water tank density as a high resolution proxy to model inner-urban *Aedes Aegypti* distributions presumably useful for a more targeted planning of vector control interventions in the future.



**Figure 1:** Overview of the proposed framework to support future vector control including the required open accessible study data (grey), the semi-supervised self-training loop for water tank detection model fine-tuning (red), the evaluation of stated research questions (blue), and the ground truth evaluation set (green).

### 2.1 Materials

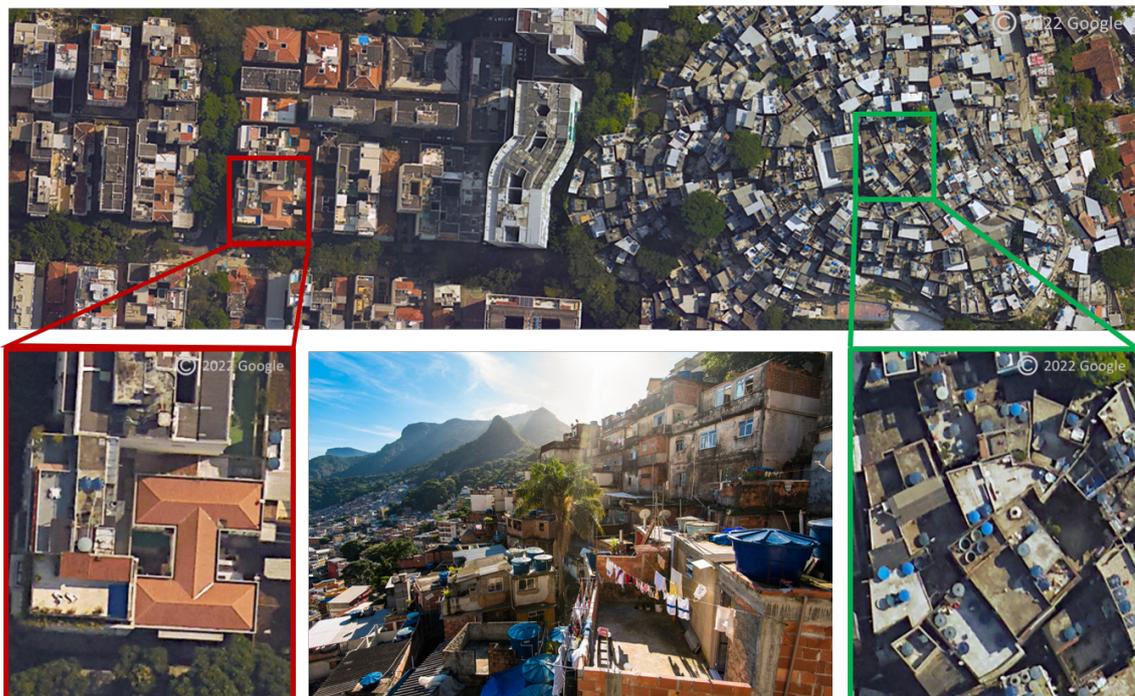
#### 2.1.1 Study site

We applied the proposed workflow to the city of Rio de Janeiro, which is one of the highest effected mega cities for mosquito borne diseases worldwide (Gibson et al., 2014; Wilson, 2011). The city belongs to the endemic regions for *Aedes*

*Aegypti* transmitted diseases due to its year long tropical climate (Franco dos Santos et al., 2022). With a population of around 6.75 million people and a high connectivity to other urban areas in Latin America, the second biggest city of Brazil has often been a starting point for larger uncontrolled disease outbreaks (Da Silva Jr. et al., 2002; Luiz Tadeu Moraes Figueiredo, 2004). The proximity of different types of urban structure, such as favela and other residential areas, and the topography of the city account for a high variability of possible *Aedes Aegypti* breeding sites. This makes the city of Rio de Janeiro an interesting use case for our proposed method.

### 2.1.2 Study object

Our study object are water tanks, often used for drinking water storage in the city of Rio de Janeiro and other Latin American cities. Water tanks are known to be one of the main breeding spots for *Aedes Aegypti* (Trewin et al., 2021). In the city of Rio de Janeiro they are part of vector control measures as well as entomological surveillance systems (Secretaria de Vigilância em Saúde, 2013). However, by far the majority of water tanks are not monitored due to the labor intense process. Since the locations of water tanks in the city of Rio de Janeiro are not mapped, it was so far not possible to investigate the relationship between water tank presence and mosquito abundance. Water tanks usually have a radius of 1 to 2 meters and are objects in the format of a cylinder with a approximated height of 1.5 meters. They appear dark blue, whereas some older ones might appear with brighter color due to sun bleaching (cf. Figure 2).



**Figure 2:** Rainwater tanks for water supply in the city of Rio de Janeiro occurring in high spatial variability due to different urban structure types like residential areas (bottom left) and favelas (bottom right).

The urban appearance of water tanks can be strongly correlated with the social structure of cities as shown by Cunha et. al. (Cunha et al., 2021). As expected

and further revealed by visual inspection, this was also the case for the city of Rio de Janeiro, where water tanks appeared more frequently in socially weaker parts of the city such as favelas due to the lack of piped water access. Thereby, the close proximity of favelas and other urban structures in the city of Rio de Janeiro leads presumably to a high spatial variance of water tanks. When observed, water tanks were primarily located in backyards and on rooftops. Due to their complex installation and heavy weight, it can be assumed that the position of water tanks does not vary much over time. All this makes the object detection of water tanks based on satellite imagery an interesting task to close the targeted research gaps.

### 2.1.3 Study data

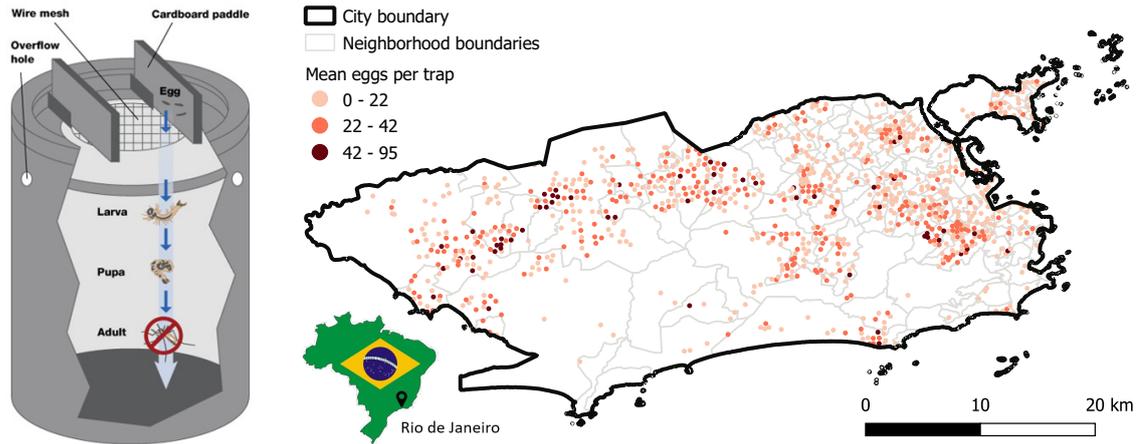
We used three data sets for our approach, namely: satellite imagery, land use land cover (LULC) maps, and entomological surveillance data (cf. Figure 1). Satellite imagery downloaded from the Microsoft Bing Tile Map Service (TMS) API (Microsoft, 2022) served as our main input data source. For the high resolution detection of water tanks we chose the highest available zoom level of 22 with an image resolution of 0.0373 meter per pixel. We retrieved 10,668,699 image patches of  $256 \times 256$  pixels. In addition, we used administrative LULC data (Municipality of Rio de Janeiro, 2022) to derive information about the location and size of different urban structures, which was used to stratify water tank labeling.

As evaluation data we had access to an entomological surveillance database from January 2019 until December 2021 provided by the health ministry of Rio de Janeiro. The purpose of entomological surveillance is to monitor the distribution and impact of vector control measures. The reliability of this data is highly affected by the spatial coverage and temporal execution frequency. Examples for entomological surveillance measurements are mosquito count or index data for various development stages of *Aedes Aegypti*: eggs, larvae, pupa, and adult. For our use case we used data collected with 1,207 ovitraps distributed around the city (cf. Figure 3). These are traps filled with water of around 20 cm radius used in the city of Rio de Janeiro for the collection of *Aedes Aegypti* eggs and larvae. When mature, the mosquitoes cannot escape these traps. The amount of eggs and larvae was collected on a monthly basis. As an evaluation indicator for our proposed method we used the “mean eggs per trap” (MET) rate aggregated over monthly time steps for *Aedes Aegypti*.

## 2.2 Methods

### 2.2.1 Manual water tank labeling

Since water tank locations were not available in sufficient amount and quality from open datasets, manual labeling was necessary. We manually labeled 2,000 water tanks in favelas and another 2,000 in residential areas for the training purpose. Labeling was done in QGIS (QGIS Development Team, 2022). The strata (favela/residential areas) were derived from the LULC map (Municipality of Rio de Janeiro, 2022). We labeled another 1,000 water tanks for validation of the model. Unlike the training labels, the labels for the test data were sampled across all kind of urban structures to analyze the robustness and generalization

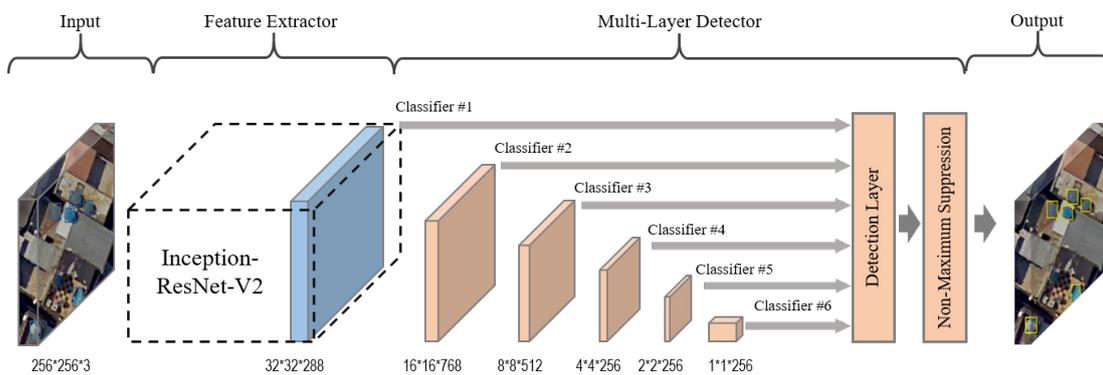


**Figure 3:** Sketch of the ovitrap used by the entomological surveillance system (left) and corresponding locations of the traps in the city of Rio de Janeiro. The color indicates the “mean eggs per trap” (MET) rate for *Aedes Aegypti* between January 2019 and December 2021 (right.)

of our object detection model. All manual labels including are provided in the supplementary material.

### 2.2.2 Supervised training

We trained a Single Shot Multibox Detection (SSD) Network (Liu et al., 2015) on the manual water tank labels. The SSD was retrieved from the TensorFlow Object Detection Model Zoo API (Google, 2022). SSD networks are single-stage object detector architectures that have been successfully applied for the detection and mapping of geospatial objects of diverse size and shape. In addition, they offer a good balance between training time and accuracy when compared to two-stage object detection networks like Faster R-CNNs as shown in Model Zoo (Google, 2022). The output of an SSD network is a list of predicted features and the corresponding probability scores.



**Figure 4:** Single-Stage Object Detection Network consisting of Inception-ResNet-V2 as feature extractor and multi layer detector with Non-Maximum Suppression layer as used for water tank detection models. The numbers at the bottom describes the dimension of the raster bands used at the different stages of the network. The output consists of bounding boxes for detected water tanks together with a confidence scale.

The SSD network we used consists of mainly two parts (see Figure 4). First a

feature extractor, which is in our case a backbone network with 164 layers, namely the Inception-ResNet-V2 network, allowing the model to learn deterministic features (e.g. colors and shapes) for common object detection tasks (Liu et al., 2015). Second a multi-layer detector together with a Non-Maximum Suppression (NMS) layer was used to create multi-scale detection boxes and to calculate the confidence scores of classification. This was needed for calculating the training loss. We used a SSD network pre-trained on the Microsoft COCO dataset (Lin et al., 2014) as the starting point for the model training to reduce training effort. The initial training process used 4,000 manual water tank labels and 20,000 training iterations. The corresponding training hyperparameters, also used for later model fine-tuning through semi-supervised self-training, were listed in Table 1.

**Table 1:** Hyperparameters used for training

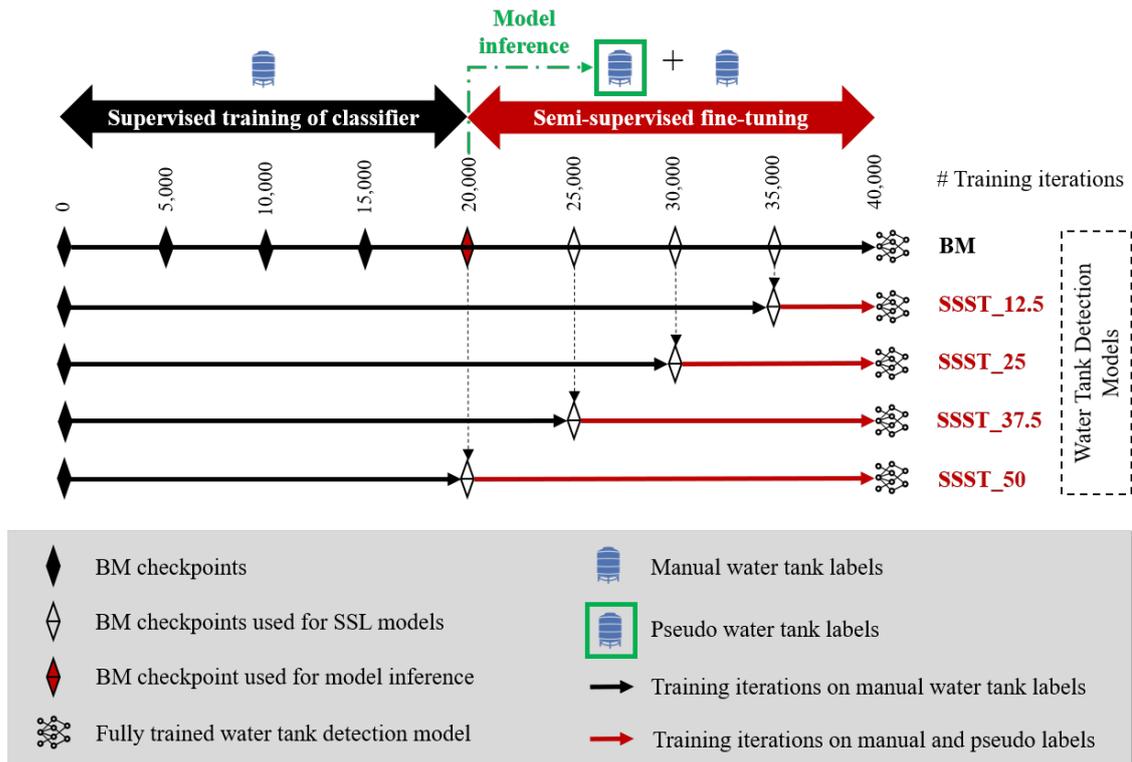
<b>Batch Size</b>	24
<b>Learning Rate</b>	0.0004
<b>IoU Threshold</b>	0.5
<b>Optimizer</b>	RMSprop
<b>Optimizer Momentum</b>	0.9
<b>Optimizer Decay</b>	0.9
<b>Optimizer Epsilon</b>	0.1

### 2.2.3 Pseudo water tank labeling

After the initial training process we generated additional pseudo water tank labels. Therefore, we applied the initially trained supervised model to predict water tanks in an unlabeled region. The selected region was approximately 15,000 ha and included all urban structure types, not just favelas and residential areas such as the manual water tank labeling process. Only water tank predictions with high confidence were used as additional pseudo labels to further fine-tune the water tank detection model. In our study, a confidence level of 80% was used as a lower threshold. This resulted in 10,800 additional water tank labels for the fine-tuning of our water tank detection model.

### 2.2.4 Semi-supervised self-training

The merged label set, combining the manual training labels and the pseudo labels, was used to fine-tune five new instances of our initial supervised model. The new instances differed in the number of semi-supervised training iterations in which the model was shown the additional pseudo-labels (cf. Figure 5). For consistency, all new fine-tuned instances were trained for further 20,000 iterations. This resulted in five water tank detection models, all trained for 40,000 training iterations using the same number of manual water tank labels. All models were built upon the same initial supervised water tank detection model used for model inference and thus pseudo label creation. One of the new instances were fine-tuned without using the pseudo labels. This model was named the base model (BM). It was trained to evaluate the change in model performance reached through self-training.



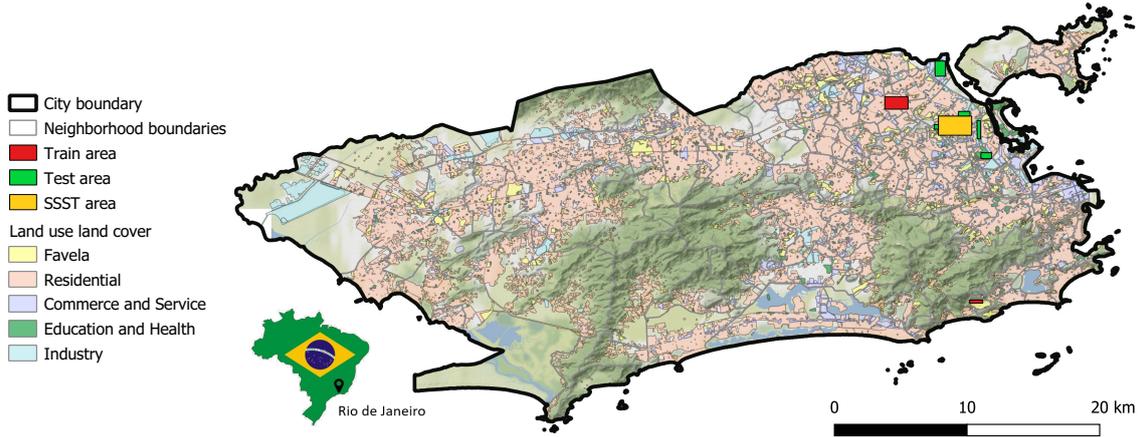
**Figure 5:** Overview of trained models, water tank label sets used for training, and corresponding length of training epochs. BM = supervised base model, SSST - semi-supervised self-trained models.

### 2.2.5 Evaluation of semi-supervised self-training

Performance between all five water tank detection models was evaluated based on precision, recall, and their harmonic mean, the F1-score. Precision is defined as the ratio of the true positive objects to all detected objects. Recall describes the fraction of relevant objects that are successfully retrieved. The performance indicators were calculated based on the comparison between the intersection of the bounding boxes of the predictions and of the validation labels. The level of agreement of the two boxes was based on the Intersection over Union (IoU) value. The IoU takes values between 0 and 1. For a value of 0 the two boxes do not overlap at all. For a value of 1 they overlap completely. An IoU Value of 0.5 or higher for a detected object was considered a true positive. An IoU value lower than 0.5 as a false positive. In order to measure generalization capabilities of the models, we compared the change in model performance, in terms of the F1 score, between test sets from urban structures used in the manual label set and urban structures not used in the manual label set.

### 2.2.6 Large-scale predictions

The model with the best F1-score was used to predict water tanks for the whole metropolitan area of Rio de Janeiro. The prediction used over 10 million satellite image patches in parallel tasks. For data management we used the map-proxy API (Omniscale GmbH & Co. KG, 2022). This allowed to store satellite imagery in subset folder structure. As for the detected water tanks, we pushed



**Figure 6:** The Land use land cover (LULC) in the city of Rio de Janeiro with train, test and semi-supervised self-training (SSST) region and topography in the background.

our predictions on each image patch to a PostGIS database. The database was then used for a post-processing step of filtering predictions by confidence scores.

### 2.2.7 Evaluation of water tank density as entomological proxy for *Aedes Aegypti*

To quantitatively evaluate our RQ2 of how well water tank density can capture the inner-urban distribution of *Aedes Aegypti* measured by entomological surveillance data, we ran a negative binomial generalized linear regression model (GLM) with a log-link function (Hilbe, 2012). The negative binomial GLM was selected as it allows the model to account for the overdispersion present in the count data. Corresponding equations are defined in Formula 1. As our response variable we selected the “mean eggs per trap” (MET) rate. As our explanatory variable we used water tank counts surrounding ovitrap locations in the range of 200 meters which corresponds to the estimated *Aedes Aegypti* flight range. We used all 1,207 ovitrap locations with information on the MET rate.

$$\begin{aligned}
 MET_i &\sim NB(\hat{\mu}_i, \hat{\theta}) \\
 \mathbb{E}(MET_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta} \\
 \text{Var}(MET_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta}^2 \\
 \log(\hat{\mu}_i) &= \hat{\beta}_0 + \hat{\beta}_1 * Watertank_i
 \end{aligned} \tag{1}$$

## 3 Results and Discussion

### 3.1 Comparison of training strategies

The results shown here are the outcome of the developed semi-supervised self-training approach for the large-scale detection of water tanks in the city of Rio de Janeiro (cf. Figure 1). The workflow consisted of three major configuration points: First the targeted selection of suitable SSST regions, secondly the choice of an appropriate confidence threshold for pseudo label filtering and thirdly the testing of various ratios of supervised and semi-supervised training iterations (cf. Figure 5). Sensitivity analysis for changing the first two configuration points were excluded from this result section. However, these parameters

were chosen carefully during experimental design to enable model improvements through self-training. During experimental design, SSST regions for model inference were chosen to be small enough to save computational time, but large enough to generate a sufficient amount of pseudo labels required to fine-tune the object detection model. In addition, SSST regions were selected to cover all urban structure types present in the city of Rio de Janeiro to ensure a robust object detection for large-scale predictions. The confidence threshold for the filtering of model inference outcomes and thus for the generation of pseudo labels was chosen with respect to the model performance of our supervised base model with 20,000 training iterations. The results for the third configuration point of the workflow, namely the variation of different ratios of supervised to semi-supervised training iterations, were analyzed in more detail and discussed in the following.

An increasing training time on the merged label set of manual and pseudo water tank labels continuously improved the F1-score of our object detection models (Table 2). The best water tank detection model was the model that used the additional pseudo labels for the longest SSST time (50% of the total 40,000 iteration, SSST-50) with an overall F1-score of 0.84 averaged over all test labels. This significant improvement of 22% compared to the supervised base model indicates a good balance of precision and recall. All SSST models showed a slightly decrease in recall compared to the supervised base model - i.e the proportion of correctly detected water tanks to the sum of all true water tanks decreased. However, this was countered by major improvements in precision, as the amount of correct water tank predictions on all predictions was higher for all four SSST models used.

**Table 2:** Performance metrics of trained water tank detection models. The SSST-12.5/25/37.5/50 model used the pseudo labels during 12.5/25/37.5/50% of the 40,000 training iteration respectively.

Models	Precision(%)	Recall (%)	F1	% F1 Improvement
<b>BM</b>	<b>0.59</b>	<b>0.85</b>	<b>0.69</b>	-
SSST-12.5	0.84	0.69	0.76	+10%
SSST-25	0.9	0.7	0.79	+14%
SSST-37.5	0.86	0.78	0.82	+19%
SSST-50	0.86	0.82	0.84	+22%

For the best (SSST-50) model, the relative increase in F1-score (cf. Table 3) was more obvious for urban structure types excluded in the manual label set (e.g. Commerce and Service, Education and Health, Industry) than for the urban structure types included in the manual label set (Favela, Residential). The F1-score improved, however, for all urban structure types. This makes the SSST-50 model more applicable for large-scale predictions than the supervised base model. These results were consistent with our expectations, namely that SSST models benefit from the additional knowledge collected by the machine itself, leading to more precise and robust water tank predictions across different urban structures relevant for large-scale predictions. The trained SSST-50 model is provided in the supplementary materials of this work.

However, we also identified several limitations in the results. First, not all urban structure types were used for model evaluation. We focused only on five of eleven land classes included in the LULC map where we expect human population

**Table 3:** Goodness of fit indicators for the base model and the best performing SSST model for different urban structures. The performance was based on independent test data points.

Method	BM			SSST-50			
	Precision(%)	Recall(%)	F1	Precision(%)	Recall(%)	F1	% F1 Improvement
Favela	0.63	0.85	0.73	0.87	0.78	0.83	+14%
Residential	0.59	0.86	0.70	0.81	0.82	0.82	+17%
Industry	0.51	0.85	0.64	0.8	0.80	0.8	+25%
Education and Health	0.54	0.93	0.68	0.91	0.91	0.91	+34%
Commerce and Service	0.66	0.76	0.71	0.90	0.78	0.84	+18%
Average weighted by instance	0.59	0.85	0.69	0.86	0.82	0.84	+22%

and thus the largest risk for infections by *Aedes Aegypti*. The second limitation results from the manual labeling process. We generated the test set on the basis of satellite imagery instead of a field study. Non-visible water tanks underneath shelters were thus not included in our test labels for model evaluation. However, we assume that a field study for the labeling of water tanks would not mitigate the achieved performance improvement of the semi-supervised self-training approach. Much more likely, it would have an impact on the absolute performance metrics, but to the same extent on those of the supervised BM as on those of the SSST models. The third limitation of our study is the low amount of manual training labels (4,000) compared to the amount of pseudo labels used for training (10,800). This implies a relatively high risk of an inappropriate training with potentially incorrect pseudo labels which can accumulate the error in the iterative self-training process. To reduce such a bias, one could either develop a debiased self-training algorithm similar to the one proposed by Chen et al. (2022) or apply co-training of classifiers originally proposed by Blum and Mitchell (1998).



**Figure 7:** Example for false negative, true negative, false positive, and true positive water tank predictions. Water tanks identified by the best performing SSST model are indicated by green bounding boxes together with the confidence of the prediction.

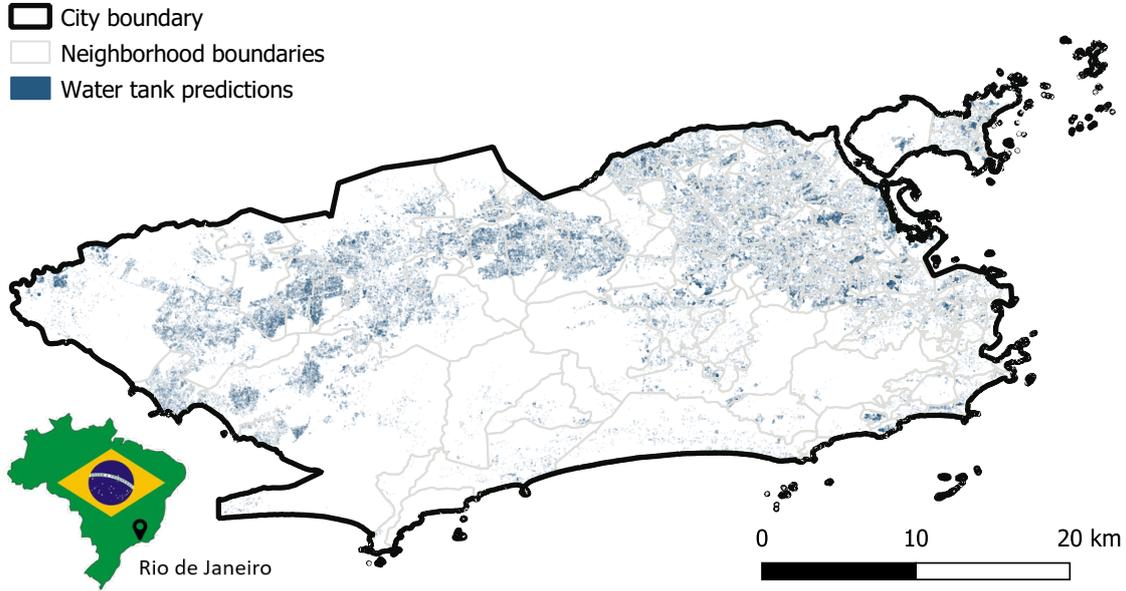
Further limitations of our study become apparent when visually inspecting raw prediction images of the SSST-50 model (cf. Figure 7). Common false negative predictions included water tanks in the shade or partial shade. To minimize the amount of these false negative predictions one could further fine-tune the SSST-50 model by feeding it with more shaded water tank labels. It is noteworthy that the number of objects in our study area which appear similar to water tanks was quite high resulting in high numbers of false positives. While similar objects such as blue cars and rooftop ventilators were rarely labeled as water tanks by our models, circular water pools or blue sunshades on beaches were frequently false positives. The false positive detection of water pools could be solved by applying a size filter. The detection of blue sunshades on beaches could be eliminated by applying an automatic land use map based filtering. However, these solution methods would only work to a limited extent. For very small water pools and blue sunshades not located on beaches this solution method becomes obsolete. Another solution would be the filtering of predictions by confidence score as applied during SSST training.

Further improvements of our models might be achieved by changing parameters of our semi-supervised self-training framework. This includes the size of the areas used for supervised model predictions to generate pseudo labels, the confidence threshold score applied for pseudo label filtering, the overall training time for object detection models, and the corresponding point for conducting the semi-supervised self-training loop. The training of a two-stage object detector like Faster-RCNNs as proposed by Cao et al. (2019) could also be an option for further investigations.

An alternative method for the reduction of manual labelling effort for object detection could be data augmentation. Data augmentation describes the technique of increasing the training set by creating slightly modified copies of provided training samples, for example by changing the rotation of the label (Shorten and Khoshgoftaar, 2019; van Dyk and Meng, 2001). It is a widely used method especially applied to avoid overfitting. However, for our use case of generating a robust model for large-scale predictions over various urban structure types, semi-supervised self-training seems to be more suitable. Instead of creating label copies, self-training can create completely new water tank labels (pseudo label) that can appear in different shape, color, and with varying shadow coverage. In addition, it allows to incorporate background features in the training process, like different rooftop types or water tank densities, not necessarily present in the limited manual label set used. All these additionally features gathered during pseudo label generation via self-training are extremely relevant, when trying to train robust object detector using Convolutional-Neural-Networks (CNNs). Especially for applying these models on over 10 million satellite image patches covering all types of urban structures. Semi-supervised self-training can avoid overfitting similar to data augmentation (Nartey et al., 2020). Of course, do both methods, data augmentation and self-training, allow a cost-sensitive creation of additional labels, which is relevant for our use case to minimize the manual labelling effort and associated cost and time. However, the capability of learning additionally background features, not present in the manual label set, is only possible through self-training in an automatic manner. Nevertheless, self-training requires a relatively high configuration effort to be successful compared to data augmentation

techniques as described in the beginning of this result section.

### 3.2 Modeling of urban mosquito abundance



**Figure 8:** Water tanks predicted by the best performing SSST model for the case study region. For orientation, the administrative boundaries of the neighborhoods are overlaid. The inset map to the lower left indicates the position of Rio de Janeiro in Brazil.

As a highlight of this work, water tanks predicted by the SSST-50 model were distributed throughout the whole metropolitan area of Rio de Janeiro with a high spatial variability (cf. Figure 8). The occurrence of water tanks was strongly dependent on inhabited areas. Forest areas were almost empty of water tanks. Water tank density within single neighborhoods also varied strongly. In addition to Figure 8, we provide a raster layer with a spatial resolution of 200 meters in the supplementary materials of this work. This consists of a raster value for water tank counts with confidence score above 90 percent and is intended to represent the spatial variance of water tanks at the resolution of an estimated *Aedes Aegypti* flight range for the whole city of Rio de Janeiro which could be used for urban mosquito modeling.

**Table 4:** Coefficients, standard errors, and p-values for negative binomial generalized linear regression model. Regression coefficients and standard errors are reported at the link scale. The dispersion parameter  $\theta$  indicates underdispersion.

	Estimate	Std. Error	p-value
Intercept	1.535	0.709	$< 2e - 16$
water tank count in 200 meter ovitrap buffer	0.058	0.014	$< 2e - 16$
$\theta$	0.649	0.027	-

The results of our negative binomial generalized linear regression model (cf. Table 4) indicated that water tank density was a highly significant proxy for modeling the *Aedes Aegypti* MET rate. This was in line with our expectations and

implicates that water tank density maps can be a useful indicator to enrich entomological surveillance data and thus support future vector control by providing more continuous and high resolution insights for urban mosquito distributions. The explained deviance for this regression model was 0.11. It was measured by Cohen's pseudo- $R^2$  (Cohen, 2013) (cf. Formula 2) indicating that about 11% of the deviance in the response are explained by the model. The deviance function of the negative binomial GLM captured the increasing variance with the mean that is typical for count data. The dispersion parameter captured thereby how strong the variance increases with the mean relative to a Poisson GLM, for which the variance equals the mean. The theta value of 0.649 corresponded to a significant overdispersion. This can be explained by the large number of zero values in the entomological dataset, which is why a negative binomial GLM was applied. Another reason for this is the low number of predictors used to model urban *Aedes Aegypti* distribution. However, other potentially relevant predictors have deliberately not been included in the model, which also explains the low value of the explained deviance. The addition of further explanatory variables is planned for follow-up activities.

$$\begin{aligned} \text{Cohen's pseudo } R^2 &= 1 - \frac{\text{model deviance}}{\text{null model deviance}} \\ \text{Negative binomial model deviance} &= 2 \sum (y \cdot \log(\frac{y}{\mu}) - (y + k - 1) \log(\frac{y + k - 1}{\mu + k - 1})) \end{aligned} \quad (2)$$

## 4 Conclusion

The emergence of open-accessible big spatial data in combination with modern computing technologies has great potential to revolutionize the treatment of emergent infectious diseases transmitted by *Aedes Aegypti*. Especially for those that are missing effective vaccines and are therefore treated mainly by local vector control, namely dengue, Zika and chikungunya which cause thousands of deaths each year. In this paper we demonstrated how deep learning based object detection models in combination with open-accessible satellite imagery can be applied to extract a fine-grained and informative proxy for the urban modelling of *Aedes Aegypti* distribution, namely water tanks. Such models are essential to derive more targeted vector control interventions, allow cost savings in entomological surveillance, and most importantly a more efficient overall disease control. The results of this paper indicate that the burden of manual labeling necessary for large-scale and robust water tank detection can be substantially reduced by the development of a semi-supervised self-training workflow without compromising model accuracy. This increases not only real world applicability of our water tank detection model, but also its robust transferability to other *Aedes Aegypti* endemic cities where water tanks are common mosquito breeding sites. The measured significance in the association between water tank density and abundance of *Aedes Aegypti* showed the potential of the generated indicator to augment entomological surveillance gaps that occur when limited mosquito flight ranges are considered. The developed urban-specific indicator can thus bring novel insights into the high spatial variability of urban *Aedes Aegypti* distributions that can hardly be explained by commonly used low resolution features for *Aedes Aegypti* mapping. However, as the abundance of *Aedes Aegypti* depends on other predictors such as climatic conditions, upcoming research will explore the predictive

power of water tank density in combination with these indicators. With these combined models for the fine-scale mapping of *Aedes Aegypti* distributions we hope to reveal hidden patterns not only with regard to urban *Aedes Aegypti* populations, but also for inner-urban pathogen transmission for dengue, Zika, and chikungunya. With these major contributions of our interdisciplinary research we hope to create new pathways for the science of computational eco-epidemiology and provide useful data sets as well as methods to public health authorities especially in the city of Rio de Janeiro, Brazil.

## Data statement

The supplementary data to this article can be found online at: <https://heidata.uni-heidelberg.de/privaturl.xhtml?token=fbcd51cb-af56-48b7-8450-46eeb8880fae>. It contains our best water tank detection model, manual water tank labels used for training and testing as well as a high-resolution raster of water tank density for the city of Rio de Janeiro.

## Acknowledgements

The authors would like to take this opportunity to thank the editors and reviewers for their valuable comments and suggestions. The entomological data was provided by the health ministry of the city of Rio de Janeiro. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) [grant number 451956976]. Sven Lautenbach acknowledges funding by the Klaus-Tschira Stiftung, Germany. The authors acknowledge support by the High Performance and Cloud Computing Group at the Center for Data Processing of the University of Tübingen, the state of Baden-Württemberg through bwHPC, and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Credit authorship contribution statement

Steffen Knoblauch: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing - original draft, writing - review & editing, visualization, supervision, project administration, funding acquisition. Hao Li: conceptualization, software, writing - review & editing. Sven Lautenbach: conceptualization, writing - review & editing, supervision, project administration, funding acquisition. Yara Elshiaty: software. Antônio A. de A. Rocha: resources. Bernd Resch: funding acquisition. Dorian Arifi: funding acquisition. Thomas Jänisch: funding acquisition. Ivonne Morales: funding

acquisition. Alexander Zipf: supervision, project administration, funding acquisition.



# Bibliography

- Agarwal, A., Chaudhuri, U., Chaudhuri, S., and Seetharaman, G. (2014). “Detection of potential mosquito breeding sites based on community sourced geotagged images”. DOI: 10.1117/12.2058121.
- Amorim, J. H. and Birbrair, A. (2022). “Dengue vaccines: where are we now and where we are going?” *The Lancet Infectious Diseases* 22.6, pp. 756–757. DOI: 10.1016/S1473-3099(21)00753-2.
- Andersson, V. O., Ferreira Birck, M. A., and Araujo, R. M. (2018). “Towards Predicting Dengue Fever Rates Using Convolutional Neural Networks and Street-Level Images”. *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8. DOI: 10.1109/IJCNN.2018.8489567.
- Banerjee, S., Aditya, G., and Saha, G. K. (2015). “Household Wastes as Larval Habitats of Dengue Vectors: Comparison between Urban and Rural Areas of Kolkata, India”. *PLoS one* 10.10, e0138082. DOI: 10.1371/journal.pone.0138082.
- Blum, A. and Mitchell, T. (1998). “Combining labeled and unlabeled data with co-training”. *Proceedings of the eleventh annual conference on Computational learning theory - COLT' 98*. Ed. by P. Bartlett and Y. Mansour. New York, New York, USA: ACM Press, pp. 92–100. DOI: 10.1145/279943.279962.
- Bomfim, R., Pei, S., Shaman, J., Yamana, T., Makse, H. A., Andrade, J. S., Lima Neto, A. S., and Furtado, V. (2020). “Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas”. *Journal of the Royal Society, Interface* 17.171, p. 20200691. DOI: 10.1098/rsif.2020.0691.
- Boser, A., Sousa, D., Larsen, A., and MacDonald, A. (2021). “Micro-climate to macro-risk: mapping fine scale differences in mosquito-borne disease risk using remote sensing”. *Environmental Research Letters* 16.12, p. 124014. DOI: 10.1088/1748-9326/ac3589.
- Bowman, L. R., Runge-Ranzinger, S., and McCall, P. J. (2014). “Assessing the relationship between vector indices and dengue transmission: a systematic review of the evidence”. *PLoS neglected tropical diseases* 8.5, e2848. DOI: 10.1371/journal.pntd.0002848.
- Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y., and Wu, Z. (2019). “An Improved Faster R-CNN for Small Object Detection”. *IEEE Access* 7, pp. 106838–106846. DOI: 10.1109/ACCESS.2019.2932731.
- Caputo, B., Manica, M., Filipponi, F., Blangiardo, M., Cobre, P., Delucchi, L., Marco, C. M. de, Iesu, L., Morano, P., Petrella, V., Salvemini, M., Bianchi, C., and Della Torre, A. (2020). “ZanzaMapp: A Scalable Citizen Science Tool to Monitor Perception of Mosquito Abundance and Nuisance in Italy and Beyond”. *International journal of environmental research and public health* 17.21. DOI: 10.3390/ijerph17217872.
- Chang, A. Y., Parrales, M. E., Jimenez, J., Sobieszczyk, M. E., Hammer, S. M., Copenhaver, D. J., and Kulkarni, R. P. (2009). “Combining Google Earth and GIS map-

- ping technologies in a dengue surveillance system for developing countries”. *International journal of health geographics* 8, p. 49. DOI: 10.1186/1476-072X-8-49.
- Chavasse, D. (2002). “Know your enemy: Some facts about the natural history of Malawi’s Anopheles mosquitoes and implications for malaria control”. 14.1, pp. 7–8.
- Chen, B., Jiang, J., Wang, X., Wan, P., Wang, J., and Long, M. (2022). “Debiased Self-Training for Semi-Supervised Learning”. *arXiv*. DOI: 10.48550/arXiv.2202.07136.
- Cho, H., Low, R. D., Fischer, H. A., and Storksdieck, M. (2021). “The STEM Enhancement in Earth Science “Mosquito Mappers” Virtual Internship: Outcomes of Place-Based Engagement with Citizen Science”. *Frontiers in Environmental Science* 9. DOI: 10.3389/fenvs.2021.682669.
- Chuck Rosenberg, Martial Hebert, and Henry Schneiderman (2005). “Semi-Supervised Self-Training of Object Detection Models”. DOI: 10.1184/R1/6560834.v1.
- Cohen (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. DOI: 10.4324/9780203774441.
- Colón-González, F. J., Sewe, M. O., Tompkins, A. M., Sjödin, H., Casallas, A., Rocklöv, J., Caminade, C., and Lowe, R. (2021). “Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study”. *The Lancet Planetary Health* 5.7, e404–e414. DOI: 10.1016/S2542-5196(21)00132-7.
- Cunha, H. S., Schlauser, B. S., Wildemberg, P. F., Fernandes, E. A. M., Dos Santos, J. A., Lage, M. d. O., Lorenz, C., Barbosa, G. L., Quintanilha, J. A., and Chiaravalloti-Neto, F. (2021). “Water tank and swimming pool detection based on remote sensing and deep learning: Relationship with socioeconomic level and applications in dengue control”. *PloS one* 16.12, e0258681. DOI: 10.1371/journal.pone.0258681.
- Da Queiroz, E. R. S. and Medronho, R. d. A. (2022). “Overlap between dengue, Zika and chikungunya hotspots in the city of Rio de Janeiro”. *PloS one* 17.9, e0273980. DOI: 10.1371/journal.pone.0273980.
- Da Silva Jr., J. B., Siqueira Jr., J. B., Coelho, G. E., Vilarinhos, P. T., and Pimenta Jr., F. G. (2002). “Dengue in Brazil: current situation and prevention and control activities”.
- Dandabathula, G. (2019). “AUTOMATIC DETECTION OF OVERHEAD WATER TANKS FROM SATELLITE IMAGES USING FASTER-RCNN”. *International Journal of Advanced Research in Computer Science* 10.5, pp. 34–37. DOI: 10.26483/ijarcs.v10i5.6466.
- David, M. R., Lourenço-de-Oliveira, R., and Freitas, R. M. de (2009). “Container productivity, daily survival rates and dispersal of *Aedes aegypti* mosquitoes in a high income dengue epidemic neighbourhood of Rio de Janeiro: presumed influence of differential urban structure on mosquito biology”. *Memorias do Instituto Oswaldo Cruz* 104.6, pp. 927–932. DOI: 10.1590/S0074-02762009000600019.
- Dias, T. M., Alves, V. C., Alves, H. M., Pinheiro, L. F., Pontes, R., Araujo, G. M., Lima, A. A., and Prego, T. M. (2018). “Autonomous Detection of Mosquito-Breeding Habitats Using an Unmanned Aerial Vehicle”, pp. 351–356. DOI: 10.1109/LARS/SBR/WRE.2018.00070.
- Ebi, K. L. and Nealon, J. (2016). “Dengue in a changing climate”. *Environmental research* 151, pp. 115–123. DOI: 10.1016/j.envres.2016.07.026.
- European Centre for Disease Prevention and Control (2016). *Aedes aegypti - Factsheet for experts*. Ed. by European Centre for Disease Prevention and Control.
- Fernandes, E., Wildemberg, P., and Dos Santos, J. (2020). “Water Tanks and Swimming Pools Detection in Satellite Images: Exploiting Shallow and Deep-Based Strategies”, pp. 117–122. DOI: 10.5753/wvc.2020.13491.

- Franco dos Santos, S. A., Karam, H. A., Filho, A. J. P., Da Silva, J. C. B., Rojas, J. L. F., Suazo, J. M. A., Panduro, I. L. V., and Peña, C. A. S. (2022). “Dengue Climate Variability in Rio de Janeiro City with Cross-Wavelet Transform”. *Journal of Environmental Protection* 13.03, pp. 261–276. DOI: 10.4236/jep.2022.133016.
- Getachew, D., Tekie, H., Gebre-Michael, T., Balkew, M., and Mesfin, A. (2015). “Breeding Sites of *Aedes aegypti*: Potential Dengue Vectors in Dire Dawa, East Ethiopia”. *Interdisciplinary perspectives on infectious diseases* 2015, p. 706276. DOI: 10.1155/2015/706276.
- Gibson, G., Souza-Santos, R., Pedro, A. S., Honório, N. A., and Sá Carvalho, M. (2014). “Occurrence of severe dengue in Rio de Janeiro: an ecological study”. *Revista da Sociedade Brasileira de Medicina Tropical* 47.6, pp. 684–691. DOI: 10.1590/0037-8682-0223-2014.
- Google (2022). *TensorFlow Object Detection Model Zoo*.
- Gwitira, I., Murwira, A., Zengeya, F. M., and Shekede, M. D. (2018). “Application of GIS to predict malaria hotspots based on *Anopheles arabiensis* habitat suitability in Southern Africa”. *International Journal of Applied Earth Observation and Geoinformation* 64, pp. 12–21. DOI: 10.1016/j.jag.2017.08.009.
- Haas-Stapleton, E. J., Barretto, M. C., Castillo, E. B., Clausnitzer, R. J., and Ferdan, R. L. (2019). “Assessing Mosquito Breeding Sites and Abundance Using An Unmanned Aircraft”. *Journal of the American Mosquito Control Association* 35.3, pp. 228–232. DOI: 10.2987/19-6835.1.
- Haddawy, P., Wettayakorn, P., Nonthaleerak, B., Su Yin, M., Wiratsudakul, A., Schöning, J., Laosiritaworn, Y., Balla, K., Euaungkanakul, S., Quengdaeng, P., Choknitipakin, K., Traivijitkhun, S., Erawan, B., and Kraising, T. (2019). “Large scale detailed mapping of dengue vector breeding sites using street view images”. *PLoS neglected tropical diseases* 13.7, e0007555. DOI: 10.1371/journal.pntd.0007555.
- Harrington, L. C., Scott, T. W., Lerdthusnee, K., Coleman, R. C., Costero, A., Clark, G. G., Jones, J. J., Kitthawee, S., Kittayapong, P., Sithiprasasna, R., and Edman, J. D. (2005). “Dispersal of the dengue vector *Aedes aegypti* within and between rural communities”. *The American journal of tropical medicine and hygiene* 72.2, pp. 209–220. DOI: 10.4269/ajtmh.2005.72.209.
- Hilbe, J. M. (2012). *Negative Binomial Regression*. Cambridge University Press. DOI: 10.1017/CB09780511973420.
- Hladish, T. J., Pearson, C. A. B., Toh, K. B., Rojas, D. P., Manrique-Saide, P., Vazquez-Prokopec, G. M., Halloran, M. E., and Longini, I. M. (2020). “Designing effective control of dengue with combined interventions”. *Proceedings of the National Academy of Sciences of the United States of America* 117.6, pp. 3319–3325. DOI: 10.1073/pnas.1903496117.
- Honório, N. A., Da Silva, W. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003). “Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil”. *Memorias do Instituto Oswaldo Cruz* 98.2, pp. 191–198. DOI: 10.1590/s0074-02762003000200005.
- Kantor, I. N. (2018). “Dengue, zika, chikungunya y el desarrollo de vacunas”. *Medicina* 78.1, pp. 23–28.
- Knerer, G., Currie, C. S. M., and Brailsford, S. C. (2020). “The economic impact and cost-effectiveness of combined vector-control and dengue vaccination strategies in Thailand: results from a dynamic transmission model”. *PLoS neglected tropical diseases* 14.10, e0008805. DOI: 10.1371/journal.pntd.0008805.

- Limkittikul, K., Brett, J., and L’Azou, M. (2014). “Epidemiological trends of dengue disease in Thailand (2000-2011): a systematic literature review”. *PLoS neglected tropical diseases* 8.11, e3241. DOI: 10.1371/journal.pntd.0003241.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). “Microsoft COCO: Common Objects in Context”. DOI: 10.48550/arXiv.1405.0312.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2015). “SSD: Single Shot MultiBox Detector”. DOI: 10.48550/arXiv.1512.02325.
- Lobo, N. F., Achee, N. L., Greico, J., and Collins, F. H. (2018). “Modern Vector Control”. *Cold Spring Harbor perspectives in medicine* 8.1. DOI: 10.1101/cshperspect.a025643.
- Lorenz, C., Chiaravalloti-Neto, F., Oliveira Lage, M. de, Quintanilha, J. A., Parra, M. C., Dibo, M. R., Fávoro, E. A., Guirado, M. M., and Nogueira, M. L. (2020). “Remote sensing for risk mapping of *Aedes aegypti* infestations: Is this a practical task?” *Acta tropica* 205, p. 105398. DOI: 10.1016/j.actatropica.2020.105398.
- Louis, V. R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y., and Dambach, P. (2014). “Modeling tools for dengue risk mapping - a systematic review”. *International journal of health geographics* 13, p. 50. DOI: 10.1186/1476-072X-13-50.
- Low, R. D., Nelson, P. V., Soeffing, C., and Clark, A. (2021). “Adopt a Pixel 3 km: A Multiscale Data Set Linking Remotely Sensed Land Cover Imagery With Field Based Citizen Science Observation”. *Frontiers in Climate* 3. DOI: 10.3389/fc1im.2021.658063.
- Low, R. D., Schwerin, T. G., Boger, R. A., Soeffing, C., Nelson, P. V., Bartlett, D., Ingle, P., Kimura, M., and Clark, A. (2022). “Building International Capacity for Citizen Scientist Engagement in Mosquito Surveillance and Mitigation: The GLOBE Program’s GLOBE Observer Mosquito Habitat Mapper”. *Insects* 13.7. DOI: 10.3390/insects13070624.
- Luiz Tadeu Moraes Figueiredo (2004). *Dengue in Brazil: Past, Present and Future Perspective - Dengue bulletin*. Vol. 27. WHO Regional Office for South-East Asia.
- Machault, V., Yébakima, A., Etienne, M., Vignolles, C., Palany, P., Tourre, Y., Guérécheau, M., and Lacaux, J.-P. (2014). “Mapping Entomological Dengue Risk Levels in Martinique Using High-Resolution Remote-Sensing Environmental Data”. *ISPRS International Journal of Geo-Information* 3.4, pp. 1352–1371. DOI: 10.3390/ijgi3041352.
- McFeeters, S. (2013). “Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach”. *Remote Sensing* 5.7, pp. 3544–3561. DOI: 10.3390/rs5073544.
- Mehra, M., Bagri, A., Jiang, X., and Ortiz, J. (2016). “Image Analysis for Identifying Mosquito Breeding Grounds”. *2016 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*. IEEE, pp. 1–6. DOI: 10.1109/SECONW.2016.7746808.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. (2019). “The current and future global distribution and population at risk of dengue”. *Nature microbiology* 4.9, pp. 1508–1515. DOI: 10.1038/s41564-019-0476-8.
- Microsoft (2022). *Zoom levels and tile grid*.

- Morrison, A. C., Gray, K., Getis, A., Astete, H., Sihuincha, M., Focks, D., Watts, D., Stancil, J. D., Olson, J. G., Blair, P., and Scott, T. W. (2004). “Temporal and geographic patterns of *Aedes aegypti* (Diptera: Culicidae) production in Iquitos, Peru”. *Journal of medical entomology* 41.6, pp. 1123–1142. DOI: 10.1603/0022-2585-41.6.1123.
- Municipality of Rio de Janeiro (2022). *Land use data*. Ed. by Rio de Janeiro.
- Muñoz, Á. G., Chourio, X., Rivière-Cinnamond, A., Diuk-Wasser, M. A., Kache, P. A., Mordecai, E. A., Harrington, L., and Thomson, M. C. (2020). “AeDES: a next-generation monitoring and forecasting system for environmental suitability of *Aedes*-borne disease transmission”. *Scientific reports* 10.1, p. 12640. DOI: 10.1038/s41598-020-69625-4.
- Nartey, O. T., Yang, G., Wu, J., and Asare, S. K. (2020). “Semi-Supervised Learning for Fine-Grained Classification With Self-Training”. *IEEE Access* 8, pp. 2109–2121. DOI: 10.1109/ACCESS.2019.2962258.
- Omniscale GmbH & Co. KG (2022). *MapProxy*. Ed. by Omniscale GmbH & Co. KG.
- Pan American Health Organization (2019). “Technical document for the implementation of interventions based on generic operational scenarios for *Aedes aegypti* control; 2019”.
- Paploski, I. A. D., Rodrigues, M. S., Mugabe, V. A., Kikuti, M., Tavares, A. S., Reis, M. G., Kitron, U., and Ribeiro, G. S. (2016). “Storm drains as larval development and adult resting sites for *Aedes aegypti* and *Aedes albopictus* in Salvador, Brazil”. *Parasites & vectors* 9.1, p. 419. DOI: 10.1186/s13071-016-1705-0.
- Passos, W. L., Araujo, G. M., Lima, A. A. de, Netto, S. L., and Da Silva, E. A. B. (2020). “Automatic Detection of *Aedes aegypti* Breeding Grounds Based on Deep Networks with Spatio-Temporal Consistency”. DOI: 10.48550/arXiv.2007.14863.
- QGIS Development Team (2022). *QGIS*.
- Rocklöv, J. and Dubrow, R. (2020). “Climate change: an enduring challenge for vector-borne disease prevention and control”. *Nature immunology* 21.5, pp. 479–483. DOI: 10.1038/s41590-020-0648-y.
- Runge-Ranzinger, S., McCall, P. J., Kroeger, A., and Horstick, O. (2014). “Dengue disease surveillance: an updated systematic literature review”. *Tropical medicine & international health : TM & IH* 19.9, pp. 1116–1160. DOI: 10.1111/tmi.12333.
- Sallam, M. F., Fizer, C., Pilant, A. N., and Whung, P.-Y. (2017). “Systematic Review: Land Cover, Meteorological, and Socioeconomic Determinants of *Aedes* Mosquito Habitat for Risk Mapping”. *International journal of environmental research and public health* 14.10. DOI: 10.3390/ijerph14101230.
- Schenkel, J., Taele, P., Goldberg, D., Horney, J., and Hammond, T. (2020). “Identifying Potential Mosquito Breeding Grounds: Assessing the Efficiency of UAV Technology in Public Health”. *Robotics* 9.4, p. 91. DOI: 10.3390/robotics9040091.
- Schrauf, S., Tschismarov, R., Tauber, E., and Ramsauer, K. (2020). “Current Efforts in the Development of Vaccines for the Prevention of Zika and Chikungunya Virus Infections”. *Frontiers in immunology* 11, p. 592. DOI: 10.3389/fimmu.2020.00592.
- Schwartz, L. M., Halloran, M. E., Durbin, A. P., and Longini, I. M. (2015). “The dengue vaccine pipeline: Implications for the future of dengue control”. *Vaccine* 33.29, pp. 3293–3298. DOI: 10.1016/j.vaccine.2015.05.010.
- Secretaria de Vigilância em Saúde (2013). “Levantamento Rápido de Índices para *Aedes Aegypti* (LIRAA) para vigilância entomológica do *Aedes*”.
- Semenza, J. C., Rocklöv, J., and Ebi, K. L. (2022). “Climate Change and Cascading Risks from Infectious Disease”. *Infectious diseases and therapy* 11.4, pp. 1371–1390. DOI: 10.1007/s40121-022-00647-3.

- Shorten, C. and Khoshgoftaar, T. M. (2019). “A survey on Image Data Augmentation for Deep Learning”. *Journal of Big Data* 6.1. DOI: 10.1186/s40537-019-0197-0.
- Su Yin, M., Bicout, D. J., Haddawy, P., Schöning, J., Laosiritaworn, Y., and Sa-Anghchai, P. (2021). “Added-value of mosquito vector breeding sites from street view images in the risk mapping of dengue incidence in Thailand”. *PLoS neglected tropical diseases* 15.3, e0009122. DOI: 10.1371/journal.pntd.0009122.
- Taborda, A., Chamorro, C., Quintero, J., Carrasquilla, G., and Londoño, D. (2022). “Cost-effectiveness of a Dengue Vector Control Intervention in Colombia”. *The American journal of tropical medicine and hygiene*. DOI: 10.4269/ajtmh.20-0669.
- Trewin, B. J., Parry, H. R., Pagendam, D. E., Devine, G. J., Zalucki, M. P., Darbro, J. M., Jansen, C. C., and Schellhorn, N. A. (2021). “Simulating an invasion: unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, *Aedes aegypti*, in Brisbane, Australia”. *Biological invasions* 23.12, pp. 3891–3906. DOI: 10.1007/s10530-021-02619-z.
- Uusitalo, R., Siljander, M., Culverwell, C. L., Mutai, N. C., Forbes, K. M., Vapalahti, O., and Pellikka, P. K. (2019). “Predictive mapping of mosquito distribution based on environmental and anthropogenic factors in Taita Hills, Kenya”. *International Journal of Applied Earth Observation and Geoinformation* 76, pp. 84–92. DOI: 10.1016/j.jag.2018.11.004.
- van Dyk, D. A. and Meng, X.-L. (2001). “The Art of Data Augmentation”. *Journal of Computational and Graphical Statistics* 10.1, pp. 1–50. DOI: 10.1198/10618600152418584.
- Vasconcelos, D., Yin, M. S., Wetjen, F., Herbst, A., Ziemer, T., Förster, A., Barkowsky, T., Nunes, N., and Haddawy, P. (2021). “Counting Mosquitoes in the Wild”. *Proceedings of the Conference on Information Technology for Social Good*. New York, NY, USA: ACM, pp. 43–48. DOI: 10.1145/3462203.3475914.
- Vezzani, D. (2007). “Review: artificial container-breeding mosquitoes and cemeteries: a perfect match”. *Tropical medicine & international health : TM & IH* 12.2, pp. 299–313. DOI: 10.1111/j.1365-3156.2006.01781.x.
- WHO (2012). *Global strategy for dengue prevention and control, 2012-2020*. Geneva, Switzerland: World Health Organization.
- WHO (2017). “Global Vector Control Response 2017-2030”.
- WHO (2022). *Vector-borne diseases*. Ed. by WHO.
- Wilke, A. B. B., Vasquez, C., Carvajal, A., Medina, J., Chase, C., Cardenas, G., Mutebi, J.-P., Petrie, W. D., and Beier, J. C. (2020). “Proliferation of *Aedes aegypti* in urban environments mediated by the availability of key aquatic habitats”. *Scientific reports* 10.1, p. 12925. DOI: 10.1038/s41598-020-69759-5.
- Wilson, A. L., Courtenay, O., Kelly-Hope, L. A., Scott, T. W., Takken, W., Torr, S. J., and Lindsay, S. W. (2020). “The importance of vector control for the control and elimination of vector-borne diseases”. *PLoS neglected tropical diseases* 14.1, e0007831. DOI: 10.1371/journal.pntd.0007831.
- Wilson, M. E. (2011). “Chapter 4 Megacities and Emerging Infections: Case Study of Rio de Janeiro, Brazil”. *Megacities & Global Health*. Ed. by O. A. Khan and G. Pappas. American Public Health Association. DOI: 10.2105/9780875530031ch04.
- Youssefi, F., Javad Valadan Zoej, M., Ali Hanafi-Bojd, A., Borahani Dariane, A., Khaki, M., and Safdarinezhad, A. (2022). “Predicting the location of larval habitats of *Anopheles* mosquitoes using remote sensing and soil type data”. *International Journal of Applied Earth Observation and Geoinformation* 108, p. 102746. DOI: 10.1016/j.jag.2022.102746.
- Yukich, J. O., Lengeler, C., Tediosi, F., Brown, N., Mulligan, J.-A., Des Chavasse, Stevens, W., Justino, J., Conteh, L., Maharaj, R., Erskine, M., Mueller, D. H.,

---

Wiseman, V., Ghebremeskel, T., Zerom, M., Goodman, C., McGuire, D., Urrutia, J. M., Sakho, F., Hanson, K., and Sharp, B. (2008). “Costs and consequences of large-scale vector control for malaria”. *Malaria journal* 7, p. 258. DOI: 10.1186/1475-2875-7-258.



## Publication II: High-resolution Mapping of Urban *Aedes aegypti* Immature Abundance through Breeding Site De- tection based on Satellite and Street View Im- agery

**Authors.** Steffen Knoblauch, Myat Su Yin, Krittin Chatrinan, Antonio Rocha, Peter Haddawy, Filip Biljecki, Sven Lautenbach, Bernd Resch, Dorian Arifi, Thomas Jänisch, Ivonne Morales, Alexander Zipf

**Outlet.** *Scientific Reports: Urbanization and Communicable Diseases*

**Status.** Published on 2024-08-06 under CC-BY 4.0

### Abstract.

Identification of *Aedes aegypti* breeding hotspots is essential for the implementation of targeted vector control strategies and thus the prevention of several mosquito-borne diseases worldwide. Training computer vision models on satellite and street view imagery in the municipality of Rio de Janeiro, we analyzed the correlation between the density of common breeding grounds and *Aedes aegypti* infestation measured by ovitraps on a monthly basis between 2019 to 2022. Our findings emphasized the significance ( $p \leq 0.05$ ) of micro-habitat proxies generated through object detection, allowing to explain high spatial variance in urban abundance of *Aedes aegypti* immatures. Water tanks, non-mounted car tires, plastic bags, potted plants, and storm drains positively correlated with *Aedes aegypti* egg and larva counts considering a 1,000 m mosquito flight range buffer around 2,700 ovitrap locations, while dumpsters, small trash bins, and large trash bins exhibited a negative association. This complementary application of satellite and street view imagery opens the pathway for high-resolution interpolation of entomological surveillance data and has the potential to optimize vector control strategies. Consequently it supports the mitigation of emerging infectious diseases transmitted by *Aedes aegypti*, such as dengue, chikungunya, and Zika, which cause thousands of deaths each year.

**Keywords.** *Aedes aegypti* · Rio de Janeiro · Satellite · Street view · Object detection · Ovitrap

## 1 Introduction

The mosquito species *Aedes aegypti* is responsible for transmitting several communicable diseases, such as dengue, yellow fever, chikungunya, and Zika (Wilke et al., 2020). It has become an increasing global threat due to environmental changes associated with climate change, urban growth, and resistance to insecticides (Messina et al., 2019; Semenza et al., 2022). Dengue fever alone accounted for 390 million infections worldwide in 2020, marking a 30-fold increase over the last fifty years (Ebi and Nealon, 2016; Glassman et al., 2022). For this reason, numerous attempts have been made to enhance entomological surveillance methods for *Aedes aegypti* in order to predict patterns of potential disease outbreaks and conduct more targeted vector control (Louis et al., 2014; Sallam et al., 2017). However, the bioecology of *Aedes aegypti* turns the development of

accurate monitoring techniques into a challenging task. *Aedes aegypti* is an urban favouring mosquito that breeds in small artificial water containers such as potted plants, and trash, which are often of ephemeral nature. This, combined with the bioecological assumption about a limited *Aedes aegypti* flight range of below 1,000 m without the assistance of wind (David et al., 2009; Honório et al., 2003), can result in a high spatial variability of abundance. High spatial variability is challenging to capture with traditional sample-based entomological field surveys (Knoblauch et al., 2023). The financial cost of such labor-intensive surveillance methods is also substantial, underscoring the urgent need for alternative mapping solutions, especially for urban areas of *Aedes aegypti*-endemic countries in the Global South where most infections occur (Bhatt et al., 2013).

The increasing availability of openly accessible big spatial data, in combination with modern computing technologies, can help address these issues (Knoblauch and Moritz, 2023). Digital techniques enable the large-scale interpolation of entomological surveillance data at a low financial cost. This enables the extrapolation of knowledge gathered from entomological sample locations into a continuous space, considering micro-scale changes in urban circumstances and the constrained flight range of mosquitoes. Consequently, these advancements could optimize the allocation of vector control resources, including more targeted spraying of insecticides and educational campaigns on local communities aimed at eliminating prevalent breeding sites (Boser et al., 2021; Limkittikul et al., 2014; Runge-Ranzinger et al., 2014; Runge-Ranzinger et al., 2016). The systematic reviews by Louis et al. (Louis et al., 2014) and Sallam et al. (Sallam et al., 2017) summarized how satellite imagery and remote sensing techniques have successfully been applied in the past to estimate the spatial variance in *Aedes aegypti* abundance on both a global and local scale. They provided an extensive overview of hypothesis-driven indicators and modeling approaches that are instrumental in generating spatial suitability models for *Aedes aegypti*. However, both identified a gap in generating and evaluating urban indicators to capture *Aedes aegypti* distributions at mosquito flight range scale.

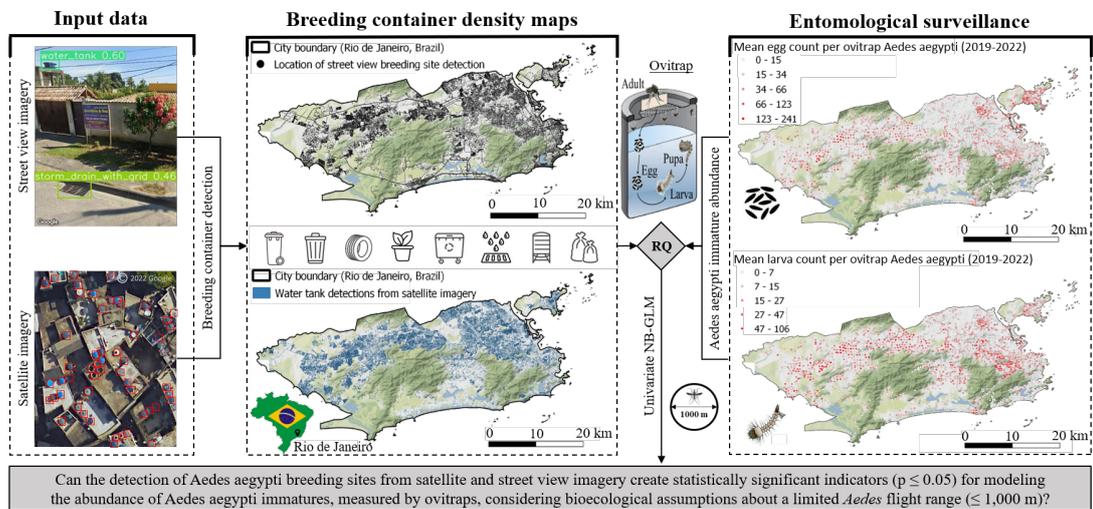
In this paper, we therefore propose a workflow of first generating high-resolution proxies to model *Aedes aegypti* abundance, and second, evaluating them with entomological surveillance data collected via ovitraps over a time period of four years. More precisely, we applied state-of-the-art computer vision models on satellite and street view imagery to detect common *Aedes aegypti* breeding sites. We chose these image datasets for their open accessibility, high resolution, and georeferencing, enabling a city-wide environmental analysis and the generation of urban micro-habitat indicators to estimate *Aedes aegypti* suitability at mosquito flight range scale. The joint application of both datasets promises to generate synergies. While both datasets have individually proven useful for this area of application in our previous studies (Haddawy et al., 2019; Knoblauch et al., 2023; Su Yin et al., 2021), their combined usefulness in this application field has not yet been investigated, to our knowledge. Therefore, our specific aim is to evaluate the following research question (RQ) to support future vector control of *Aedes aegypti*:

- **RQ:** Can the detection of *Aedes aegypti* breeding sites from satellite and street view imagery create statistically significant indicators ( $p \leq 0.05$ ) for modeling the abundance of *Aedes aegypti* immatures, measured by ovitraps,

considering bioecological assumptions about a limited *Aedes* flight range ( $\leq 1,000$  m)?

## 2 Methods

Our experiment consists mainly of two parts (cf. Figure 1): first, the detection of common breeding grounds for *Aedes aegypti* mosquitoes in urban areas, and second, the evaluation of container density for inference on urban *Aedes aegypti* abundance, considering limited mosquito flight range. We applied this workflow to the municipality of Rio de Janeiro, an endemic place for *Aedes aegypti* in Brazil, which is one of the worldwide hotspots for dengue, chikungunya, and Zika outbreaks (Gibson et al., 2014; Wilson, 2011).



**Figure 1:** Workflow of evaluating the density of *Aedes aegypti* breeding container detections for modeling immature mosquito abundance at flight range scale in the city of Rio de Janeiro, Brazil. The mapping of *Aedes aegypti* breeding containers was carried out using satellite and street view imagery by applying and fine-tuning single-stage object detection networks (left). Container densities were calculated within a circular flight range buffer of 1,000 m around ovitrap locations. For the evaluation of the research question, univariate negative binomial regression models were trained using temporally aggregated egg and larva counts from entomological surveillance (middle). Entomological surveillance data about immature abundance of *Aedes aegypti* was collected by the municipal health ministry of Rio de Janeiro (right). ©2023 Google

### 2.1 Computer vision models for the detection of *Aedes aegypti* breeding habitat

The selection of breeding containers in this study was guided by a priori expectations derived from existing literature (Arana-Guardia et al., 2014; Cavalcanti et al., 2016; Medronho et al., 2009; Paploski et al., 2016; Simard et al., 2005; Souza et al., 2017; Spiegel et al., 2007; Valença et al., 2013; Vezzani and Schweigmann, 2002), centered around their presumed influence on the abundance of *Aedes aegypti* immatures. The generation of water tank counts as a micro-habitat proxy derived from satellite imagery was extensively described in our previous research study (Knoblauch et al. (2023)). In this previous work, we conceptualized a

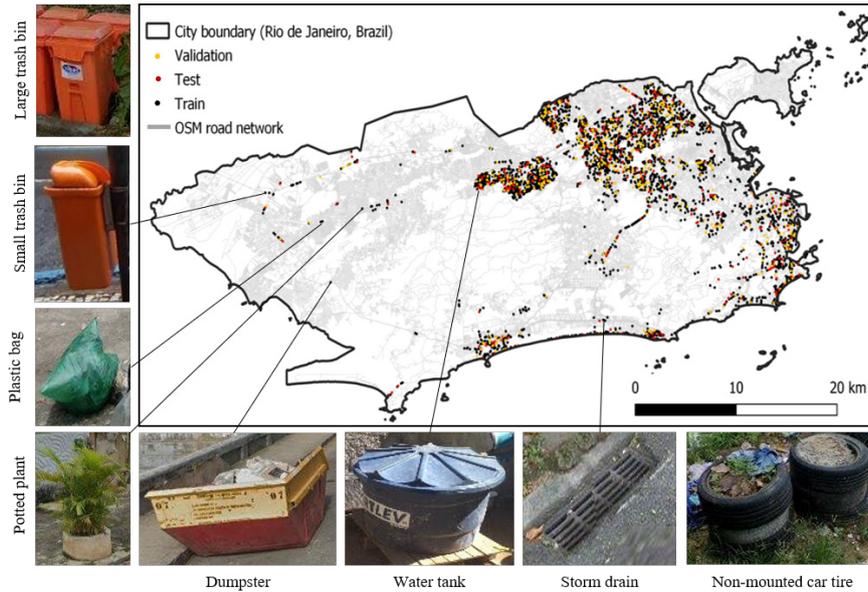
semi-supervised self-training algorithm to minimize the manual labeling effort for automated water tank detection in urban areas based on satellite imagery. We used a Single-Stage Object Detection network consisting of Inception-ResNet-V2 as a feature extractor and a multi-layer detector with a Non-Maximum Suppression layer pre-trained on the Microsoft COCO dataset (Lin et al., 2014). We fine-tuned this model using 4,000 manually labeled water tanks along with 10,400 pseudo water tank labels, encompassing various urban structure types, generated by the model during the training process. In our case, pseudo labels represented the results of model inference at 20,000 training iterations, applying a confidence threshold of 0.8. In total, the neural network was trained for 40,000 iterations: 20,000 initial iterations using manual labels only and 20,000 subsequent iterations using both manual and pseudo labels, which refers to a semi-supervised self-training procedure.

In the present study, we additionally fine-tuned a multi-class object detector to map further *Aedes aegypti*-specific habitats as an extension of prior research. These habitats include potted plants, large and small trash bins, plastic bags, non-mounted car tires, water tanks, dumpsters, and storm drains (cf. Figure 2). To detect these objects, we used street view images retrieved from Google’s Street View Static API (Google LLC, 2023). A 50 m downloading interval for 360-degree street view images calculated from the OSM road network was deemed appropriate for the detection of mosquito breeding sites, following the approach used in other studies (Haddawy et al., 2019; Su Yin et al., 2021). As of August 8th, 2023, this method yielded a total of 467,605 available street view images, which were utilized for labeling and city-wide container detection. The timestamps of the retrieved images ranged from January 2010 until 2023, with a share of 51% for images taken between 2022 and 2023, 15% from 2021, 19% from 2020 and 15% from before 2020. The downloaded image resolution was 600x500 pixels. For the supervised training of our multi-class object detector we manually labeled 7,578 breeding containers on 3,979 images using the graphical image annotation tool ‘labelImg’ (TuzuTa Lin, 2023). To minimize the manual labelling effort we implemented additional data augmentation techniques for instances of the ‘dumpster’ container class, which were observed infrequently within our dataset. We applied PCA color augmentation, horizontal flip and 180 degree rotation. The labelled dataset was then randomly divided into 80% for training, 10% for validation, and 10% for testing, resulting in 3,152, 454, and 373 image subsets, respectively (cf. Table 1).

**Table 1:** Counts of images and labels for instances of *Aedes aegypti* breeding container detected in street view imagery, with differentiation between the train, validation, and test sets.

Dataset	Images	Labels	Instance							
			Dumpster	Large trash bin	Small trash bin	Non-mounted car tire	Plastic bag	Potted plant	Storm drain	Water tank
Train	3,152	5,729	310	428	400	625	757	1,990	606	613
Validation	454	965	69	72	61	87	189	291	97	99
Test	373	884	67	60	57	129	173	197	81	120

Based on the street view imagery we fine-tuned a YOLOv5 model, which was pre-trained on the Microsoft COCO dataset (Lin et al., 2014), specifically YOLOv5x, known for speed, accuracy, efficiency, adaptive architecture and scale-invariant detection (cf. Figure 3). The applied model consisted of a CSPNet

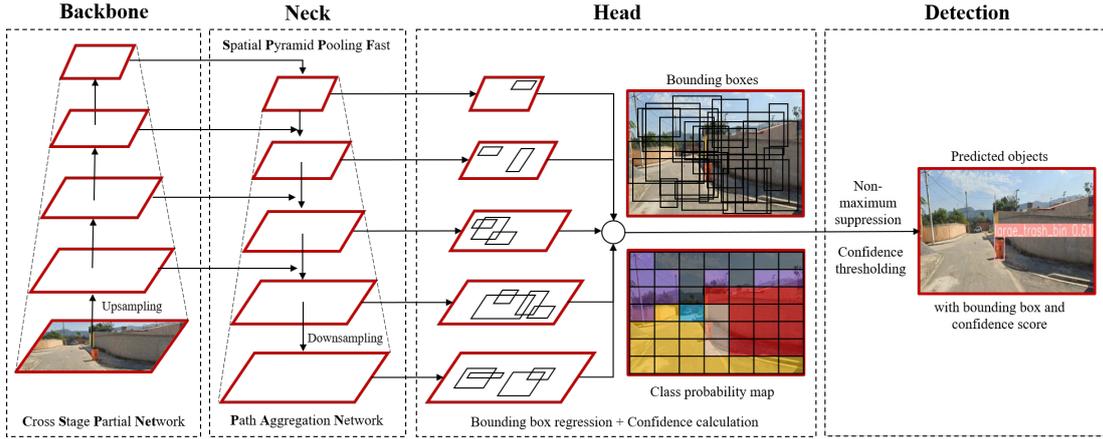


**Figure 2:** The images depicted identify breeding containers, accompanied by a map illustrating the coordinates of randomly selected train, test, and validation sets. These sets were chosen as subsets from a complete dataset of coordinates at 50 m intervals, encompassing the entire Open Street Map (OSM) road network in the municipality of Rio de Janeiro as of August 8th, 2023. Each train, test, and validation point corresponds to the downloading of five street view images, capturing a comprehensive 360-degree view at each location. This dataset compilation facilitated the training of object detection networks specifically tailored to identify *Aedes aegypti* breeding containers within the urban landscape. ©2023 Google

enhancing inter-layer information flow (Wang et al., 2019), SPPF for multi-scale object analysis (He et al., 2014), and PAN for parameter aggregation from different backbone levels (Liu et al., 2018). During training, we fitted key parameters, utilizing AdamW for stability (Loshchilov and Hutter, 2017). In three iterations of 300 epochs, we optimized the learning rate, adjusting it from  $5e^{-5}$  to  $1e^{-5}$ , to enhance the efficiency of the model. Model convergence was reached after 900 epochs applying a patience parameter of 20. An iterative decrease of the focal loss parameter from 0.5 to 0.2 was implemented to cope with feature imbalance. Feature imbalance can lead to higher miss-classification rates for minority class instances (Krawczyk, 2016). The selected hyperparameters for training were listed in Table A. 1.

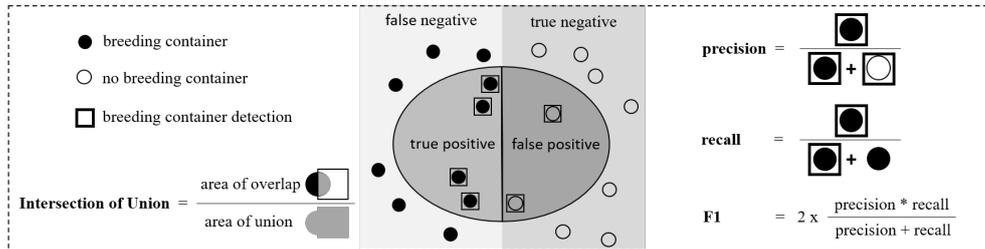
## 2.2 Evaluation metrics for container detection and workflow of city-wide prediction

Precision, recall, their harmonic mean, the F1-score, and the mean average precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5) were utilized to assess the object detection model performance (cf. Figure 4). Precision is defined as the ratio of the True Positive objects to all detected objects, while recall describes the fraction of relevant objects that are successfully retrieved. The performance metrics were computed based on the comparison between the intersection of the bounding boxes of the predictions and of the validation labels. This evaluation depended on the IoU value, which ranges between 0 and 1. An IoU Value of 0.5 or higher for a detected object was considered a True Positive,



**Figure 3:** Schematic YOLOv5x architecture applying upsampling for semantic enrichment and downsampling to augment image resolution. The backbone component shaped feature maps at various levels of granularity. Subsequently, the neck module merged these feature maps and forwards them to the prediction head. In this stage, the features were utilized to perform precise box and class predictions. ©2023 Google

while an IoU value lower than 0.5 indicated a False Positive. Evaluated models were deployed to map the locations of *Aedes aegypti* breeding containers across the whole metropolitan area of Rio de Janeiro. In carrying out this task, street view images were processed in batches to sequentially predict bounding boxes and probabilities. Subsequent post-processing steps included non-maximum suppression and thresholding, with the application of a confidence score equal to or above 0.3. For the detection of water tanks in satellite imagery, we utilized over 10 million patches at zoom level 22 from the Bing Tile Map Service (Microsoft, 2023). In this case, predictions were processed in parallel tasks. To organize the data, we employed the mapproxy API (Omniscale GmbH & Co. K.G. O.T., 2023), facilitating the storage of satellite imagery within a structured subset folder. Object detections for each image patch were pushed to a Post-GIS database. The database was then used for a post-processing step to filter predictions with confidence scores of 0.7.



**Figure 4:** Schematic explanation of evaluation metrics applied to implemented *Aedes aegypti* breeding container detection networks.

### 2.3 Inference on *Aedes aegypti* immature abundance

To quantitatively evaluate the research question concerning how well the density of each detected mosquito breeding container can represent the spatial distribution of *Aedes aegypti* immatures within urban areas, as measured by entomological surveillance data, we ran univariate negative binomial generalized linear

regression models (GLMs) employing log-link functions (Hilbe, 2012). The selection of the negative binomial GLM was motivated by its capacity to account for the observed overdispersion in the entomological count data. For each of the nine detected breeding container types, two univariate models were conducted: one employing the 'mean egg per trap' (MET) rate as the response variable (Mean = 19.8, Standard deviation = 20.12), and the other utilizing the 'mean larva per trap' (MLT) rate as the response variable (Mean = 10.96, Standard deviation = 10.94). The associated mathematical formulations are delineated in Equation 1. The entomological response variables  $Y_i$  were averaged over 48 months of ovitrap surveillance ranging from January 2019 to December 2022. This was performed to yield robust spatial measurements over time, mitigating potential biases that could arise from the manual ovitrap collection process.

$$\begin{aligned}
 Y_i &\sim NB(\hat{\mu}_i, \hat{\theta}) \\
 \mathbb{E}(Y_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta} \\
 \text{Var}(Y_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta}^2 \\
 \log(\hat{\mu}_i) &= \hat{\beta}_0 + \hat{\beta}_1 * \textit{Breeding Container Count}_i
 \end{aligned} \tag{1}$$

This approach led to the creation of two distinct models for each of the nine detected *Aedes aegypti* breeding containers. In these models, container density served as the independent variable, calculated using mosquito flight range buffers of 1,000 m around ovitrap locations. For breeding containers identified through street view imagery, the counts were further normalized based on the number of retrieved images within each circular flight range buffer. This normalization accounted for observed variations in street coverage and image availability across different spatial locations. Our analysis incorporated data from a total of 2,700 ovitrap locations, denoted by  $i$  in Equation 1, each providing information on monthly egg and larva counts. The entomological data was provided upon request by the health ministry of the city of Rio de Janeiro. To scrutinize the robustness of our research findings concerning the estimated maximum flight ranges of *Aedes aegypti*, a sensitivity analysis was conducted by employing alternative buffer sizes of 250 and 500 m.

### 3 Results and Discussion

#### 3.1 Evaluation of breeding container detection

The multi-class object detection network trained on street view imagery achieved an overall F1 score of 0.878, indicating a balanced precision-recall trade-off (cf. Table 3). When examining specific classes, the breeding container class 'dumpster' showed the highest F1 score of 0.950, supported by a precision of 0.946 and a recall of 0.955. Similarly, the 'large trash bin' container type also has a high precision (0.933) and recall (0.930), contributing to an F1 score of 0.931. Although the 'plastic bag' container type demonstrates relatively lower precision (0.706), its recall (0.792) and F1 score (0.747) remain reasonable. On the other hand, the 'potted plant' container type achieves a high recall (0.959) alongside a corresponding F1 score of 0.865. The results for water tank detection in satellite imagery were extensively described in our previous research work

(Knoblauch et al. (2023)). In this previous research, the object detection model yielded a precision score of 0.864, a recall of 0.823, and an F1 score of 0.843 on independent test datasets.

**Table 2:** Goodness of fit indicators for the YOLOv5 model trained on street view imagery applying a confidence threshold of 0.3. The performance was based on independent test data points.

Breeding container	YOLOv5			
	Precision (%)	Recall (%)	F1	mAP@0.5
Water tank	0.801	0.867	0.833	0.895
Non-mounted car tire	0.876	0.837	0.856	0.901
Storm drain	0.884	0.941	0.912	0.955
Plastic bag	0.706	0.792	0.747	0.79
Potted plant	0.788	0.959	0.865	0.897
Large trash bin	0.933	0.93	0.931	0.959
Small trash bin	0.892	0.947	0.919	0.962
Large trash bin	0.933	0.93	0.931	0.959
Dumpster	0.946	0.955	0.950	0.98
Average weighted by instance	0.853	0.904	0.878	0.917

Interestingly, the ‘dumpster’ container class achieved the highest F1 score despite a smaller training set, necessitating augmentation techniques. Being larger containers with distinct characteristics might aid accurate identification in dumpsters. In contrast, the ‘plastic bag’ class recorded the lowest F1 score among all classes. This could be attributed to the inherent variability in plastic bag attributes like shape, size, and color. From 93,521 citywide coordinates, our model detected 2,490 dumpsters, 7,927 large trash bins, 6,092 small trash bins, 24,034 non-mounted car tires, 43,334 plastic bags, 54,117 potted plants, 39,807 storm drains and 5,898 water tanks from street view imagery.

Upon examining the results, we also identified several limitations. Detection of potted plants behind open fences was challenging as the fence’s pattern texture blends with the potted plant objects, leading to erroneous detection results (cf. Figure 5). Vehicles parked along streets further complicated detection, potentially obstructing views of breeding containers. In detecting the water tank from the street view images, we observed that high-rise surroundings amplified the difficulty of identifying water tanks due to potential occlusions by neighboring structures. Other False Negative examples included sun-bleached water tanks, closely spaced storm drains, and overlaying containers such as plastic bags, water tanks, potted plants or non-mounted car tires. True Negative detections contained for example miscellaneous small containers in garbage heaps, plants without pots, sealed storm drains, and lattice trash bins. The most occurring False Positive cases included drainpipes detected as non-mounted car tires, truck loads detected as dumpsters and larger stones used as road boundaries falsely detected as either plastic bags or non-mounted car tires. An object class labeled as ‘miscellaneous small containers’, representing for example trash piles, was excluded during the training phase due to the absence of clear 3D object features and its varied appearance. These characteristics made it challenging to capture this potential breeding site using our object detection network for street view im-

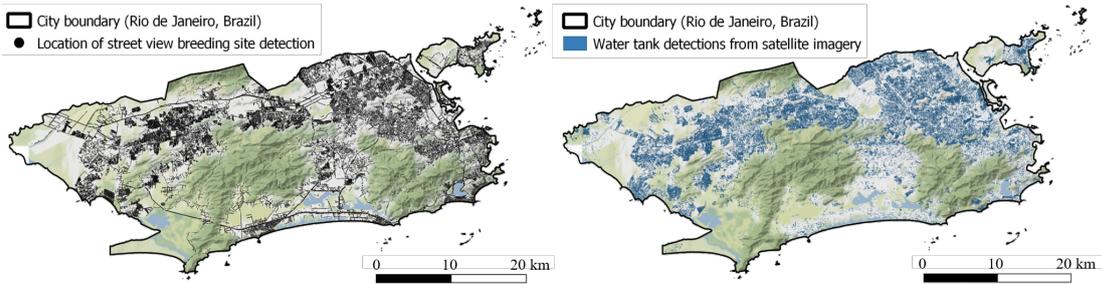
agery. An additional training of scene recognition model could potentially address these limitations associated with this object class.

In our previous research work on water tank detection in satellite imagery (Knoblauch et al. (2023)), common False Negative predictions included water tanks in shaded or partially shaded areas. To mitigate the occurrence of False Negative predictions, one could enhance the precision of the water tank detection network by augmenting its training data with additional instances of shaded water tank labels. It is important to highlight that our study area contained a substantial number of objects resembling water tanks, leading to a notable prevalence of False Positives. While our models rarely misclassified similar objects like blue cars and rooftop ventilators as water tanks, circular water pools and blue sunshades on beaches consistently resulted in False Positives. Addressing the False Positive detections of water pools could involve the implementation of a size filter, while False Positives associated with blue sunshades on beaches could be alleviated through the application of an automatic land use map-based filter. However, it is crucial to recognize that these solutions have inherent limitations. Specifically, they may not be effective for very small water pools and blue sunshades not situated on beaches, rendering the proposed methods obsolete. An alternative approach would involve filtering predictions based on confidence scores.



**Figure 5:** Example for False Negative, True Negative, False Positive, and True Positive breeding container predictions utilizing street view imagery. Detected breeding containers were indicated by bounding boxes, with distinct colors assigned to each container class. These visual representations were generated based on the confidence scores derived from a fine-tuned YOLOv5 model. In the two first image rows, white and black dashed bounding boxes were manually added to point to the locations of False Negative (white) and True Negative (black) examples, respectively, to enhance explanation. ©2023 Google

As a highlight of the present research work, we created density maps of *Aedes aegypti* breeding containers detected over the entire municipal area of Rio de Janeiro using the combined dataset of satellite and street view imagery (cf. Figure 6). All detected breeding containers exhibited a widespread distribution across the study region, characterized by substantial spatial variation. Satellite imagery played a crucial role in detecting breeding containers located in residential backyards and on top of buildings, while street view imagery complementary identified such containers on streets, beneath trees, and within sheltered areas (cf. Table 3). While the detection of mosquito breeding sites in satellite imagery was limited to water tanks due to the image resolution of 0.0373 m per pixel (Knoblauch et al., 2023), street view images enabled the detection of even small breeding containers like plastic bags. However, satellite imagery has the advantage of providing continuous spatial coverage. The availability of street view images within the road network of the municipality of Rio de Janeiro was restricted, especially in narrow and impassable streets that are common in favelas. Generally, the prevalence of artificial mosquito breeding containers was strongly linked to inhabited regions, with forested areas notably lacking such containers.



**Figure 6:** Large-scale *Aedes aegypti* breeding site detection from 461,152 street view and satellite imagery for the metropolitan area of Rio de Janeiro, Brazil. Left map shows location of retrieved street view images used for 360-degree breeding site detection and right maps highlights water tank density detected from satellite imagery generated in Knoblauch et al. (2023).

**Table 3:** Qualitative attribute comparison of satellite and street view imagery, underscoring the essential need for a complementary application of these digital sources when mapping *Aedes aegypti* breeding sites.

Attribute	Street view imagery	Satellite imagery
Spatial coverage	Limited coverage defined by visited road network. Ground-level perspective below shelters and tree canopy.	Complete coverage. Monitoring of inaccessible backyards and rooftops.
Image resolution	Detection of small and large breeding container.	Fails at capturing smaller breeding sites.
Open-accessibility	Open-accessibility with monthly downloading limit. Open-source alternatives without limits.	Open-accessibility at limited resolution. Commercial high-resolution alternatives.
Temporal updates	Infrequent, non-collective updates of images.	Infrequent, collective updates of all images.
Provider	<b>Google</b> (Google LLC, 2023)/ <b>Mapilio</b> (Mapilio, 2023)/ <b>Mapillary</b> (Mapillary, 2023)/ <b>KartaView</b> (KartaView, 2023)/...	<b>Microsoft</b> (Microsoft, 2023)/ <b>NASA</b> (NASA, 2023)/ <b>Copernicus</b> (Copernicus, 2023)/ <b>Planet</b> (Planet, 2023)/...

### 3.2 Modeling of immature *Aedes aegypti* infestation

The results of our negative binomial linear regression models (cf. Table 4) indicated that all detected breeding containers of this study were highly significant

( $p \leq 0.05$ ) proxies for modeling urban *Aedes aegypti* immature abundance while considering limited mosquito flight range below 1,000 m. This was in line with our expectations and implies that breeding site density maps can be a useful indicator to enrich entomological surveillance data and thus support future vector control by providing more continuous and high resolution insights for urban mosquito distributions.

**Table 4:** Coefficients, standard errors, and p-values for univariate negative binomial generalized linear regression models applying a *Aedes* flight range buffer of 1,000 m around 2,700 ovitrap locations. Regression coefficients and standard errors were reported at the link scale. The p-value of the intercept was  $< 2e^{-16}$  for all models. (Water tank\* = Water tanks detected with satellite imagery)

Breeding container	Response	Intercept		Breeding Container			Theta		Metrics	
		Estimate	Std. Error	Estimate	Std. Error	p-value	Estimate	Std. Error	Pseudo R2	AIC
Water tank*	MET	2.8493	0.0296	0.0026	0.0004	$2.03e^{-09}$	1.1990	0.0363	0.0123	17645
	MLT	2.1046	0.0298	0.0053	0.0004	$< 2e^{-16}$	1.2937	0.0435	0.0496	14551
Water tank	MET	2.8448	0.0272	0.2580	0.0358	$5.49e^{-13}$	1.2045	0.0365	0.0172	17634
	MLT	2.2253	0.0280	0.3077	0.0369	$< 2e^{-16}$	1.2505	0.0417	0.0225	14620
Non-mounted car tire	MET	2.9169	0.0230	0.2872	0.0505	$1.26e^{-08}$	1.1964	0.0362	0.0110	17650
	MLT	2.2941	0.0234	0.4027	0.0506	$1.78e^{-15}$	1.2506	0.0417	0.0223	14620
Storm drain	MET	2.8028	0.0431	0.4343	0.0926	$2.7e^{-06}$	1.1927	0.0361	0.0082	17657
	MLT	2.1101	0.0441	0.6704	0.0945	$1.29e^{-12}$	1.2431	0.0414	0.0176	14632
Plastic bag	MET	2.8694	0.0316	0.2583	0.0558	$3.7e^{-06}$	1.1909	0.0360	0.0068	17660
	MLT	2.1996	0.0324	0.4223	0.0566	$8.44e^{-14}$	1.2436	0.0414	0.0179	14631
Potted plant	MET	2.8723	0.0328	0.1870	0.0440	$2.11e^{-05}$	1.1908	0.0360	0.0067	17661
	MLT	2.2224	0.0336	0.2837	0.0454	$4.03e^{-10}$	1.2369	0.0411	0.0134	14642
Small trash bin	MET	3.1111	0.0245	-1.6963	0.1775	$< 2e^{-16}$	1.2178	0.0370	0.0276	17607
	MLT	2.4966	0.0252	-1.3805	0.1865	$1.32e^{-13}$	1.2428	0.0414	0.0177	14632
Large trash bin	MET	3.0765	0.0270	-1.0363	0.1953	$1.12e^{-07}$	1.1958	0.0362	0.0105	17651
	MLT	2.4863	0.0278	-1.0553	0.2063	$3.13e^{-07}$	1.2325	0.0410	0.0102	14650
Dumpster	MET	3.0561	0.0243	-2.4497	0.4461	$4e^{-08}$	1.1973	0.0363	0.0116	17648
	MLT	2.4648	0.0249	-2.4589	0.4636	$1.13e^{-07}$	1.2340	0.0410	0.0112	14648

Water tanks, non-mounted car tires, storm drains, plastic bags, and potted plants consistently displayed positive coefficient estimates for both response variables, whereas the coefficient estimates for small and large trash bins, as well as dumpsters, consistently demonstrated negativity across both model variations. These findings aligned with the intuitive understanding that an increased presence of trash bins of any kind correlates with a reduced prevalence of uncontained refuse piles, thereby mitigating the potential for additional mosquito breeding sites. The correlation between the density of plastic bags and all three trash container classes was found to be negative, namely -0.1 for the dumpster class, -0.03 for large trash bins, and -0.3 for small trash bins. In addition, small and large trash bins, as well as dumpsters, are usually closed containers that rarely fill with water when it rains, which underlines their significance ( $p \leq 0.05$ ) and negative association with entomological data about *Aedes aegypti* immature abundance. Furthermore, these containers are regularly emptied by refuse collection services, ensuring that they often remain dry and unsuitable for mosquito breeding, thus contributing to mosquito control efforts.

When analyzing the results independently from the response variable, it was observed that models using water tank density derived from satellite and street view imagery consistently led to the lowest Akaike information criterion (AIC), indicating a superior fit to the data across both immature abundance stages. Conversely, models employing the density of potted plants displayed the highest AIC values in relation to the MET rate, while models utilizing the density of large trash bins exhibited the highest AIC values in relation to the MLT rate. The ex-

tent of explained deviance in the regression models pertaining to the MLT rate generally exhibited higher values compared to those associated with the MET rate. Specifically, the MLT model, utilizing water tank density derived from satellite imagery, achieved the highest explained deviance at 0.05 as quantified by Cohen's pseudo- $R^2$  (Cohen, 2013) (cf. Equation 2). This indicates that approximately 5% of the variance in the response variable is accounted for by the univariate model.

$$\begin{aligned} \text{Cohen's pseudo } R^2 &= 1 - \frac{\text{model deviance}}{\text{null model deviance}} \\ \text{Negative binomial model deviance} &= 2 \sum (y \cdot \log(\frac{y}{\mu}) - (y + k - 1) \log(\frac{y+k-1}{\mu+k-1})) \end{aligned} \quad (2)$$

The deviance function of the negative binomial GLM captured the increasing variance with the mean that is typical for count data. The dispersion parameter captures how much the variance increases with the mean relative to a Poisson GLM, where the variance equals the mean. The theta values of all univariate regression models in this study indicated a substantial overdispersion. This overdispersion can be attributed to two primary factors. First, the dataset on entomological observations contained a substantial number of zero values, necessitating the adoption of a negative binomial GLM to account for excess variation. Second, the limited inclusion of predictors in modeling the urban distribution of *Aedes aegypti* also contributed to the observed low value of explained deviance. It is worth noting that certain potentially relevant predictors have been intentionally omitted from the model, further contributing to the constrained explanatory power. The incorporation of additional explanatory variables is planned for subsequent phases of this research.

The outcomes of the performed sensitivity analysis (cf. Table A. 2), scrutinizing different assumed maximum flight ranges of *Aedes aegypti* (250 m, 500 m, 1,000 m), confirmed the robustness of the results outlined in Table 4. Similar to the results for a 1,000 m *Aedes aegypti* maximum flight range, at a maximum flight range of 500 m, all container types exhibited significant p-values ( $p \leq 0.05$ ) for both egg and larva counts. The same trend was observed for the assumed maximum *Aedes aegypti* flight range of 250 m, except for the container types dumpster, storm drain, and water tank detected from satellite imagery. Notably, the findings concerning water tanks from satellite imagery at 250 m scale show a slight contrast to our previous findings in Knoblauch et al. (2023), where a different time frame for entomological data was utilized; however, significance was detected at a flight range scale of 200 m. This divergence of these findings underscores the considerable influence of the selected time period of entomological surveillance on the validation of such results. The coefficients for small and large trash bins, as well as the dumpster category, remained negative also at lower estimated maximum *Aedes aegypti* flight ranges. Intriguingly, the coefficient for potted plants shifted from positive to negative when simulating a maximum flight range of 250 m for *Aedes aegypti*. Overall, there was an evident upward trend in significance (indicated by a downward trend in p-values) across all container classes, with larger buffer sizes, representing simulations of larger flight ranges, showing higher significance levels. Essentially, larger buffer areas augment the probability of encountering containers, consequently yielding more dependable

statistical outcomes in our methodology for modeling ovitrap count data with digital proxies. For a more nuanced understanding of the relationship between assumed maximum *Aedes aegypti* flight range and significance values, models implementing soft constraints could be considered, such as Bayesian models.

The collective findings presented in this study offer a comprehensive overview and extension of our prior research about urban mosquito mapping (Haddawy et al., 2019; Knoblauch et al., 2023; Su Yin et al., 2021). For the first time the results underscore the practical efficacy of integrating satellite and street view imagery for identifying mosquito breeding sites in urban areas, emphasizing the distinctive advantages of each method. A further alternative data source for mapping mosquito breeding containers in urban areas could be drone imagery, which offers both continuous spatial coverage and images in high resolution for small breeding container detection (Passos et al., 2022; Passos et al., 2023). However, it is essential to note that generating drone imagery incurs substantial costs and labor, thereby limiting the applicability in diverse global urban settings. A common limitation across all three data sources is their inability to detect breeding containers located inside buildings. Consequently, the digital strategies outlined in this study cannot fully replace on-site entomological surveillance. Instead, our approach aims to complement manual monitoring efforts by augmenting them with high-resolution digital information. Citizen Science offers a promising avenue to address this limitation, fostering public participation, including crowdsourced mapping, to enhance data collection and monitoring, particularly of indoor breeding sites. The primary challenge in utilizing digital data sources for mosquito mapping lies in achieving temporal alignment with entomological surveillance for modeling purposes.

Another challenge associated with digital data sources, such as satellite and street view imagery, pertains to the potential obsolescence of information and the insights derived from it. Street view images, in particular, are infrequently updated (Hou and Biljecki, 2022). It is also crucial to consider the transient nature and shifting locations of identified containers, especially for plastic bags, potted plants, non-mounted car tires, large trash bins, and dumpsters, which may have introduced a potential bias to the measured significance values of these container classes in our results. Conversely, water tanks, small city trash bins attached to streetlights, and storm drains are presumed to have relatively stable locations over time, leading to more reliable results. Furthermore, the calculated container densities in this study may be influenced by citywide solid waste collections or vector control campaigns, wherein breeding containers may have been removed before images were captured. In future studies, investigating the relationship between image timestamps and such interventions, as well as exploring alternative data sources (cf. Table 3), could be beneficial. Crowd-sourced platforms such as Mapillary (Mapillary, 2023) and KartaView (KartaView, 2023) may particularly offer more continuous image updates (Biljecki et al., 2023).

In summary, this study demonstrated the enhanced efficiency in managing urban diseases such as dengue through the application of digital techniques. The increasing availability of spatial big data, such as satellite and street view imagery, presents a considerable opportunity for obtaining high-resolution indicators for mapping urban mosquito suitability beyond entomological sample points and allows interpolations without violating biological assumptions about limited

mosquito flight ranges in the future. The proposed approach can be combined with further urban-specific mosquito proxies for enabling more targeted vector control. A task that is challenging with entomological surveillance alone. The proposed method can thus not only reduce surveillance costs but also facilitates the potential interruption of infection chains at earlier stages of an outbreak than with conventional methods.

## Data statement

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. Restrictions apply only to the sharing of entomological surveillance data collected by the Municipal Health Ministry of Rio de Janeiro.

## Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) [grant number 451956976].

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Credit authorship contribution statement

S.K. conceptualized the study. S.K. and M.S.Y. conducted the experiment. S.K. analyzed and interpreted the data. S.K. drafted the manuscript. S.K., M.S.Y., and K.C. worked on software for object detection. A.R. supported acquisition of entomological data. S.K, M.S.Y, P.H, F.B., B.R reviewed and edited the manuscript. S.L., B.R., T.J., I.M, D.A., and S.K. supported acquisition of funding. A.Z. conducted and supervised the study. All authors read and approved the manuscript and its subsequent revisions prior to submission.

# Appendix

## Appendix A

**Table A. 1:** Hyperparameters used for training. The training and detection processes were conducted using Google Colab, a cloud computing platform offering 53 GB of Random Access Memory (RAM), 8 CPU cores, a Tesla V100 GPU, and 150 GB of disk space.

<b>Batch Size</b>	24
<b>Learning Rate</b>	0.0004
<b>IoU Threshold</b>	0.5
<b>Optimizer</b>	RMSprop
<b>Optimizer Momentum</b>	0.9
<b>Optimizer Decay</b>	0.9
<b>Optimizer Epsilon</b>	0.1

**Table A. 2:** Coefficients, standard errors, and p-values for univariate negative binomial generalized linear regression models applying a mosquito flight range buffer of 250 and 500 m around 2,700 ovitrap locations. Regression coefficients and standard errors were reported at the link scale. The p-value of the intercept was  $< 2e^{-16}$  for all models. (Water tank\* = Water tanks detected with satellite imagery)

Breeding container	Response	Modelled flight range	Intercept			Breeding Container			Theta		Metrics	
			Estimate	Std. Error	p-value	Estimate	Std. Error	p-value	Estimate	Std. Error	Pseudo R2	AIC
Water tank*	MET	≤250m	2.9734	0.0256	0.0013	0.0018	0.4450	1.1824	0.0357	0.0002	17677.0000	
		≤500m	2.9655	0.0229	0.0085	0.0047	0.0410	1.1831	0.0358	0.0008	17676.0000	
	MLT	≤250m	2.3954	0.0261	-0.0001	0.0018	0.9510	1.2172	0.0403	$1.304e^{-6}$	14675.0000	
		≤500m	2.3312	0.0234	0.0258	0.0047	$3.43e^{-8}$	1.2262	0.0407	0.0063	14660.0000	
Water tank	MET	≤250m	2.9450	0.0235	0.0839	0.0262	0.0014	1.1874	0.0359	0.0041	17667.0000	
		≤500m	2.8928	0.0252	0.1785	0.0312	$1.05e^{-08}$	1.1971	0.0363	0.0116	17648.0000	
	MLT	≤250m	2.3576	0.0241	0.0759	0.0267	0.0045	1.2219	0.0405	0.0032	14667.0000	
		≤500m	2.2799	0.0259	0.2182	0.0321	$1e^{-11}$	1.2411	0.0413	0.0161	14635.0000	
Non-mounted car tire	MET	≤250m	2.9732	0.0205	0.0606	0.0223	0.0066	1.1871	0.0359	0.0037	17668.0000	
		≤500m	2.9599	0.0211	0.1163	0.0311	0.0002	1.1883	0.0360	0.0048	17666.0000	
	MLT	≤250m	2.3840	0.0210	0.0510	0.0224	0.0229	1.2215	0.0405	0.0028	14668.0000	
		≤500m	2.3609	0.0215	0.1494	0.0314	$1.96e^{-6}$	1.2288	0.0408	0.0079	14656.0000	
Storm drain	MET	≤250m	2.9883	0.0252	-0.0078	0.0419	0.8530	1.1822	0.0357	$1.47e^{-5}$	17678.0000	
		≤500m	2.9069	0.0309	0.1916	0.0583	0.0010	1.1869	0.0359	0.0037	17668.0000	
	MLT	≤250m	2.4173	0.0259	-0.0655	0.0438	0.1350	1.2186	0.0404	0.0010	14673.0000	
		≤500m	2.2785	0.0316	0.2797	0.0593	$2.44e^{-6}$	1.2277	0.0408	0.0072	14657.0000	
Plastic bag	MET	≤250m	2.9626	0.0228	0.0565	0.0272	0.0375	1.1844	0.0358	0.0017	17673.0000	
		≤500m	2.9524	0.0249	0.0754	0.0342	0.0273	1.1842	0.0358	0.0016	17674.0000	
	MLT	≤250m	2.3757	0.0234	0.0465	0.0282	0.0982	1.2189	0.0404	0.0011	14672.0000	
		≤500m	2.3381	0.0255	0.1254	0.0346	0.0003	1.2234	0.0406	0.0042	14665.0000	
Potted plant	MET	≤250m	3.0241	0.0238	-0.0796	0.0253	0.0016	1.1869	0.0359	0.0037	17668.0000	
		≤500m	2.9210	0.0270	0.1058	0.0305	0.0005	1.1878	0.0359	0.0043	17667.0000	
	MLT	≤250m	2.4277	0.0245	-0.0687	0.0262	0.0088	1.2211	0.0405	0.0027	14669.0000	
		≤500m	2.3076	0.0277	0.1420	0.0314	$5.97e^{-6}$	1.2277	0.0408	0.0072	14658.0000	
Small trash bin	MET	≤250m	3.0202	0.0210	-0.6147	0.0963	$1.77e^{-10}$	1.2039	0.0365	0.0164	17636.0000	
		≤500m	3.0274	0.0227	-0.5604	0.1320	$2.18e^{-5}$	1.1907	0.0360	0.0067	17661.0000	
	MLT	≤250m	2.4299	0.0216	-0.6224	0.1005	$5.95e^{-10}$	1.2415	0.0414	0.0159	14636.0000	
		≤500m	2.4275	0.0233	-0.4449	0.1383	0.0013	1.2227	0.0406	0.0038	14666.0000	
Large trash bin	MET	≤250m	3.0038	0.0211	-0.2511	0.0788	0.0014	1.1872	0.0359	0.0039	17668.0000	
		≤500m	3.0058	0.0227	-0.2305	0.1155	0.0459	1.1842	0.0358	0.0015	17674.0000	
	MLT	≤250m	2.4106	0.0216	-0.2168	0.0806	0.0072	1.2212	0.0405	0.0027	14669.0000	
		≤500m	2.4224	0.0233	-0.3204	0.1192	0.0072	1.2212	0.0405	0.0027	14669.0000	
Dumpster	MET	≤250m	2.9850	0.0205	0.0161	0.1230	0.8960	1.1822	0.0357	$9.98e^{-6}$	17678.0000	
		≤500m	3.0061	0.0214	-0.6987	0.2279	0.0022	1.1871	0.0359	0.0038	17668.0000	
	MLT	≤250m	2.3990	0.0210	-0.1589	0.1290	0.2180	1.2185	0.0404	0.0008	14673.0000	
		≤500m	2.4179	0.0219	-0.8032	0.2349	0.0006	1.2241	0.0406	0.0047	14664.0000	



# Bibliography

- Arana-Guardia, R., Baak-Baak, C. M., Loroño-Pino, M. A., Machain-Williams, C., Beaty, B. J., Eisen, L., and García-Rejón, J. E. (2014). “Stormwater drains and catch basins as sources for production of *Aedes aegypti* and *Culex quinquefasciatus*”. *Acta tropica* 134, pp. 33–42. DOI: 10.1016/j.actatropica.2014.01.011.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I. (2013). “The global distribution and burden of dengue”. *Nature* 496.7446, pp. 504–507. DOI: 10.1038/nature12060.
- Biljecki, F., Zhao, T., Liang, X., and Hou, Y. (2023). “Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives”. *International Journal of Applied Earth Observation and Geoinformation* 122, p. 103385. DOI: 10.1016/j.jag.2023.103385.
- Boser, A., Sousa, D., Larsen, A., and MacDonald, A. (2021). “Micro-climate to macro-risk: mapping fine scale differences in mosquito-borne disease risk using remote sensing”. *Environmental Research Letters* 16.12, p. 124014. DOI: 10.1088/1748-9326/ac3589.
- Cavalcanti, L. P. d. G., Oliveira, R. d. M. A. B., and Alencar, C. H. (2016). “Changes in infestation sites of female *Aedes aegypti* in Northeast Brazil”. *Revista da Sociedade Brasileira de Medicina Tropical* 49.4, pp. 498–501. DOI: 10.1590/0037-8682-0044-2016.
- Cohen (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. DOI: 10.4324/9780203774441.
- Copernicus (2023). *Copernicus service catalogue*.
- David, M. R., Lourenço-de-Oliveira, R., and Freitas, R. M. de (2009). “Container productivity, daily survival rates and dispersal of *Aedes aegypti* mosquitoes in a high income dengue epidemic neighbourhood of Rio de Janeiro: presumed influence of differential urban structure on mosquito biology”. *Memorias do Instituto Oswaldo Cruz* 104.6, pp. 927–932. DOI: 10.1590/s0074-02762009000600019.
- Ebi, K. L. and Nealon, J. (2016). “Dengue in a changing climate”. *Environmental research* 151, pp. 115–123. DOI: 10.1016/j.envres.2016.07.026.
- Gibson, G., Souza-Santos, R., Pedro, A. S., Honório, N. A., and Sá Carvalho, M. (2014). “Occurrence of severe dengue in Rio de Janeiro: an ecological study”. *Revista da Sociedade Brasileira de Medicina Tropical* 47.6, pp. 684–691. DOI: 10.1590/0037-8682-0223-2014.
- Glassman, R., Scarpino, S., and Gilmour, J. (2022). *The Increasing Burden of Dengue Fever in a Changing Climate: Multiple effects of climate change affect the incidence and severity of dengue fever*. Ed. by The Rockefeller Foundation.
- Google LLC (2023). *Street View Static API*.
- Haddawy, P., Wettayakorn, P., Nonthaleerak, B., Su Yin, M., Wiratsudakul, A., Schönning, J., Laosiritaworn, Y., Balla, K., Euaungkanakul, S., Quengdaeng, P., Chokni-

- tipakin, K., Traivijitkhun, S., Erawan, B., and Kraissang, T. (2019). “Large scale detailed mapping of dengue vector breeding sites using street view images”. *PLoS neglected tropical diseases* 13.7, e0007555. DOI: 10.1371/journal.pntd.0007555.
- He, K., Zhang, X., Ren, S., and Sun, J. (2014). “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition”. DOI: 10.48550/arXiv.1406.4729.
- Hilbe, J. M. (2012). *Negative Binomial Regression*. Cambridge University Press. DOI: 10.1017/CB09780511973420.
- Honório, N. A., Da Silva, W. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003). “Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil”. *Memorias do Instituto Oswaldo Cruz* 98.2, pp. 191–198. DOI: 10.1590/s0074-02762003000200005.
- Hou, Y. and Biljecki, F. (2022). “A comprehensive framework for evaluating the quality of street view imagery”. *International Journal of Applied Earth Observation and Geoinformation* 115, p. 103094. DOI: 10.1016/j.jag.2022.103094.
- KartaView (2023). *OpenStreetCam API Documentation*.
- Knoblauch, S. and Moritz, M. (2023). *Geospatial Innovation’s Potential for Addressing Mosquito-Borne Diseases*. Ed. by United Nations World Data Forum.
- Knoblauch, S., Su Yin, M., Chatrinan, K., Rocha, A. A. d. A., Haddawy, P., Biljecki, F., Lautenbach, S., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2023). “High-resolution mapping of urban *Aedes aegypti* immature abundance through breeding site detection based on satellite and street view imagery (under review)”. *Scientific Reports: Collection on Urbanization and communicable diseases*.
- Krawczyk, B. (2016). “Learning from imbalanced data: open challenges and future directions”. *Progress in Artificial Intelligence* 5.4, pp. 221–232. DOI: 10.1007/s13748-016-0094-0.
- Limkittikul, K., Brett, J., and L’Azou, M. (2014). “Epidemiological trends of dengue disease in Thailand (2000-2011): a systematic literature review”. *PLoS neglected tropical diseases* 8.11, e3241. DOI: 10.1371/journal.pntd.0003241.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). “Microsoft COCO: Common Objects in Context”. DOI: 10.48550/arXiv.1405.0312.
- Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). “Path Aggregation Network for Instance Segmentation”. DOI: 10.48550/arXiv.1803.01534.
- Loshchilov, I. and Hutter, F. (2017). “Decoupled Weight Decay Regularization”. DOI: 10.48550/arXiv.1711.05101.
- Louis, V. R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y., and Dambach, P. (2014). “Modeling tools for dengue risk mapping - a systematic review”. *International journal of health geographics* 13, p. 50. DOI: 10.1186/1476-072X-13-50.
- Mapilio (2023). *Mapilio API Documentation*.
- Mapillary (2023). *Mapillary API Documentation*.
- Medronho, R. A., Macrini, L., Novellino, D. M., Lagrotta, M. T. F., Câmara, V. M., and Pedreira, C. E. (2009). “*Aedes aegypti* immature forms distribution according to type of breeding site”. *The American journal of tropical medicine and hygiene* 80.3, pp. 401–404. DOI: 10.4269/ajtmh.2009.80.401.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. (2019). “The current and future global distribution

- and population at risk of dengue”. *Nature microbiology* 4.9, pp. 1508–1515. DOI: 10.1038/s41564-019-0476-8.
- Microsoft (2023). *Zoom levels and tile grid of Bing Map Tiles service*.
- NASA (2023). *World View Earth Data*.
- Omniscale GmbH & Co. K.G. O.T. (2023). *MapProxy*.
- Paploski, I. A. D., Rodrigues, M. S., Mugabe, V. A., Kikuti, M., Tavares, A. S., Reis, M. G., Kitron, U., and Ribeiro, G. S. (2016). “Storm drains as larval development and adult resting sites for *Aedes aegypti* and *Aedes albopictus* in Salvador, Brazil”. *Parasites & vectors* 9.1, p. 419. DOI: 10.1186/s13071-016-1705-0.
- Passos, W. L., Araujo, G. M., Lima, A. A. de, Netto, S. L., and Da Silva, E. A. (2022). “Automatic detection of *Aedes aegypti* breeding grounds based on deep networks with spatio-temporal consistency”. *Computers, Environment and Urban Systems* 93, p. 101754. DOI: 10.1016/j.compenvurbsys.2021.101754.
- Passos, W. L., Da Barreto, C. S., Araujo, G. M., Haque, U., Netto, S. L., and Da Silva, E. A. (2023). “Toward improved surveillance of *Aedes aegypti* breeding grounds through artificially augmented data”. *Engineering Applications of Artificial Intelligence* 123, p. 106488. DOI: 10.1016/j.engappai.2023.106488.
- Planet (2023). *Planet Homepage*.
- Runge-Ranzinger, S., McCall, P. J., Kroeger, A., and Horstick, O. (2014). “Dengue disease surveillance: an updated systematic literature review”. *Tropical medicine & international health : TM & IH* 19.9, pp. 1116–1160. DOI: 10.1111/tmi.12333.
- Runge-Ranzinger, S., Kroeger, A., Olliaro, P., McCall, P. J., Sánchez Tejada, G., Lloyd, L. S., Hakim, L., Bowman, L. R., Horstick, O., and Coelho, G. (2016). “Dengue Contingency Planning: From Research to Policy and Practice”. *PLoS neglected tropical diseases* 10.9, e0004916. DOI: 10.1371/journal.pntd.0004916.
- Sallam, M. F., Fizer, C., Pilant, A. N., and Whung, P.-Y. (2017). “Systematic Review: Land Cover, Meteorological, and Socioeconomic Determinants of *Aedes* Mosquito Habitat for Risk Mapping”. *International journal of environmental research and public health* 14.10. DOI: 10.3390/ijerph14101230.
- Semenza, J. C., Rocklöv, J., and Ebi, K. L. (2022). “Climate Change and Cascading Risks from Infectious Disease”. *Infectious diseases and therapy* 11.4, pp. 1371–1390. DOI: 10.1007/s40121-022-00647-3.
- Simard, F., Nchoutpouen, E., Toto, J. C., and Fontenille, D. (2005). “Geographic distribution and breeding site preference of *Aedes albopictus* and *Aedes aegypti* (Diptera: culicidae) in Cameroon, Central Africa”. *Journal of medical entomology* 42.5, pp. 726–731. DOI: 10.1093/jmedent/42.5.726.
- Souza, R. L., Mugabe, V. A., Paploski, I. A. D., Rodrigues, M. S., Moreira, P. S. D. S., Nascimento, L. C. J., Roundy, C. M., Weaver, S. C., Reis, M. G., Kitron, U., and Ribeiro, G. S. (2017). “Effect of an intervention in storm drains to prevent *Aedes aegypti* reproduction in Salvador, Brazil”. *Parasites & vectors* 10.1, p. 328. DOI: 10.1186/s13071-017-2266-6.
- Spiegel, J. M., Bonet, M., Ibarra, A.-M., Pagliccia, N., Ouellette, V., and Yassi, A. (2007). “Social and environmental determinants of *Aedes aegypti* infestation in Central Havana: results of a case-control study nested in an integrated dengue surveillance programme in Cuba”. *Tropical medicine & international health : TM & IH* 12.4, pp. 503–510. DOI: 10.1111/j.1365-3156.2007.01818.x.
- Su Yin, M., Bicout, D. J., Haddawy, P., Schöning, J., Laosiritaworn, Y., and Sa-Angchai, P. (2021). “Added-value of mosquito vector breeding sites from street view images in the risk mapping of dengue incidence in Thailand”. *PLoS neglected tropical diseases* 15.3, e0009122. DOI: 10.1371/journal.pntd.0009122.

- TuzuTa Lin (2023). *labelImg: Graphical image annotation tool and label object bounding boxes in images*.
- Valença, M. A., Marteis, L. S., Steffler, L. M., Silva, A. M., and Santos, R. L. C. (2013). “Dynamics and characterization of *Aedes aegypti* (L.) (Diptera: Culicidae) key breeding sites”. *Neotropical entomology* 42.3, pp. 311–316. DOI: 10.1007/s13744-013-0118-4.
- Vezzani, D. and Schweigmann, N. (2002). “Suitability of containers from different sources as breeding sites of *Aedes aegypti* (L.) in a cemetery of Buenos Aires City, Argentina”. *Memorias do Instituto Oswaldo Cruz* 97.6, pp. 789–792. DOI: 10.1590/S0074-02762002000600006.
- Wang, C.-Y., Liao, H.-Y. M., Yeh, I.-H., Wu, Y.-H., Chen, P.-Y., and Hsieh, J.-W. (2019). “CSPNet: A New Backbone that can Enhance Learning Capability of CNN”. DOI: 10.48550/arXiv.1911.11929.
- Wilke, A. B. B., Vasquez, C., Carvajal, A., Medina, J., Chase, C., Cardenas, G., Mutebi, J.-P., Petrie, W. D., and Beier, J. C. (2020). “Proliferation of *Aedes aegypti* in urban environments mediated by the availability of key aquatic habitats”. *Scientific reports* 10.1, p. 12925. DOI: 10.1038/s41598-020-69759-5.
- Wilson, M. E. (2011). “Chapter 4 Megacities and Emerging Infections: Case Study of Rio de Janeiro, Brazil”. *Megacities & Global Health*. Ed. by O. A. Khan and G. Pappas. American Public Health Association. DOI: 10.2105/9780875530031ch04.

## Publication III: Urban *Aedes aegypti* suitability indicators

**Authors.** Steffen Knoblauch, Rutendo Tadiwa Mukaratirwa, Paulo Filemon Paolucci Pimenta, Annelies Wilder-Smith, Joacim Rocklöv, Sven Lautenbach, Sukanya Randhawa, Oliver J. Brady, Antonio Rocha, Peter Dambach, Thomas Jänisch, Bernd Resch, Filip Biljecki, Till Bärnighausen, Alex Zipf

**Outlet.** *The Lancet Planetary Health*

**Status.** Submitted on 2024-06-05 and under review.

**Abstract.**

Controlling *Aedes aegypti* stands as the primary strategy in curtailing the global threat of vector-borne viral infections such as dengue fever, responsible for around 400 million infections and 40,000 fatalities annually. Effective interventions necessitate a precise understanding of *Ae. aegypti* spatiotemporal distribution and behavior, particularly in urban settings where most infections occur. However, conventionally applied sample-based entomological surveillance systems often fail to capture the high spatial variability of *Ae. aegypti* that can arise from heterogeneous urban landscapes and limited *Aedes* flight range. This study aims at addressing this challenge by leveraging emerging geospatial big data, including openly available satellite and street view imagery, to locate common *Ae. aegypti* breeding habitats. This data enabled to infer the seasonal suitability for *Ae. aegypti* eggs and larvae at a spatial resolution of 200 m within the municipality of Rio de Janeiro. The proposed micro- and macro-habitat indicators for immature *Ae. aegypti* explained the distribution of *Ae. aegypti* ovitrap egg counts by up to 73% and larval counts by up to 75%. Spatiotemporal interpolations of ovitrap counts, utilizing suitability indicators, provided high-resolution insights into the spatial variability of urban immature *Ae. aegypti* that could not be captured with sample-based surveillance techniques alone. The potential of the proposed method lies in synergizing entomological field measurements with digital indicators on urban landscape to guide vector control and address the prevailing spread of *Ae. aegypti*-transmitted viruses. Estimating *Ae. aegypti* distributions considering habitat size is particularly important for targeting novel vector control interventions such as *Wolbachia*.

**Keywords.** big spatiotemporal data · digital urban landscape · suitability indicator · immature *Aedes aegypti* · vector control · Rio de Janeiro

## 1 Introduction

The mosquito species *Ae. aegypti* is the primary vector of yellow fever, dengue fever, Zika, and chikungunya, causing thousands of deaths each year (Camara et al., 2016; Honório et al., 2009; World Mosquito Program, 2023). It favors breeding in artificial water containers commonly encountered near human settlements, such as water tanks (Trewin et al., 2021), discarded tires (Getachew et al., 2015), and storm drains (Paploski et al., 2016). Suitable habitat areas for *Ae. aegypti* will expand due to global trends such as climate change and increasing urbanization (Colón-González et al., 2021; Semenza et al., 2022). The WHO has estimated that by 2080, over 60 percent of the world’s population will live in areas that are likely to be populated by the potential disease vector *Ae. aegypti* (Ebi and Nealon, 2016; Messina et al., 2019). Yellow fever stands out among *Aedes*-borne diseases as the only one for which effective vaccines are globally available. Currently, there is no effective vaccine for Zika, but recently vaccines for dengue fever and chikungunya have been licensed, yet global access and uptake at this stage is low. Therefore, vector control, involving the process of eliminating vector breeding habitats and the application of insecticides to maintain mosquito populations at acceptable level, remains the most effective counter measure (Wilson et al., 2020).

To achieve more efficient and cost-effective vector control in the future, accurate *Ae. aegypti* suitability maps are essential (Boser et al., 2021; Limkittikul et al., 2014; Reiter, 2007). However, generating spatially continuous maps of *Ae. aegypti* for large metropolitan areas proves to be challenging. The limited flight range, assessed to be below 1000 m without the assistance of wind (Getis et al., 2003; Harrington et al., 2005; Honório et al., 2003; Moore and Brown, 2022), and the heterogeneous urban landscape, which influences the availability of breeding sites, can lead to high spatial variability in *Ae. aegypti* abundance (Kache et al., 2022; Louis et al., 2014; Sallam et al., 2017). Capturing this potential variability with conventionally applied sample-based entomological surveillance systems is difficult. In other words, it would require dense coverage of ovitraps or a manual surveillance system at *Aedes* habitat size. Nevertheless, the increasing availability of extensive geospatial data, such as satellite and street view imagery, can aid in bridging this gap (Knoblauch and Moritz, 2023; Knoblauch et al., 2023; Lorenz et al., 2020a; Lorenz et al., 2020b). Particularly noteworthy in this context are indicator-driven interpolation techniques for entomological surveillance data collected during field campaigns (Parra et al., 2022).

Extensive research has been conducted on modeling *Ae. aegypti* abundance at different life cycle stages (egg, larva, pupa, adult) with various sets of proxies on different spatiotemporal scales. Among these studies, inference models for immature *Ae. aegypti* abundance at the mosquito flight range scale have appeared to exert the most considerable impact on local vector control planning (Louis et al., 2014; Sallam et al., 2017). These models concentrate on the early stages of a mosquito’s life cycle, enabling more effective intervention, while also considering the bioecological characteristics of the vector (Boser et al., 2021; Limkittikul et al., 2014; Liu et al., 2017). Lorenz et al. (2020a) and Bailly et al. (2021) proposed methods that take into account the limited flight range of *Ae. aegypti* when modeling its suitability for single neighborhoods in São José do Rio Preto (Brazil) and Cayenne (French Guiana). Sun et al. (2021) applied a flight range model on a larger scale, covering the entire city of Singapore. This study also

incorporated temporal features for spatiotemporal larva abundance maps, similar to the approach by Costa et al. (2015) applied to a small property in the city of Rio de Janeiro (Brazil). Among the identified research on urban *Ae. aegypti* interpolation techniques considering limited mosquito flight range, the research of Portella Ornelas de Melo, Diogo et al. (2012) in Belo Horizonte (Brazil) stands as the only study employing larval survey and ovitrap data together for species distribution modeling.

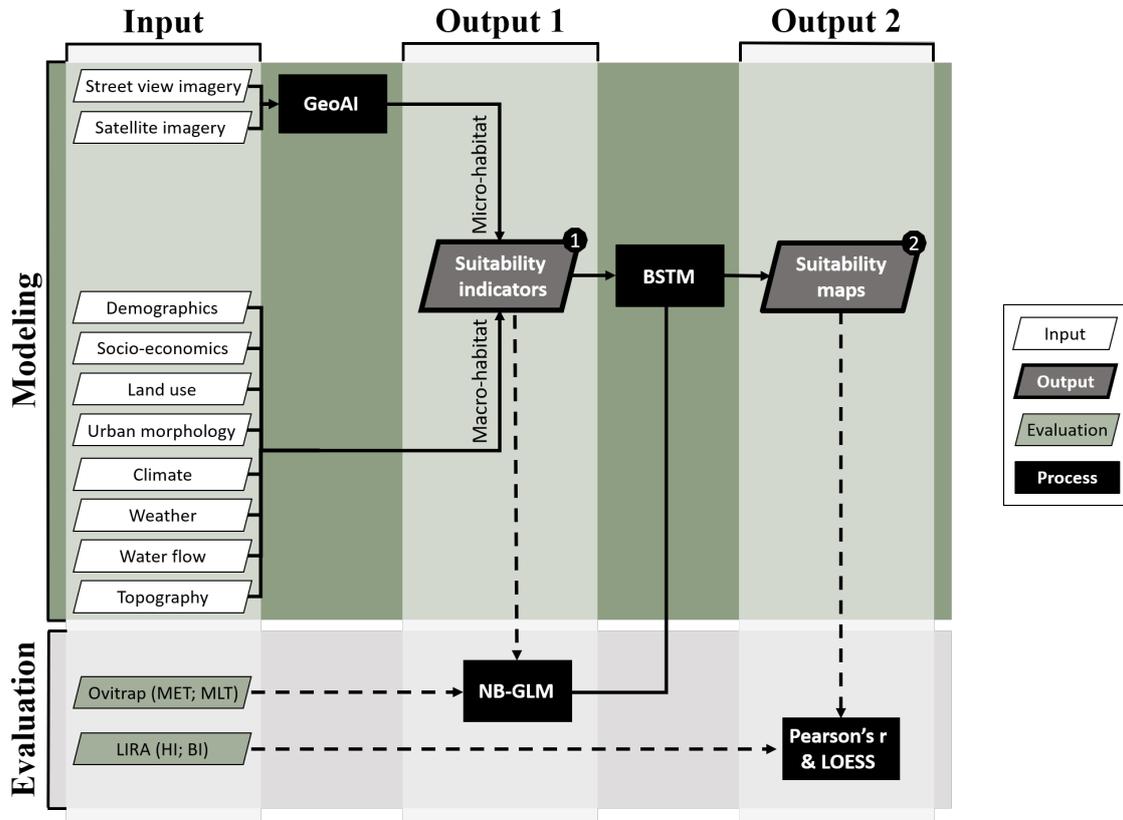
However, as far as our knowledge extends, research that adopts a holistic approach, encompassing i) the restricted flight range of *Ae. aegypti*, ii) the amalgamation of entomological surveillance data from larval surveys and ovitraps, iii) the integration of temporal dynamics, and iv) the extrapolation of *Ae. aegypti* suitability across an entire municipality, is yet to be conducted.

In this paper, we aim to address this research gap by employing a Bayesian spatiotemporal model to construct seasonal *Ae. aegypti* suitability maps at the *Aedes* flight range scale for the municipality of Rio de Janeiro, Brazil. As a prerequisite for this, we generated micro- and macrohabitat suitability indicators for immature *Ae. aegypti*, capturing, for example, *Ae. aegypti* breeding container density and rainwater accumulation, from openly available geospatial data sources. The predictive power of these generated suitability indicators was evaluated using a negative-binomial GLM alongside monthly egg and larval counts from ovitraps. The suitability maps for immature *Ae. aegypti*, built upon ovitrap counts and generated indicators, were assessed alongside the seasonal House and Breteau indices from the Rapid Assay of the Larval Index for *Aedes aegypti* (LIRAA) (cf. Figure 1).

## 2 Material and Methods

Here, we propose a novel framework for the spatiotemporal mapping of immature *Ae. aegypti* suitability in urban spaces (cf. Figure 1). Our framework mainly consists of two outputs: (i) suitability indicators for immature *Ae. aegypti* and (ii) seasonal suitability maps for immature *Ae. aegypti* at *Aedes* flight range scale, generated from suitability indicators and ovitrap counts. We applied the proposed framework to the *Ae. aegypti*-endemic municipality of Rio de Janeiro (Gibson et al., 2014; Wilson, 2011). Recent data from 2023 revealed over 3 million dengue fever cases in Brazil, underlining the urgent need to enhance vector control in major Brazilian cities. We chose the year 2019 as our analysis time period because it marked the largest dengue fever outbreak in the municipality of Rio de Janeiro within the last 5 years (Secretario Municipal de Saude Rio de Janeiro, 2024b). With its year-round tropical climate (Franco dos Santos et al., 2022), a population of around 6.75 million people, and high connectivity to other urban areas in Latin America, the second-largest city in Brazil has often served as a starting point for larger, uncontrolled disease outbreaks across the Americas, including dengue fever (Moraes Figueiredo, 2004). The proximity of different types of urban structures, such as favelas and other residential areas, coupled with the city's topography, accounts for a high variability of possible *Ae. aegypti* breeding sites, making the city of Rio de Janeiro an intriguing use case for our proposed method. An extended description of the applied entomological surveillance conducted in the municipality of Rio de Janeiro, which was used in our

study to evaluate the suitability indicators for immature *Ae. aegypti*, is provided in Appendix C.

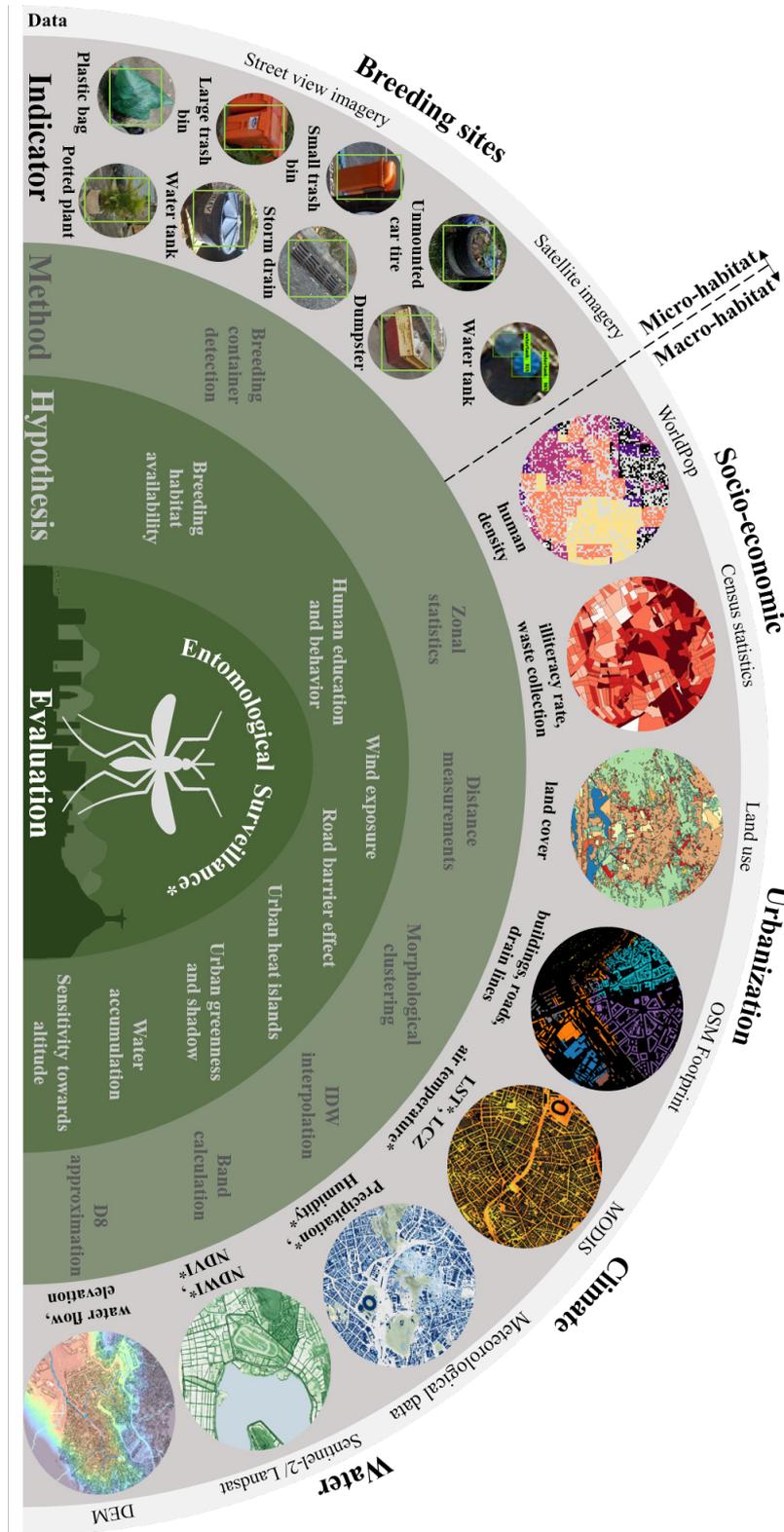


**Figure 1:** Framework for modeling and evaluating (i) immature *Ae. aegypti* suitability indicators and (ii) immature *Ae. aegypti* suitability maps at *Aedes* flight range resolution. Openly available geodata served as input for our framework. The abbreviation GeoAI refers to geospatial techniques of artificial intelligence. In this study it represents the detection and mapping of common *Ae. aegypti* breeding containers from satellite and street view imagery. The abbreviation BSTM refers to a Bayesian spatiotemporal model. In this study, we fitted a BSTM with the integrated nested Laplace approximation (INLA) to generate seasonal suitability maps for immature *Ae. aegypti* covering the whole municipality of Rio de Janeiro at *Aedes* flight range resolution using suitability indicators and ovitrap counts. Entomological surveillance data form ovitraps and the Rapid Assay of the Larval Index for *Ae. aegypti* (LIRAA) were applied for evaluation. MET and MLT stand for the mean egg per trap and mean larva per trap rate collected monthly via ovitraps. BI and HI represent the Breteau - and House indices collected during LIRAA.

## 2.1 Generation of suitability indicators

Urban suitability indicators for immature *Ae. aegypti* were selected based on availability and a priori expectation of factors influencing immature *Ae. aegypti* abundance. Spatial as well as spatiotemporal covariates with differing resolutions were retrieved to interpolate entomological surveillance data considering the limited mosquito flight range. The identified locations of common *Ae. aegypti* breeding sites were considered as micro-habitat indicators, while those indicators collected at a coarser spatial scale were classified as macro-habitat indicators (cf. Figure 2).

In this study, micro-habitat indicators were modeled concerning typical breeding sites of *Ae. aegypti*, which include artificial water containers such as water tanks, potted plants, trash bins, unmounted car tires, or dumpsters often found in close vicinity to human settlements. All these containers can harbor stagnant water after rainfall, which is highly suitable for *Ae. aegypti* oviposition and subsequent adult population development. We hypothesized that their spatial distribution and occurrence, in the form of container density, could serve as a reliable indicator for the abundance of immature *Ae. aegypti* in urban environments. In addition to micro-habitat urban suitability indicators, we hypothesized a range of additional macro-habitat urban suitability indicators for immature *Ae. aegypti*. These indicators encompass a broad range of spatiotemporal proxies describing urban landscape in terms of demography, socio-economy, land use, climate, weather, green spaces, and water availability. The corresponding hypotheses were derived from previous literature (cf. Appendix A). An extended description of the applied methods and the hypothesized spatiotemporal influences of suitability indicators for the abundance of immature *Ae. aegypti* can be found in Appendix B.



**Figure 2:** Grouped stack of self-generated urban suitability indicators for immature *Ae. aegypti*, featuring data sources, methods employed for retrieval, and formulated hypotheses for evaluation with entomological surveillance data. A more dense collection of proposed indicators with references for assumed hypotheses and data source is listed in the supplements. Indicators marked with an asterisk (\*) represent spatiotemporal factors, while those without an asterisk (\*) were considered to be temporally constant in our case study.

## 2.2 Assessment of suitability indicators for immature *Ae. aegypti*

To quantitatively evaluate how well urban indicators can capture the inner-urban distribution of immature *Ae. aegypti* abundance measured by entomological ovitrap data, we ran a negative-binomial generalized linear regression model (NB-GLM) with a log-link function (Hilbe, 2012). We ran this model for several estimated mosquito flight range scenarios to additionally test the robustness of our results. The NB-GLM was selected as it allows the model to account for the overdispersion present in the applied entomological count data (cf. Appendix C). Corresponding equations and further explanations were given in Appendix D.

To assess the capability of indicator-driven interpolations of ovitrap counts to estimate seasonal indices from LIRA, we organized monthly ovitrap data into quarters that corresponded to the four LIRAA seasons. This grouping was performed using the feature vector of the best-performing flight range buffer, identified through the NB-GLMs. Subsequently, we fitted a Bayesian spatiotemporal model with INLA (Rue and Lindgren, 2024), to generate seasonal and spatially continuous urban suitability maps for immature *Ae. aegypti* covering the whole municipality of Rio de Janeiro. A more detailed description of the modeling was provided in Appendix E.

To quantitatively assess the generated urban suitability maps for immature *Ae. aegypti*, we compared the seasonal posterior means of the spatial random effects with seasonal LIRAA measurements using scatter plots. Additionally, we calculated the Pearson's correlation coefficient and applied locally weighted scatterplot smoothing (LOESS) across all seasons. To achieve this, zonal statistics were performed on LIRAA strata for each of the four seasons (Jan-Mar; Apr-Jun; Jul-Sep; Oct-Dec) and response variables, respectively. Before calculating zonal statistics, continuous egg and larva interpolations were clipped using the urbanization area to avoid false inference and high bias, as interpolations were created using egg and larval counts from urbanised area only. The results were then compared with mean values of ovitrap counts from the field.

### 3 Results

#### 3.1 Inference capacity of suitability indicators

The results presented in Table 1 underscore the degree to which hypothesis-driven urban indicators for *Ae. aegypti* suitability can capture entomological surveillance data on immature abundance collected via ovitraps in the municipality of Rio de Janeiro for the year 2019, given the constraints of a limited *Aedes* flight range. The Cohen’s explained deviance for NB-GLMs, using the seasonal mean eggs per trap (S-MET) rate as a response variable, reached up to 0.7253 and varied only marginally ( $\pm 0.003$ ) for different simulated flight range buffers. Increasing the flight range buffer from 50m to 1000m exhibited similar patterns for models using the larval count as a response variable. In this case, the predictive performance of the collected urban indicators was slightly higher, reaching a Cohen’s pseudo- $R^2$  of 0.7473 at a flight range of 200 m. This means that 74.73% of the deviance in the response is explained by hypothesis-driven urban indicators for immature *Ae. aegypti* suitability derived from openly available geospatial data. The deviance function of the NB-GLM captured the increasing variance with the mean that is typical for count data. Upon evaluating models with respect to both designated response variables, the best performance was observed in association with a flight range buffer characterized by a diameter of 200 m.

**Table 1:** Cohen’s explained deviance for NB-GLMs using seasonal mean eggs per trap (S-MET) and seasonal mean larva per trap (S-MLT) rates of 2019 as response variables. Urban suitability indicators for immature *Ae. aegypti*, employed as explanatory variables, were collected under various *Aedes* flight range scenarios within different estimated flight range buffers around ovitrap locations. The results indicate that the proposed method allows for modeling on different spatial scales with consistent performance. The flight range scenario with the best performance for each entomological index was indicated in bold. The 200 m flight range scenario exhibited the highest Cohen’s pseudo- $R^2$  when evaluating the combined results of S-MET and S-MLT. Ovitrap locations were treated as independent observations, indicating restricted *Ae. aegypti* dispersal.

Entomological index	<i>Aedes</i> flight range scenarios	Cohen’s pseudo- $R^2$
S-MET	50m	0.7243
	100 m	0.7226
	<b>200 m</b>	<b>0.7246</b>
	400 m	0.7223
	<b>1 000 m</b>	<b>0.7253</b>
S-MLT	50m	0.7468
	100 m	0.7444
	<b>200 m</b>	<b>0.7473</b>
	400 m	0.7443
	1 000 m	0.7463

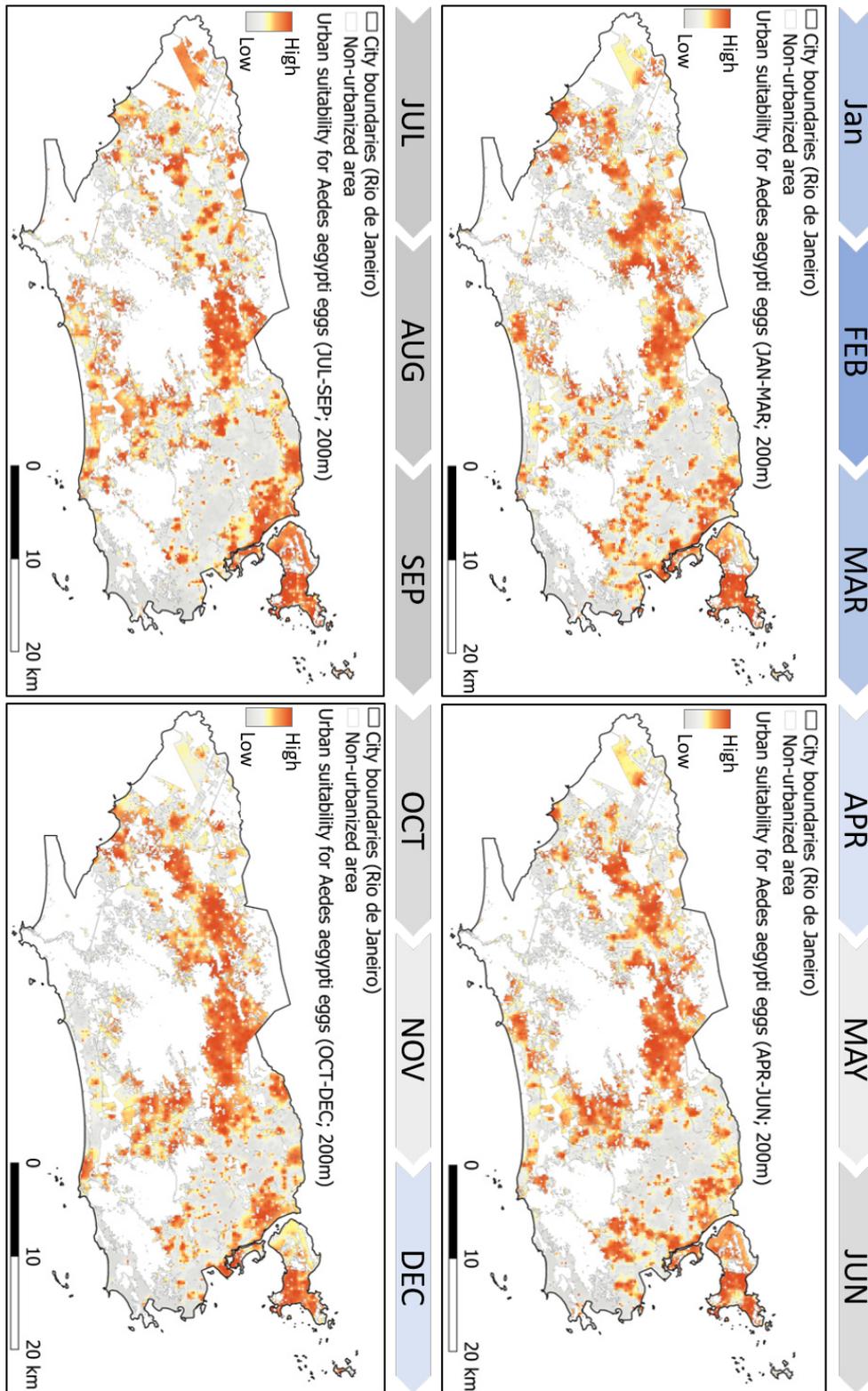
S-MET = seasonal mean egg per trap rate; S-MLT = seasonal mean larva per trap rate

### 3.2 Seasonal suitability maps for the municipality of Rio de Janeiro

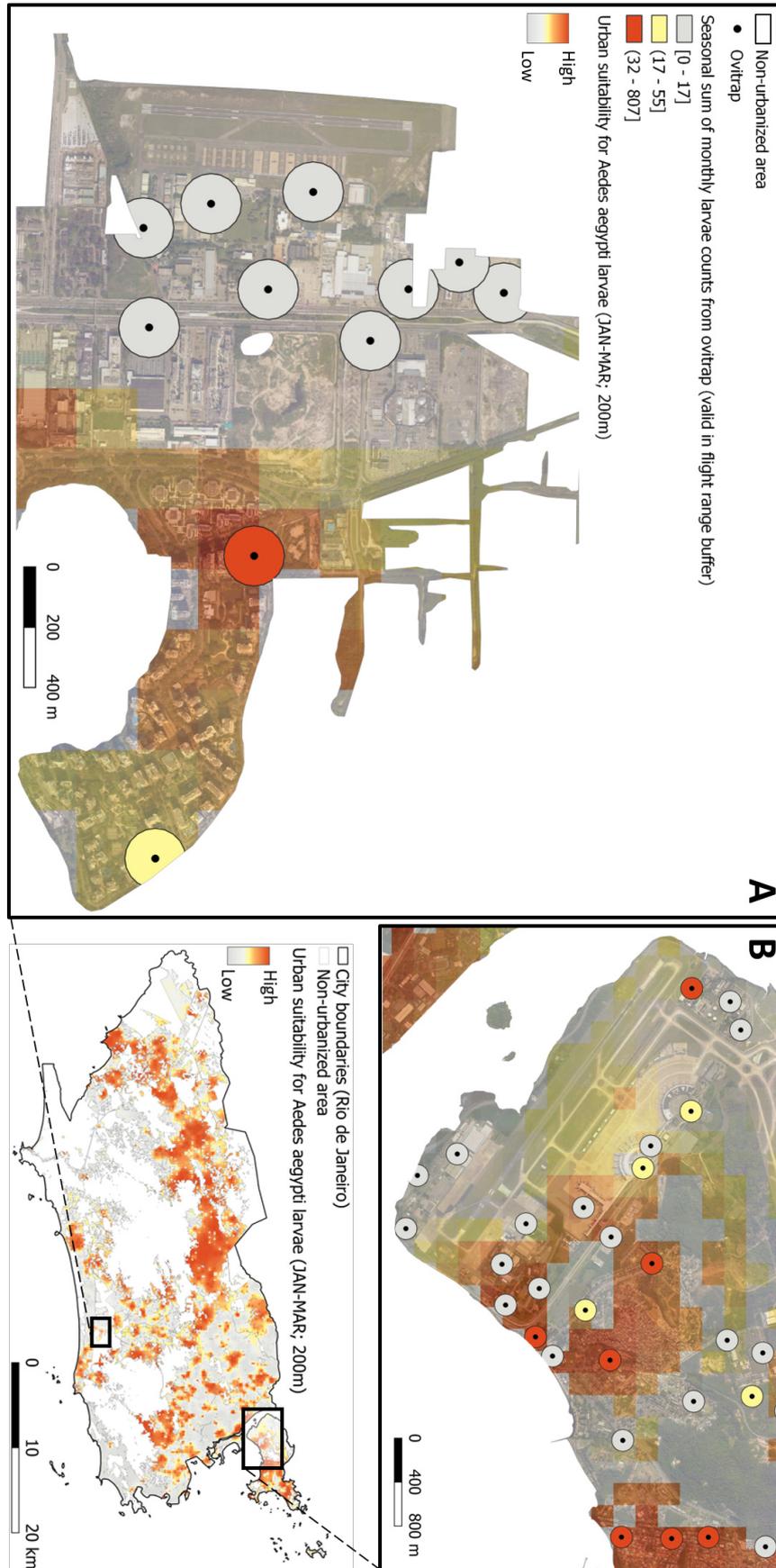
Figure 3 displays, as a highlight of this work, seasonal suitability maps for immature *Ae. aegypti* covering the municipality of Rio de Janeiro. The seasonal immature *Ae. aegypti* suitability maps accompany the spatiotemporal trend of entomological field measurements is presented in Figure C. 1. While the spatial variance of predicted immature *Ae. aegypti* suitability diverges significantly due to the small-scale heterogeneity of the urban landscape in the city of Rio de Janeiro, temporal effects are minimal owing to the year-long (sub)tropical climate conditions in southeast Brazil.

Figure 4 provides a more detailed insight into the results by illustrating how our best spatiotemporal model for *Ae. aegypti* larvae suitability (cf. Table 1) performs in interpolating entomological field measurements from ovitrap locations. Specifically, it focuses on the exemplary regions of Jacarepaguá (RRJ) and Galeão (GIG) airports, chosen for their distinct spatial heterogeneity in the urban landscape, which enables a closer examination of the results at a finer spatial scale. While ovitrap larval counts and interpolated immature *Ae. aegypti* suitability remained predominantly low around both airport runways and buildings, abundance values were higher in nearby residential regions when examining measurements from the summer season of January to March 2019. The spatial heterogeneity observed in immature *Ae. aegypti* suitability at a small scale, as depicted in the map, highlights the impracticality of relying solely on the current state-of-the-art approach of coarse entomological surveillance at ovitrap sample locations or within large LIRAA strata (cf. Figure C. 1) for targeted vector control interventions (Flores and O’Neill, 2018).

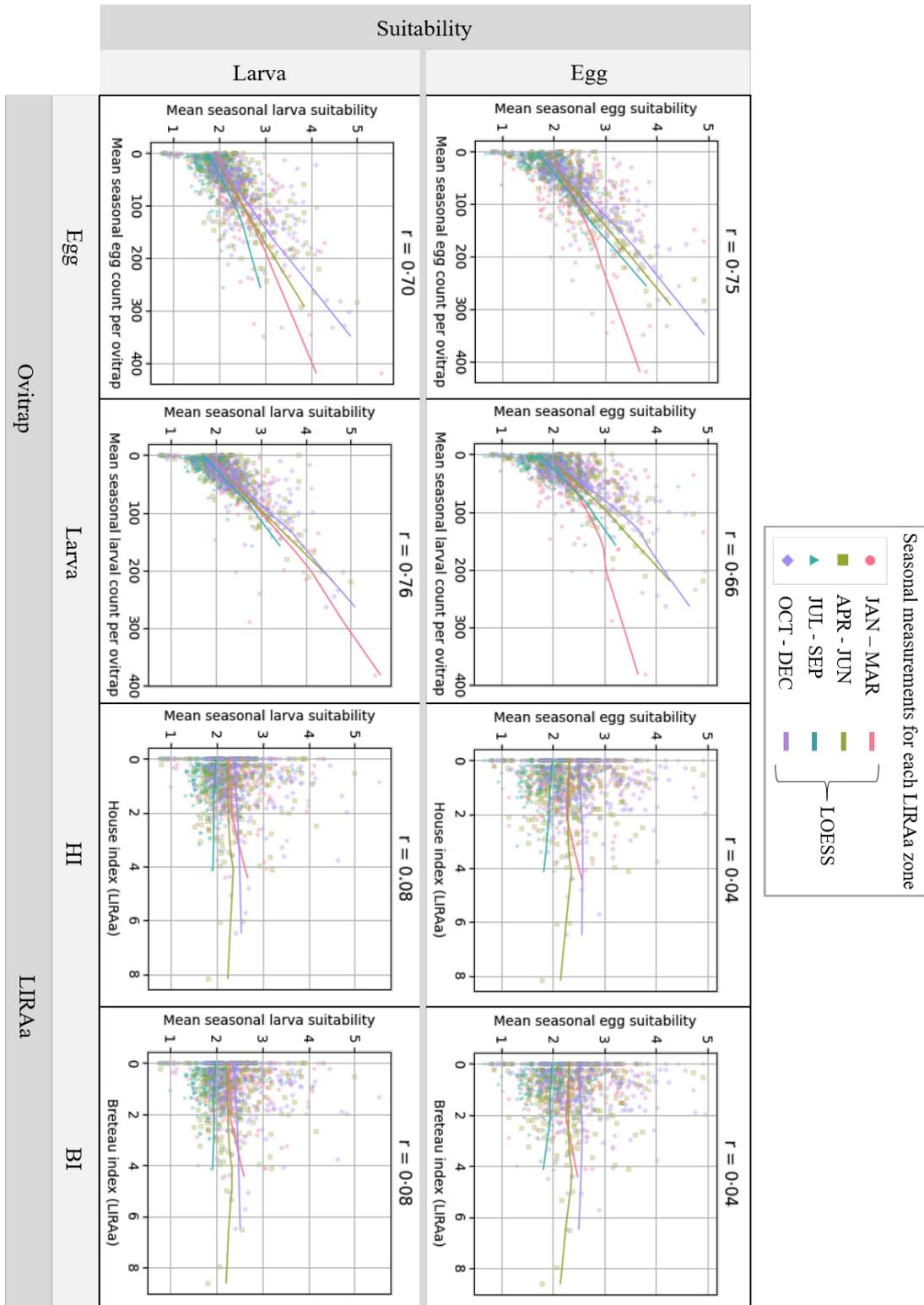
Figure 5 depicts the alignment between the generated suitability maps and spatiotemporal measurements obtained from entomological surveillance (ovitrap; LIRAA). Using locally estimated scatterplot smoothing (LOESS) analysis, we observed that the predicted suitability values did not consistently align with seasonal indices derived from LIRAA, regardless of the season (cf. Table 2). Correlation analysis between predicted suitability and ovitrap field counts, averaged for each LIRAA zone over all seasons, reveals correlation coefficients of up to 0.76 for *Ae. aegypti* larval count. However, the correlation between suitability predictions and LIRAA indices remains low, not exceeding a correlation coefficient of 0.08, as calculated between larva suitability and the LIRAA house index for the year 2019.



**Figure 3:** Seasonal suitability maps for *Ae. aegypti* eggs in 200 m resolution covering the urbanized area of the city of Rio de Janeiro for 2019, using a synergetic approach of entomological surveillance, urban landscape indicators for immature *Ae. aegypti* and bio-ecological knowledge on limited *Aedes* flight range. The months spanning from December to April align with the wet season in the municipality of Rio de Janeiro, while the period from May to November corresponds to the dry season. Seasonal suitability maps for *Ae. aegypti* larvae are shown in Appendix F



**Figure 4:** Entomological surveillance measurements from ovitraps and interpolated urban suitability for *Ae. aegypti* larvae between January and March 2019 at a 200 m scale around Jacarepaguá (RRJ) and Galeão (GIG) airport in the city of Rio de Janeiro. ©2024 Google



**Figure 5:** Pearson's correlation coefficients, scatterplots, and locally estimated scatterplot smoothing (LOESS) for seasonal measurements of ovitrap, LIRAA and immature *Ae. aegypti* suitability aggregated on LIRAA zones for the municipality of Rio de Janeiro in 2019. (BI: Breteau index, HI: house index).

**Table 2:** Pearson’s correlation coefficients for seasonal measurements of ovitrap, LIRAA and immature *Ae. aegypti* suitability aggregated on LIRAA zones for the municipality of Rio de Janeiro in 2019. The months spanning from December to April align with the wet season in the municipality of Rio de Janeiro, while the period from May to November corresponds to the dry season.

Suitability	Season	Egg (Ovitrap)	Larva (Ovitrap)	HI (LIRAA)	BI (LIRAA)
Egg	JAN-MAR	0.55	0.48	-0.01	-0.01
	APR-JUN	0.83	0.74	-0.04	-0.04
	JUL-SEP	0.73	0.60	-0.08	-0.09
	OCT-DEC	0.81	0.72	0.07	0.07
Larva	JAN-MAR	0.70	0.76	0.08	0.09
	APR-JUN	0.71	0.77	-0.01	-0.01
	JUL-SEP	0.53	0.66	-0.05	-0.05
	OCT-DEC	0.69	0.76	0.10	0.10

## 4 Discussion

### 4.1 Enhancing vector control strategies through urban landscape indicators

Entomological surveillance plays a crucial role in guiding vector control strategies aimed at mitigating the transmission of *Aedes*-borne diseases. However, traditional sample-based methodologies used in surveillance efforts often fail to capture the complex spatial dynamics of *Ae. aegypti* abundance, particularly in heterogeneous urban environments such as the municipality of Rio de Janeiro. The high spatial variability in mosquito populations is influenced by factors such as small habitat size and diverse landscape characteristics, which create varied breeding opportunities for *Ae. aegypti*. In large municipal areas, where dengue fever outbreaks are more frequent and vector control targeting is of paramount importance due to higher disease incidences, understanding the spatial distribution of *Ae. aegypti* becomes essential. Limited public health resources necessitate a strategic focus on priority areas where *Ae. aegypti* populations are most concentrated to maximize the impact of vector control efforts.

To address the challenge of spatially targeted vector control, we developed a framework to generate hypothesis-driven urban landscape indicators to model urban suitability for immature *Ae. aegypti*. These indicators derived from openly available geospatial data sources were applied to enrich entomological surveillance and create continuous urban suitability maps at *Aedes* habitat size. By integrating information on landscape characteristics with ovitrap data, our approach provided valuable insights into the spatial distribution of *Ae. aegypti* populations in the municipality of Rio de Janeiro and identified priority areas for intervention (cf. Figure 3). However, it is important to acknowledge the potential limitations associated with the use of digital indicators, including data availability, accuracy, and interpretation. Additional value was particularly generated through the complementary application of both digital indicators and entomological surveillance. An alternative selection of relevant indicators could potentially further improve our results. However, limits of the proposed framework will always be given by the bias in the entomological collection process and due to non-measurable micro-scale circumstances affecting the entomological count data, applied for indicator

validation. Despite these challenges, our findings represent a significant advancement in the field of vector control targeting and offer valuable guidance for public health practitioners in the municipality of Rio de Janeiro and policymakers in their efforts to combat *Aedes*-borne diseases.

Our study's innovation encompassed (i) the high spatial resolution of immature *Ae. aegypti* suitability maps, (ii) the incorporation of digital indicators, including the density of common *Ae. aegypti* breeding sites to model *Ae. aegypti* micro-habitats, and (iii) the comprehensive comparison of ovitrap-based field counts and suitability interpolations with block-level LIRAA indices collected across an entire municipal area, accounting for the limited flight range of *Aedes* mosquitoes. However, the current transferability of our conceptual framework to other urban areas is constrained by significant labour-costs, as the generation of spatiotemporal indicators remains resource-intensive, albeit reliant on openly available data sources. The choice of urban immature *Ae. aegypti* suitability indicators may have also influenced our main findings. Exploring different indicators could yield varied spatial distribution patterns of immature *Ae. aegypti* suitability estimates, potentially shaping the outcomes of our analysis.

Future research should investigate the sensitivity of our results to (i) diverse indicator selections, (ii) entomological surveillance data from multiple years, and (iii) other case study regions, in order to enhance the understanding of the robustness of our findings. Follow-up activities are planned to build upon our framework to derive suitability indicators for the secondary vector of dengue *Ae. albopictus*, which has been reported in Brazil for almost 30 years, for transmitting yellow fever virus (YFV), dengue virus (DENV), Zika virus (ZIKV), and chikungunya virus (CHIKV) (Pancetti et al., 2015; Ricas Rezende et al., 2020). Further research could then also analyze the feature importance of the proposed indicators across various mosquito species.

## 4.2 Implications for dengue control in the municipality of Rio de Janeiro

The “Brazilian Guidelines for Prevention and Control of Arboviruses” (Vinhali Frutuoso and Barbosa Duraes, 2023) advocate for targeted actions, particularly in large municipalities with more than one million inhabitants, such as the municipality of Rio de Janeiro. Between 2013 and 2022, 52% of probable dengue fever cases in Brazil were concentrated in municipalities with a population equal to or greater than 100 000. The Brazilian Ministry of Health recommends the implementation of several key strategies: (i) entomological monitoring using ovitraps, (ii) household residual spraying (BRI-*Aedes*), (iii) the deployment of larvicide spraying stations, (iv) the introduction of mosquitoes carrying *Wolbachia*, and (v) the utilization of sterile insect techniques to control *Ae. aegypti*. These technologies should be deployed on the basis of an action plan, which requires intra-municipal risk stratification, and should always be accompanied by home visits, depending on the area at risk, and actions to interface with society.

The presented findings in this study aim to align with this official guideline by proposing urban immature *Ae. aegypti* suitability indicators (cf. Figure 2) that can be applied for risk stratification, as shown in Figure 3. Here, the term “stratification” refers to the classification of the risk of endemic areas based on their eco-epidemiological characteristics. This approach aids in identifying areas

that require distinct approaches to arbovirus control. Risk stratification serves as a tool to organize prevention and control activities at the municipal level, whether in priority or non-priority areas, during periods of low transmission or during the preparatory phase. By providing a method for assessing the suitability of areas for *Ae. aegypti* breeding, the proposed indicators in this study facilitate the implementation of targeted vector control measures, thereby reducing the impact of arbovirus epidemics and intensifying control actions in higher-risk areas.

Our results are designed to particularly support vector control efforts in the municipality of Rio de Janeiro by creating a more accurate action plan that goes beyond relying solely on sample-based entomological surveillance and basic hotspot analysis. Instead, we propose a more advanced approach considering the heterogeneous nature of urban landscapes at the scale of *Aedes* flight range. Our generated suitability maps (cf. Figure 3) not only pinpoint priority action areas but also assign priority levels based on *Ae. aegypti* immature suitability values. This nuanced approach allows for the tailored selection of interventions, guided by priority scores alongside a cost-benefit analysis, resulting in a more efficient overall vector control strategy.

For high-priority areas, technical methods such as (i) dissemination stations containing larvicide, (ii) sterile insect techniques, or (iii) the *Wolbachia* method could be relevant to solve structural problems arising from socio-economic inequalities in water supply and solid waste collection within the municipality of Rio de Janeiro. *Wolbachia*, a naturally occurring bacterium, can be introduced into mosquito cells to curb the transmission of viruses by *Ae. aegypti* and influence mating outcomes, thereby aiding its spread and sustainability in natural mosquito populations. Larvicides have been employed in a rotation scheme in Brazil since 2012. In this scheme, the products used include *Bacillus thuringiensis israelensis* (BTI), insect growth regulators (IGR) such as *juvenile hormone analogs* (JHA) or *chitin synthesis inhibitors* (CSI), and more recently, *Spinosad*, a neurotoxic insecticide (Valle and Aguiar, 2023).

Conversely, in low-priority areas, action may be triggered only upon reaching a threshold of egg and larva density. Here, potential actions may encompass: (i) mechanical methods such as the elimination of stagnant water in common breeding site, (ii) the application of the larvicide *Temephos* to rainwater tanks, or (iii) launching health education initiatives to engage the community. When targeting community-engaged breeding site removal based on the presented suitability maps, it is vital to consider additional socio-economic gradients. Individuals from diverse socio-economic backgrounds may prioritize different actions, such as safety, food security, access to clean water and sanitation facilities, health-care services, education, and employment opportunities. Besides the tailored selection of interventions for high and low priority areas, universal vector control measures should be consistently implemented in all regions throughout the year. These measures could include house-to-house visits, inspections of strategic points such as cemeteries, tire repair shops, junkyards, scrap metal or building materials deposits, and bus garages, as well as household residual spraying for *Aedes* mosquitoes (BRI-*Aedes*).

## 5 Conclusion

In this paper, we demonstrated the potential of retrieving immature *Ae. aegypti* suitability indicators from openly available geodata, to model the urban likelihood of hosting mature *Ae. aegypti* populations considering limited *Aedes* flight range. Such high-resolution maps are essential to (i) inform and optimize targeted vector control interventions such as *Wolbachia*, (ii) allow cost savings in entomological surveillance, (iii) reduce environmental pollution, including mosquito insecticide resistance, and most importantly, (iv) provide more efficient overall disease control. The proposed synergistic method of integrating entomological surveillance with bio-ecological knowledge and digital landscape indicators yielded insights into the high spatial variability of urban immature *Ae. aegypti* distributions in the municipality of Rio de Janeiro, which cannot be captured by sample-based surveillance techniques alone. Scientific advancements were particularly achieved by this study design in the realm of spatial resolution, while temporal modeling remained coarse due to the absence of entomological field measurements at daily time intervals corresponding to the mosquito life cycle. Further investigation in other cities embracing the Digital Urban Twin concept, particularly when coupled with emerging smart trap technologies to enhance the temporal resolution of suitability inference, appears promising, notwithstanding the ambiguity surrounding the relationship between adult and larvae abundance. With this major contribution from our interdisciplinary research, we aim to create new pathways for science in computational eco-epidemiology. Additionally, we seek to provide useful datasets for future research on inner-urban pathogen transmission dynamics and to support public health authorities in the *Ae. aegypti*-endemic city of Rio de Janeiro in developing more focused vector control strategies where scalability in urban settings remains challenging.

## Data statement

The materials and datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Restrictions apply only to the sharing of entomological surveillance data collected by the Municipal Health Ministry of Rio de Janeiro, for which access should be granted directly from there.

## Acknowledgements

The work was funded by the German Research Foundation (DFG) (grant number 451956976) and the Austrian Science Fund (FWF) (grant number I 5117).

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Credit authorship contribution statement

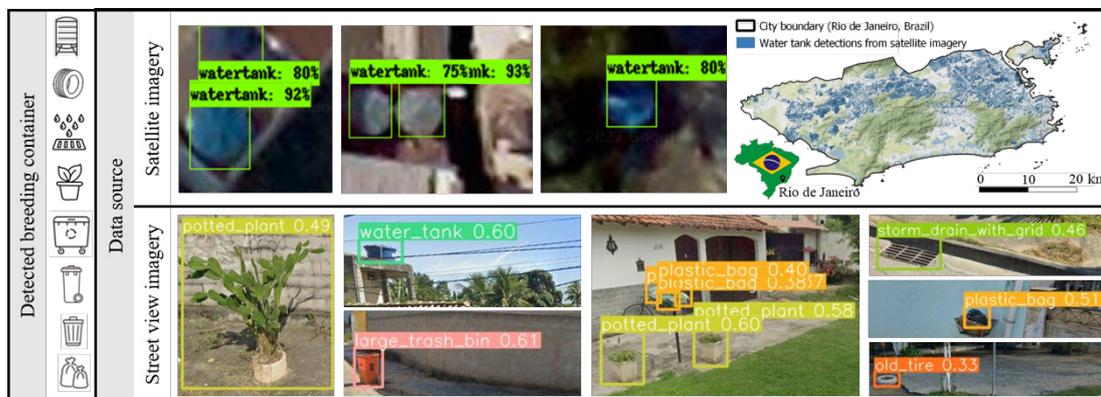
SK conceptualized the study, conducted the experiment, and drafted the manuscript. SK and RM worked together on the software for Bayesian modelling and both validated the results. SK and MY worked together on the software for object detection and both validated the results. SR provided resources for water flow modelling. SK and AR acquired entomological surveillance data. AZ, AWS, BR, FB, JR, MY, OB, PP, PD, PH, SL, SK, TB, and TJ reviewed and edited the manuscript. AZ, BR, SL, and TJ supported acquisition of funding. AZ and SL supervised the study. All authors read and approved the manuscript and its subsequent revisions prior to submission.



## Appendix B - Methods for the retrieval of suitability indicators

### Appendix B.1 - Micro-habitat indicators

The generation of water tank counts as a micro-habitat indicator derived from satellite imagery was extensively described in our previous research study for the municipality of Rio de Janeiro (Knoblauch et al., 2023). In this prior work, we devised a semi-supervised self-training algorithm to decrease the need for manual labeling in the automated detection of water tanks within urban areas using satellite imagery. The trained water tank detection model yielded a precision score of 0.864, a recall of 0.823, and an F1 score of 0.843 on independent test datasets. In addition to the city-wide water tank detection, we derived the density of trash cans, catch basins, manholes and water valves from the Mapillary API (Mapillary, 2023). For the mapping of potted plants, small and large trash bins, dumpsters, storm drains, unmounted car tires, and plastic bags, we fine-tuned a YOLOv5 model using street view imagery. Applied methods were extensively described and evaluated in Knoblauch et al. (2024). The trained computer vision model for multi-class object detection achieved a precision score of 0.853, a recall of 0.904, and an F1 score of 0.878, weighted by instance average across breeding containers in independent test datasets. Figure B. 1 illustrates some true positive examples of container detection generated in our previous studies (Knoblauch et al., 2024; Knoblauch et al., 2023). The statistical significance of these novel micro-scale suitability indicators for immature *Ae. aegypti* was established in our previous studies conducted in the municipality of Rio de Janeiro from 2019 to 2022. However, the significance was discussed to depend on the timestamp of satellite and street view imagery, solid waste collection, the ephemeral nature of some breeding container types, the spatial coverage of street view images limiting the completeness of breeding site detection (Hou and Biljecki, 2022), and interventions for large-scale breeding site removal.



**Figure B. 1:** True positive examples of *Ae. aegypti* breeding site detection applying computer vision models on satellite and street view imagery to create micro-habitat urban suitability indicators for immature *Ae. aegypti* covering the whole municipal area of Rio de Janeiro. The top-right map illustrates water tank density in the city of Rio de Janeiro, derived from object detections (Knoblauch et al., 2023). ©2024 Google

## Appendix B.2 - Macro-habitat indicators

The indicators of human population density and building density were selected as macro-habitat proxies of human influence on the *Ae. aegypti* population, considering that human activities provide artificial water containers suitable as breeding habitats (Espinosa et al., 2016; Kamgang et al., 2010; Lindsay et al., 2017; Tedjou et al., 2019; Wilson-Bahun et al., 2020). The indicators slope and water flow accumulation, defined by the Horton-Strahler number, were selected in consideration of their influence on water accumulation (Cornel et al., 2016). The water flow indicator was generated by applying a D8 approximation algorithm to 5 m elevation data provided upon request by the Urban Data Platform from PPGAU UFF (PPGAU UFF, 2023a). The indicator elevation level, including building heights, was additionally added as a covariate to account for *Ae. aegypti*'s sensitivity to altitude (Equihua et al., 2017; Liew and Curtis, 2004; Lozano-Fuentes et al., 2012; Moreno-Madriñán et al., 2014; Roslan et al., 2022; Roslan et al., 2013). The indicator local climate zones was selected to consider *Ae. aegypti* climate-sensitive reproduction and fertility rate (Azevedo et al., 2018; Jesús Crespo and Rogers, 2021). This indicator - based on urban climate estimates by Demuzere et al. (2021) - considers ten different urban built-up types (compact highrise, - midrise, - lowrise, open highrise, - midrise, - lowrise, lightweight lowrise, large lowrise, sparsely built, heavy industry) influencing shadow and heat accumulation together with seven land cover classes (dense, trees, scattered trees, bush and scrub, back rock or paved, bare soil or sand, water) in 30 m resolution. The monthly indicators of air temperature (Chang et al., 2007; Lambrechts et al., 2011; Misslin et al., 2018; Tsuda and Takagi, 2001), precipitation (Barrera et al., 2011; Li et al., 1985; Souza et al., 2010; Stewart Ibarra et al., 2013; Valdez et al., 2018; Vasconcelos et al., 2022; Vasconcelos et al., 2021), and relative humidity (Costa et al., 2010; Lega et al., 2017; Nasir et al., 2017; Reiskind and Lounibos, 2009) were derived from the Alerta system (Centro de Operacoes Rio, 2023). Therefore, we interpolated 15-minute interval measurements from 33 weather stations during the study period of 2019 to consider both climate and meteorological effects on *Ae. aegypti* populations. Additional urban heat island effects (Araujo et al., 2015; Oliveira Lemos et al., 2021; Wilk-da-Silva et al., 2018) were retrieved by Lucena et al. (2015), Peres et al. (2018), and Miranda et al. (2022) including cloud masking techniques, atmospheric correction and surface emissivity. As a further proxy to describe the habitat suitability of immature *Ae. aegypti* in urban landscape, we calculated the road network density from OpenStreetMap to consider the barrier effects of roads on mosquito populations (Kaplan et al., 2010; Regilme et al., 2021). The distance to coastal water bodies was also generated utilizing OSM to account for additional wind exposure effects with a negative influence on mosquito activity (Wong and Jim, 2017). The distance and coverage of urban drain lines were derived from a hydrographic data set (Data.Rio, 2023b) as an additional urban-specific proxy for immature *Ae. aegypti* populations. Normalized difference vegetation index (NDVI) (Britos Molinas et al., 2022; Chaves et al., 2021; Estallo et al., 2018; Estallo et al., 2008; Martín et al., 2022a; Martín et al., 2022b) and normalized difference water index (NDWI) (Britos Molinas et al., 2022; Estallo et al., 2018; Estallo et al., 2012; German et al., 2018) were computed using Sentinel-2 satellite imagery from the European Space Agency to consider vegetation types and water availability influencing *Ae. aegypti* especially

in non-built up areas. An algorithm for cloud masking was applied to calculate the mean of cloud-free pixels at a 30 m resolution from January 2019 until December 2019 using the Google Earth Engine. After band calculations, a threshold of  $\geq 0.2$  for the NDVI and  $\geq 0.3$  for the NDWI was applied to avoid false assumptions. On top of this, land cover maps were extracted from DataRioPortal (Data.Rio, 2023a) to incorporate land use classes (Albrieu-Llinás et al., 2018; Benitez et al., 2020; Egid et al., 2022; Landau and van Leeuwen, 2012; Lorenz et al., 2020a; Montagner et al., 2018; Vanwambeke et al., 2007; Westby et al., 2021; Young et al., 2017; Zahouli et al., 2017) such as the location of favelas and to calculate the minimum distance from ovitraps to forest areas to consider forest-specific climate effects such as locally increased humidity (Costa et al., 2010; Rowley and Graham, 1968). An urban morphological clustering was computed using the momepy python library (Fleischmann, 2019) and official building footprints provided upon request by the Urban Data Platform from PPGAU UFF (PPGAU UFF, 2023b). Most recent census statistics for 10,233 strata such as the amount of collected rubbish (Bonnet et al., 2020; Chumsri et al., 2020; Manrique-Saide et al., 2008; Maquart et al., 2022; Stewart Ibarra et al., 2014; Whelan et al., 2020), statistics about waste water management (Burke et al., 2010; Chan et al., 1971; Martini et al., 2019; Novaes et al., 2022), sanitation (Gomes et al., 2023), and education level (Menchaca-Armenta et al., 2018; Stefopoulou et al., 2018) as well as socio-economic indices (Liu-Helmersson et al., 2019; Lorenz et al., 2020b; Moreno-Madriñán et al., 2014; Nagao et al., 2003; Vannavong et al., 2017) were obtained from the IBGE (Instituto Brasileiro de Geografia e Estatística), IPP (Instituto Pereira Passos, Prefeitura do Rio de Janeiro), IPEA (Instituto de Pesquisa Econômica Aplicada), and the DataRioPortal.

## Appendix C - Entomological surveillance data

Two types of entomological surveillance data were applied for validation in our study (cf. Figure C. 1). Firstly, ovitrap data, comprising egg and larval counts for *Ae. aegypti*, were acquired for the training of the NB-GLM and the exploration of the first research question. This dataset was gathered by the municipal health ministry of Rio de Janeiro, covering 2 698 locations on a monthly basis, from January to December 2019. Secondly, House and Breteau indices for *Ae. aegypti* were extracted from LIRAA (Ministério da Saúde Brazil, 2013), to investigate the second research question. These indices, collected by the municipal health ministry of Rio de Janeiro, were based on 250 predefined city strata. LIRAA occurred periodically during February 3-9, 2019; May 5-11, 2019; August 4-10, 2019; and October 13-19, 2019 (Secretario Municipal de Saúde Rio de Janeiro, 2024a). The House index (HI) gauged the number of larva-infested houses relative to the total number of visited buildings during the survey, while the Breteau index (BI) represented the number of positive containers per 100 houses inspected. The municipal health ministry of Rio de Janeiro categorized the house index into three risk classes:  $HI < 0.9$  (minor),  $0.9 \leq HI \leq 3.9$  (significant), and  $HI > 3.9$  (severe).

The selection of ovitrap data for addressing the first research question and LIRAA data for the second research question was based on the spatial coverage of the entomological measurements. Ovitrap data allowed for an examination of urban suitability indicators within the vicinity of ovitrap locations, acknowledging that inferences on mosquito abundance for locations beyond the *Ae. aegypti* flight range may be invalid. LIRAA data provided a broader assessment to evaluate continuous immature *Ae. aegypti* suitability maps generated from ovitrap data and suitability indicators together. It is crucial to note that all entomological surveillance data was gathered manually, introducing potential biases due to human error, observer variability, and limitations in sampling frequency and coverage. These biases may lead to inaccuracies in estimating spatial suitability for *Ae. aegypti*, as well as errors in temporal trends, impacting the reliability of our analysis. Additionally, micro-scale factors, such as the positioning of an ovitrap in shaded or unshaded areas, can impact the observation values. The positioning of ovitraps was done in a systematic manner, more or less uniformly across the built-up areas of the municipality of Rio de Janeiro. This positioning resulted in an average distance between the two closest ovitraps of 330.38m.

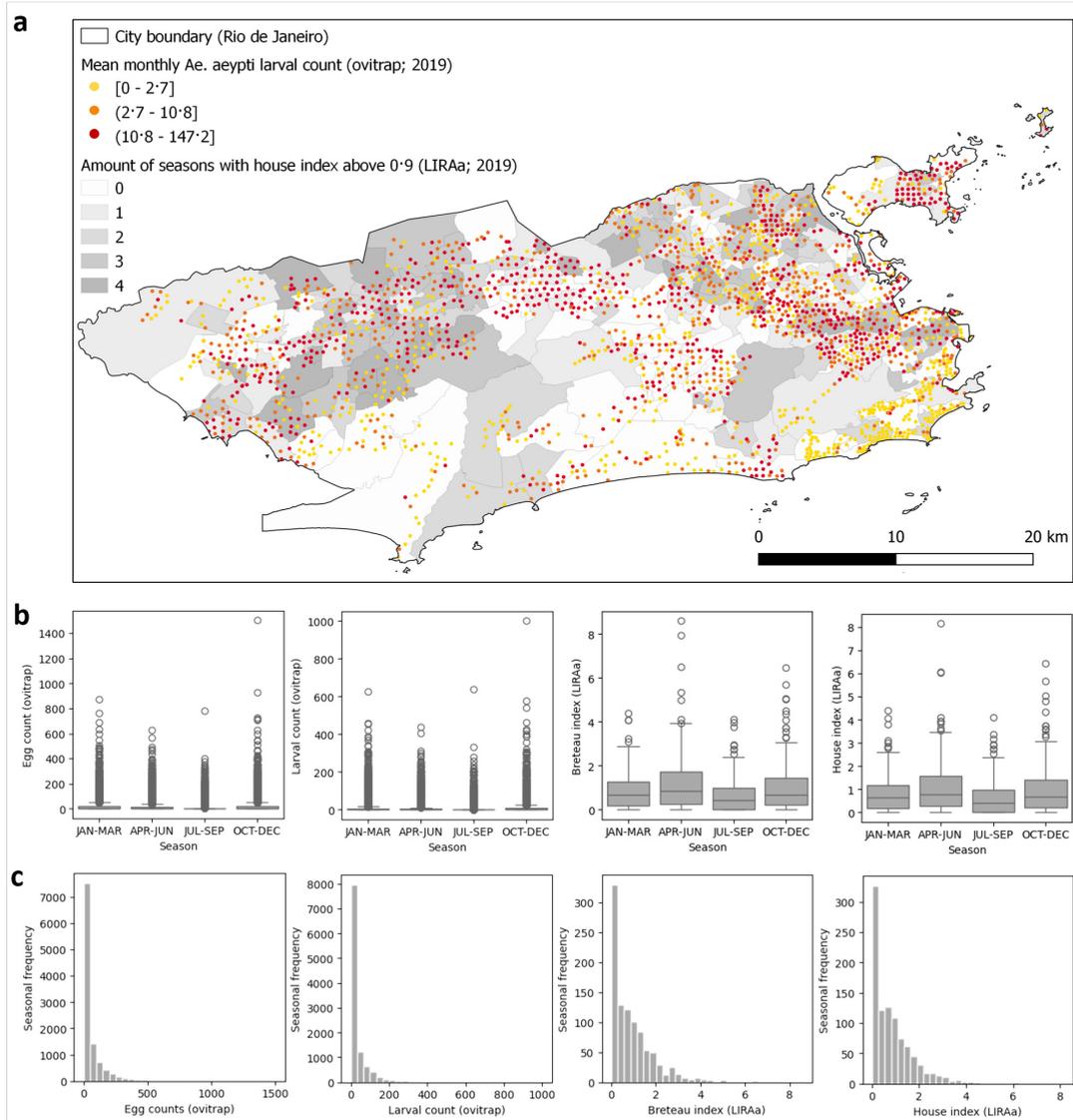
During LIRAA, the urban area was divided into 250 strata that represent homogeneous urban characteristics. According to the study design (Ministério da Saúde Brazil, 2013), one stratum consisted of 8 000 to 12 000 properties, of which 20% were inspected following a structured schema. Field agents assessed the number of eggs and larvae in all water containers present in each household surveyed. The most prominently affected container types, ranging from water tanks and ground-level deposits to furniture such as plates and vases, as well as the class of fixed deposits, tires, garbage, and natural plants such as bromeliads, were also recorded. For each container with mosquito larvae, the agents collected a sample that was sent to the Vector Laboratory of the Agency for the Control of Endemic Diseases for larval identification. Further details regarding the placement of ovitraps and the entomological surveillance in households are currently lacking in our knowledge, but could potentially be provided by the municipal

health ministry of Rio de Janeiro upon request.

The substantial spatial disparity observed between ovitrap-based suitability predictions and LIRAA entomological indices is likely attributed to variations in collection methodologies. The BI and HI indices, which serve as block-level indicators derived from a manual sampling process, contrast with ovitrap counts obtained from fixed measurement stations purposely designed as breeding sites. These stations, characterized by water retention and a dark color to attract mosquitoes, differ significantly from manual sampling methods. Whereas LIRAA indices depend on the active search for breeding sites by health agents, the ovitrap is ‘sought’ by the mosquito, making ovitrap counts a more sensitive indicator. Furthermore, the LIRAA indices typically register low values, given the temporary nature of positive recipients and their limited persistence over time. Consequently, a comprehensive and representative sample necessitates the inclusion of numerous properties, considering that the majority may lack positive recipients. The concentrated distribution of immature counts within LIRAA zones may have also skewed the comparative results. However, in operational terms, ovitraps could never replace LIRAA, as it provides additional indices like the “container type index”, monitoring infestation levels by container types not considered in this study.

In addition to the structural components mentioned, the reliability of LIRA indices relies on human components such as the dedication and expertise of field workers, encompassing their comprehension of vector biology and index calculation methods (Valle and Aguiar, 2023). In a previous study, Ribeiro et al. documented a high level of coincidence between HI and BI derived from LIRAA in the municipality of Rio de Janeiro (Ribeiro et al., 2021). However, from a biological perspective, the BI is expected to be greater than the house index because the female *Aedes* spreads the eggs in close locations. Our findings on the misalignment with ovitrap counts corroborate this assessment of the fragility of the indicators generated by LIRAA.

Another factor contributing to spatial divergence is that ovitrap-based surveillance overlooks indoor breeding sites. Additionally, ovitrap surveillance concentrates on egg and larval counts, whereas LIRAA encompasses infestation by *Ae. aegypti* pupae, which exhibit distinct lifespans and lower mortality rates. Comparable findings have been reported by Nascimento et al. (2020), who additionally observed that ovitraps provide a more rapid information due to heightened sensitivity compared to LIRAA in detecting *Ae. aegypti*. Getis et al. (2003), indicated spatial divergence between immature and adult *Ae. aegypti* populations. As the life cycle of immature *Ae. aegypti*, from emergence as 1st instar (L1) larvae to adulthood, is estimated to be around 8-10 days, varying with humidity and temperature conditions (Center for Disease Control and Prevention, 2024; Hossain et al., 2022), comparing entomological data from different surveillance techniques on a seasonal level may be too coarse (Cromwell et al., 2017; Morrison et al., 2004).



**Figure C. 1:** Graphics illustrating that in the tropical, *Aedes*-endemic municipality of Rio de Janeiro, spatial variance in *Ae. aegypti* abundance tends to exceed temporal variance at seasonal scale. The entomological surveillance data emphasizes high spatial variability in *Ae. aegypti* abundance at mosquito habitat scale, which is difficult to capture with the design of the LIRAA survey. This underscores the identified research gap for continuous suitability maps at flight range resolution considering urban landscape to advance vector control. a - Map of mean monthly *Ae. aegypti* larval count per ovitrap and amount of seasons with house index above 0.9 in 2019, indicating significant risk, as defined by LIRAA conducted for the municipality of Rio de Janeiro; b - Seasonal boxplots of entomological surveillance variables from ovitraps and LIRAA for the year 2019. The months spanning from December to April align with the wet season in the municipality of Rio de Janeiro, while the period from May to November corresponds to the dry season. The small size of the boxes in the ovitrap boxplots (left) and the presence of numerous outliers towards higher counts indicate a negative binomial distribution, with extreme values deviating significantly from the majority of the data points with zero egg and larval counts; c - Histograms showing the seasonal frequency of applied entomological datasets highlighting overdispersion for the year 2019.

## D - Evaluation of suitability indicators for immature *Ae. aegypti*

The negative-binomial generalized linear regression model (NB-GLM) with a log-link function was defined as follows:

$$\begin{aligned}
 Y_i &\sim NB(\hat{\mu}_i, \hat{\theta}) \\
 \mathbb{E}(Y_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta} \\
 \text{Var}(Y_i) &= \hat{\mu}_i * (1 - \hat{\theta}) / \hat{\theta}^2 \\
 \log(\hat{\mu}_i) &= \hat{\beta}_0 + \sum_{j=1}^{79} \hat{\beta}_j * PC_{ij}
 \end{aligned} \tag{1}$$

As our response variable  $Y_i$  we selected the seasonal mean eggs per trap (S-MET) and seasonal mean larva per trap (S-MLT) rates for the year 2019. As explanatory variables, we used all main effects of our self-generated urban indicators for immature *Ae. aegypti* suitability as listed in Table A. 1. All collected urban indicators were clipped to the geographical extent of the municipality of Rio de Janeiro. Using this stack of proxies, zonal statistics were run in different square buffers of flight range size around ovitrap locations to create five different feature vectors. Based on literature findings, we assumed that the flight range of *Ae. aegypti* is consistently lower than 1000 m for the municipality of Rio de Janeiro (Honório et al., 2003; Moore and Brown, 2022). Therefore, we defined flight ranges of 50, 100, 200, 400 and 1000 m as our flight range scenarios. The aerial coverage of ovitrap buffer regions in built-up areas showed notable variation: from 1.05% for a 50 m flight range buffer to 15.41% for a 200 m buffer, and up to 92.76% for a 1000 m buffer. The lower percentages for smaller buffers, indicative of assumptions regarding lower mosquito flight ranges, underscore the difficulty in capturing the high spatial variability of urban immature *Ae. aegypti* abundance with sample-based entomological surveillance. To remove collinearity within our feature vectors we ran a Principal component analysis (PCA). All resulting 79 principal components (PC) were utilized to run the negative-binomial GLMs for each feature vector. The combination of five different flight range buffers and the two response variables led to 10 different models, which were evaluated using Cohen's pseudo R-square (cf. Formula 2).

$$\begin{aligned}
 \text{Cohen's pseudo } R^2 &= 1 - \frac{\text{model deviance}}{\text{null model deviance}} \\
 \text{Negative - binomial model deviance} &= 2 \sum (y \cdot \log(\frac{y}{\mu}) - (y + k - 1) \log(\frac{y+k-1}{\mu+k-1}))
 \end{aligned} \tag{2}$$

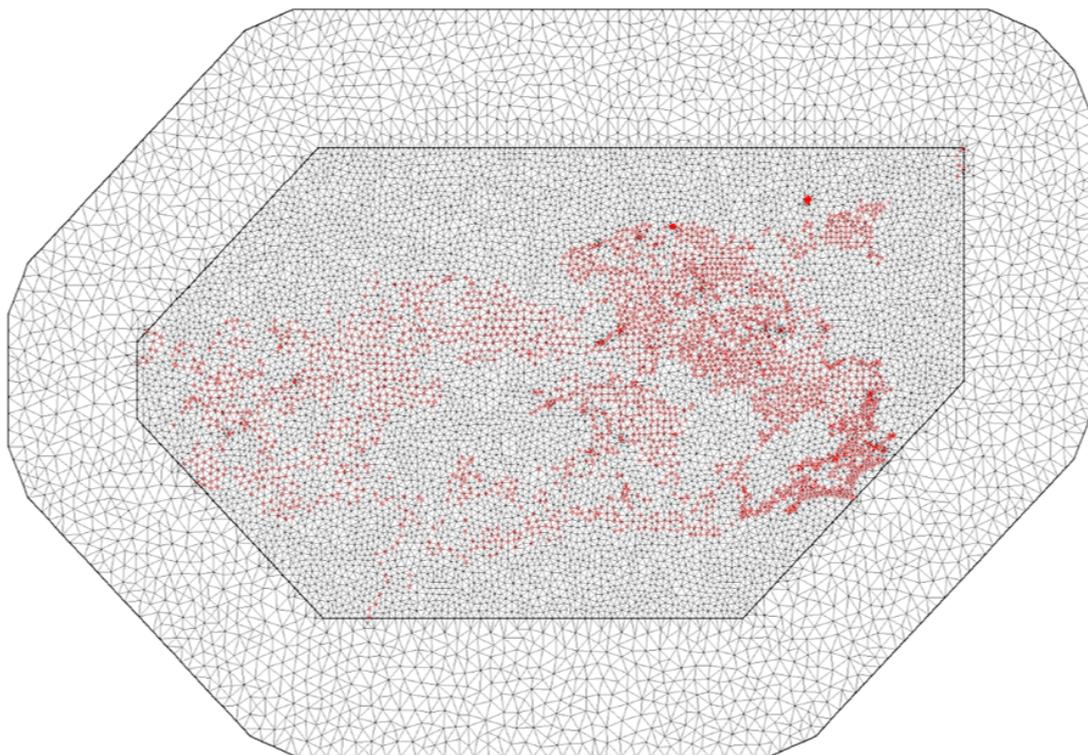
## Appendix E - Generation of urban suitability maps for immature *Ae. aegypti*

We modeled ovitrap egg and larval counts, denoted as the response variable  $Y_{it}$ , observed at distinct spatial locations  $i = 1, \dots, 2698$  and time periods  $t = 1, 2, 3, 4$  using a Bayesian spatiotemporal model with a negative-binomial probability distribution. Given the observed overdispersion in the applied entomological count data (cf. Figure C. 1), we employed a chi-squared test to evaluate the adequacy of the negative-binomial distribution as a likelihood function. The model assumes,

$$\begin{aligned}
 Y_{it} &\sim NB(\hat{\mu}_{it}, \hat{\theta}) \\
 \mathbb{E}(Y_{it}) &= \hat{\mu}_{it} * (1 - \hat{\theta}) / \hat{\theta} \\
 \text{Var}(Y_{it}) &= \hat{\mu}_{it} * (1 - \hat{\theta}) / \hat{\theta}^2 \\
 \log(\hat{\mu}_{it}) &= \hat{\beta}_0 + \sum_{j=1}^{79} \hat{\beta}_j * PC_{itj} + \xi(\mathbf{x}_i, t) \\
 \xi(\mathbf{x}_i, t) &= a\xi(\mathbf{x}_i, t - 1) + w(\mathbf{x}_i, t),
 \end{aligned} \tag{3}$$

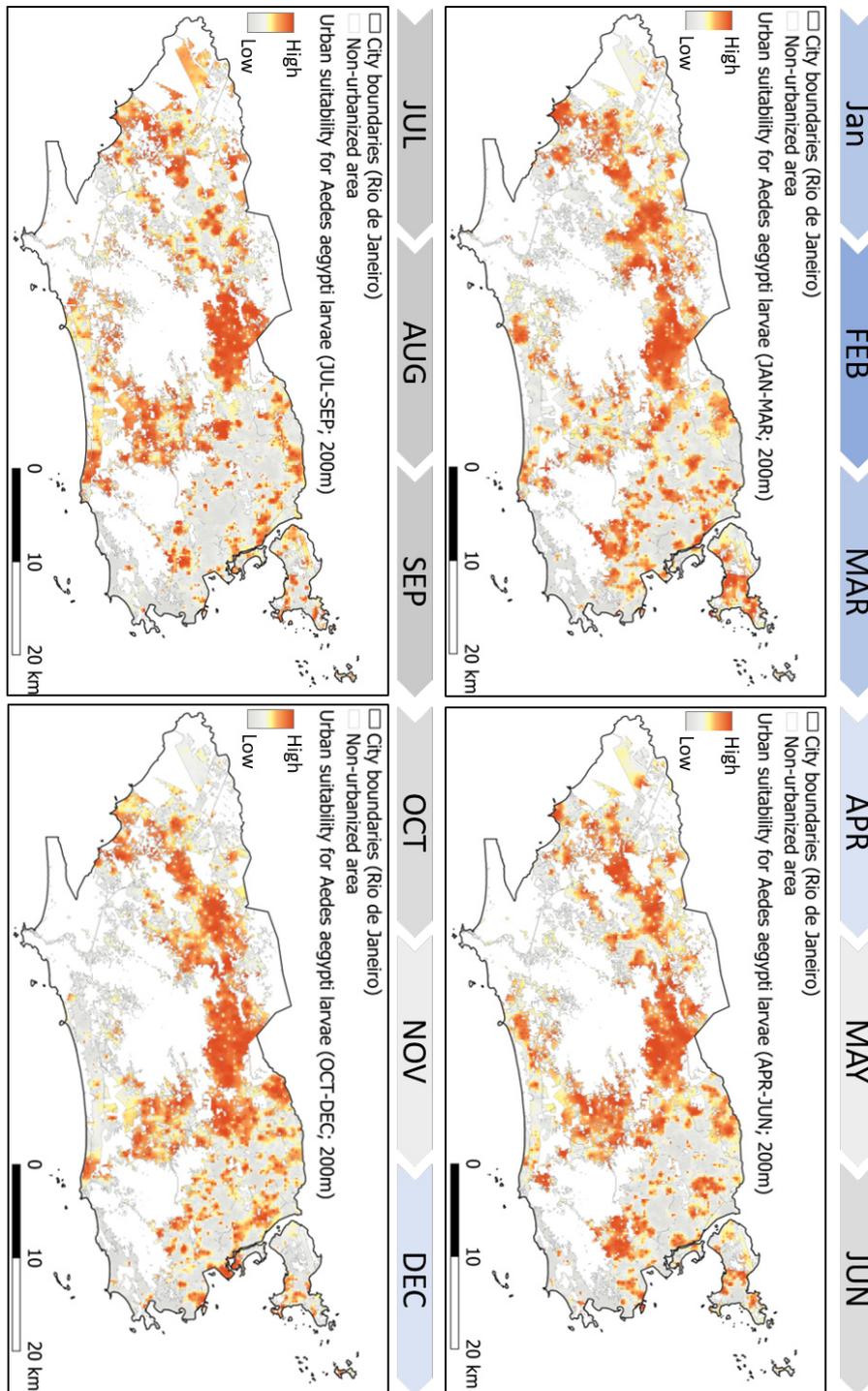
and consists of an intercept  $\hat{\beta}_0$ , principal components of spatiotemporal suitability indicators  $PC_{itj}$ , and independent and identically distributed spatiotemporal random effects  $\xi(\mathbf{x}_i, t)$  that change in time with first order autoregressive dynamics ( $|a| < 1$ ) (Blangiardo and Cameletti, 2015; Lindgren et al., 2011; Zuur et al., 2017). The model's incorporation of covariates is assumed to enhance its predictive capability and facilitates a more holistic understanding of the actors influencing immature *Ae. aegypti* suitability. The spatial model component was modeled by INLA using the Euclidean distances between ovitrap locations, a Matérn covariance function, and stochastic partial differential equations (SPDEs). Gaussian Markov random fields were built on triangle meshes considering boundary effects that could artificially inflate variance near the edges of the study area (cf. Figure E. 1). To map immature *Ae. aegypti* suitability in continuous space, the inverse distance weighting (IDW) algorithm was applied to interpolate point estimates from the mesh nodes to a uniformly distributed raster of 100 000 cells for each season of 2019, visualized using QGIS (QGIS Association, 2024).

The presented Bayesian modeling results are highly dependent on the selection of priors, particularly for building the Matérn covariance field. The most critical assumptions in our case study were made for the Penalized Complexity (PC) priors for the parameters' range and marginal standard deviation of the Matérn field, modeling, among other factors, the extent of mosquito movement in space in our case study. Besides that, passive dispersal of mosquito eggs and adults, driven by transport and trade (Bennett et al., 2019; Díaz-Nieto et al., 2016; Eritja et al., 2017; Guagliardo et al., 2015) was completely neglected in our study. Subsequent investigations may delve into methodologies for establishing these priors through more entomological surveillance and bio-ecological field studies. This would help eliminate potential bias in immature *Ae. aegypti* suitability maps and could also enable fine-tuning of the proposed framework for other mosquito species.



**Figure E. 1:** Mesh of Delaunay triangulations with 157 541 nodes created using R-INLA to represent the municipality of Rio de Janeiro and build stochastic partial differential equation (SPDE) for spatial modeling. The red nodes indicate the location of ovitraps.

## Appendix F - Seasonal suitability maps for *Ae. aegypti* larvae



**Figure F. 1:** Seasonal suitability maps for *Ae. aegypti* larvae in 200m resolution covering the urbanized area of Rio de Janeiro for 2019, using a complementary approach of entomological surveillance, urban landscape indicators for immature *Ae. aegypti* and bio-ecological knowledge on limited mosquito flight range. The months spanning from December to April align with the wet season in the municipality of Rio de Janeiro, while the period from May to November corresponds to the dry season.

# Bibliography

- Albrieu-Llinás, G., Espinosa, M. O., Quaglia, A., Abril, M., and Scavuzzo, C. M. (2018). “Urban environmental clustering to assess the spatial dynamics of *Aedes aegypti* breeding sites”. *Geospatial health* 13.1, p. 654. DOI: 10.4081/gh.2018.654.
- Alphabet Inc. (2024). *Google Earth Engine*.
- Arana-Guardia, R., Baak-Baak, C. M., Loroño-Pino, M. A., Machain-Williams, C., Beaty, B. J., Eisen, L., and García-Rejón, J. E. (2014). “Stormwater drains and catch basins as sources for production of *Aedes aegypti* and *Culex quinquefasciatus*”. *Acta tropica* 134, pp. 33–42. DOI: 10.1016/j.actatropica.2014.01.011.
- Araujo, R. V., Albertini, M. R., Costa-da-Silva, A. L., Suesdek, L., Franceschi, N. C. S., Bastos, N. M., Katz, G., Cardoso, V. A., Castro, B. C., Capurro, M. L., and Allegro, V. L. A. C. (2015). “São Paulo urban heat islands have a higher incidence of dengue than other urban areas”. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases* 19.2, pp. 146–155. DOI: 10.1016/j.bjid.2014.10.004.
- Azevedo, T. S. de, Bourke, B. P., Piovezan, R., and Sallum, M. A. M. (2018). “The influence of urban heat islands and socioeconomic factors on the spatial distribution of *Aedes aegypti* larval habitats”. *Geospatial health* 13.1, p. 623. DOI: 10.4081/gh.2018.623.
- Bailly, S., Machault, V., Beneteau, S., Palany, P., Girod, R., Lacaux, J.-P., Quenel, P., and Flamand, C. (2021). *Modeling spatiotemporal Aedes aegypti risk in French Guiana using meteorological and remote sensing data*. DOI: 10.1101/2021.08.02.21261373.
- Bar-Zeev, M. (1957). “The Effect of extreme Temperatures on different Stages of *Aedes aegypti* (L.)” *Bulletin of entomological research* 48.3, pp. 593–599. DOI: 10.1017/S0007485300002765.
- Barrera, R., Acevedo, V., and Amador, M. (2021). “Role of Abandoned and Vacant Houses on *Aedes aegypti* Productivity”. *The American journal of tropical medicine and hygiene* 104.1, pp. 145–150. DOI: 10.4269/ajtmh.20-0829.
- Barrera, R., Amador, M., and MacKay, A. J. (2011). “Population dynamics of *Aedes aegypti* and dengue as influenced by weather and human behavior in San Juan, Puerto Rico”. *PLoS neglected tropical diseases* 5.12, e1378. DOI: 10.1371/journal.pntd.0001378.
- Benitez, E. M., Ludueña-Almeida, F., Frías-Céspedes, M., Almirón, W. R., and Estallo, E. L. (2020). “Could land cover influence *Aedes aegypti* mosquito populations?” *Medical and veterinary entomology* 34.2, pp. 138–144. DOI: 10.1111/mve.12422.
- Bennett, K. L., Gómez Martínez, C., Almanza, A., Rovira, J. R., McMillan, W. O., Enriquez, V., Barraza, E., Diaz, M., Sanchez-Galan, J. E., Whiteman, A., Gittens, R. A., and Loaiza, J. R. (2019). “High infestation of invasive *Aedes* mosquitoes in used tires along the local transport network of Panama”. *Parasites & vectors* 12.1, p. 264. DOI: 10.1186/s13071-019-3522-8.

- Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley. DOI: 10.1002/9781118950203.
- Bonnet, E., Fournet, F., Benmarhnia, T., Ouedraogo, S., Dabiré, R., and Ridde, V. (2020). “Impact of a community-based intervention on *Aedes aegypti* and its spatial distribution in Ouagadougou, Burkina Faso”. *Infectious diseases of poverty* 9.1, p. 61. DOI: 10.1186/s40249-020-00675-6.
- Boser, A., Sousa, D., Larsen, A., and MacDonald, A. (2021). “Micro-climate to macro-risk: mapping fine scale differences in mosquito-borne disease risk using remote sensing”. *Environmental Research Letters* 16.12, p. 124014. DOI: 10.1088/1748-9326/ac3589.
- Britos Molinas, M. B., Melgarejo, E. G., and Arias, A. R. de (2022). “Phenotypic Variations of *Aedes aegypti* Populations and Egg Abundance According to Environmental Parameters in Two Dengue-Endemic Ecoregions in Paraguay”. *The American journal of tropical medicine and hygiene* 107.2, pp. 300–307. DOI: 10.4269/ajtmh.21-1184.
- Burke, R., Barrera, R., Lewis, M., Kluchinsky, T., and Claborn, D. (2010). “Septic tanks as larval habitats for the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus* in Playa-Playita, Puerto Rico”. *Medical and veterinary entomology* 24.2, pp. 117–123. DOI: 10.1111/j.1365-2915.2010.00864.x.
- Camara, D. C. P., Codeço, C. T., Juliano, S. A., Lounibos, L. P., Riback, T. I. S., Pereira, G. R., and Honorio, N. A. (2016). “Seasonal Differences in Density But Similar Competitive Impact of *Aedes albopictus* (Skuse) on *Aedes aegypti* (L.) in Rio de Janeiro, Brazil”. *PloS one* 11.6, e0157120. DOI: 10.1371/journal.pone.0157120.
- Campbell, L. P., Luther, C., Moo-Llanes, D., Ramsey, J. M., Danis-Lozano, R., and Peterson, A. T. (2015). “Climate change influences on global distributions of dengue and chikungunya virus vectors”. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370.1665. DOI: 10.1098/rstb.2014.0135.
- Caragata, E. P., Otero, L. M., Tikhe, C. V., Barrera, R., and Dimopoulos, G. (2022). “Microbial Diversity of Adult *Aedes aegypti* and Water Collected from Different Mosquito Aquatic Habitats in Puerto Rico”. *Microbial ecology* 83.1, pp. 182–201. DOI: 10.1007/s00248-021-01743-6.
- Cavalcanti, L. P. d. G., Oliveira, R. d. M. A. B., and Alencar, C. H. (2016). “Changes in infestation sites of female *Aedes aegypti* in Northeast Brazil”. *Revista da Sociedade Brasileira de Medicina Tropical* 49.4, pp. 498–501. DOI: 10.1590/0037-8682-0044-2016.
- Center for Disease Control and Prevention (2024). *Factsheet Aedes aegypti life cycle*.
- Centro de Operacoes Rio (2023). *Sistema Alerta Rio da Prefeitura do Rio de Janeiro*.
- Chan, K. L., Ho, B. C., and Chan, Y. C. (1971). “*Aedes aegypti* (L.) and *Aedes albopictus* (Skuse) in Singapore City: 2. Larval habitats\*”. *Bulletin of the World Health Organization* 44.5, pp. 629–633.
- Chang, L.-H., Hsu, E.-L., Teng, H.-J., and Ho, C.-M. (2007). “Differential Survival of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) Larvae Exposed to Low Temperatures in Taiwan”. *Journal of medical entomology* 44.2, pp. 205–210. DOI: 10.1093/jmedent/44.2.205.
- Chaves, L. F., Valerín Cordero, J. A., Delgado, G., Aguilar-Avenidaño, C., Maynes, E., Gutiérrez Alvarado, J. M., Ramírez Rojas, M., Romero, L. M., and Marín Rodríguez, R. (2021). “Modeling the association between *Aedes aegypti* ovitrap egg counts, multi-scale remotely sensed environmental data and arboviral cases at Puntarenas, Costa Rica (2017-2018)”. *Current research in parasitology & vector-borne diseases* 1, p. 100014. DOI: 10.1016/j.crpvbd.2021.100014.

- Chumsri, A., Jaroensutasinee, M., and Jaroensutasinee, K. (2020). “Spatial and temporal variations on the coexistence of *Aedes* and *Culex* larvae in Southern Thailand”. *Journal of Animal Behaviour and Biometeorology* 8.4, pp. 250–256. DOI: 10.31893/jabb.20033.
- Colón-González, F. J., Sewe, M. O., Tompkins, A. M., Sjödin, H., Casallas, A., Rocklöv, J., Caminade, C., and Lowe, R. (2021). “Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study”. *The Lancet. Planetary health* 5.7, e404–e414. DOI: 10.1016/S2542-5196(21)00132-7.
- Cornel, A. J., Holeman, J., Nieman, C. C., Lee, Y., Smith, C., Amorino, M., Brisco, K. K., Barrera, R., Lanzaro, G. C., and Mulligan Iii, F. S. (2016). “Surveillance, insecticide resistance and control of an invasive *Aedes aegypti* (Diptera: Culicidae) population in California”, p. 194. DOI: 10.12688/f1000research.8107.3.
- Costa, A. C., Codeço, C. T., Honório, N. A., Pereira, G. R., N Pinheiro, C. F., and Nobre, A. A. (2015). “Surveillance of dengue vectors using spatio-temporal Bayesian modeling”. *BMC medical informatics and decision making* 15, p. 93. DOI: 10.1186/s12911-015-0219-6.
- Costa, E. A. P. d. A., Santos, E. M. d. M., Correia, J. C., and Albuquerque, C. M. R. de (2010). “Impact of small variations in temperature and humidity on the reproductive activity and survival of *Aedes aegypti* (Diptera, Culicidae)”. *Revista Brasileira de Entomologia* 54.3, pp. 488–493. DOI: 10.1590/S0085-56262010000300021.
- Cromwell, E. A., Stoddard, S. T., Barker, C. M., van Rie, A., Messer, W. B., Meshnick, S. R., Morrison, A. C., and Scott, T. W. (2017). “The relationship between entomological indicators of *Aedes aegypti* abundance and dengue virus infection”. *PLoS neglected tropical diseases* 11.3, e0005429. DOI: 10.1371/journal.pntd.0005429.
- Data.Rio (2023a). *Hydrography lines*.
- Data.Rio (2023b). *Land use land cover*.
- Demuzere, M., Kittner, J., and Bechtel, B. (2021). “LCZ Generator: A Web Application to Create Local Climate Zone Maps”. *Frontiers in Environmental Science* 9. DOI: 10.3389/fenvs.2021.637455.
- Díaz-Nieto, L. M., Chiappero, M. B., Díaz de Astarloa, C., Maciá, A., Gardenal, C. N., and Berón, C. M. (2016). “Genetic Evidence of Expansion by Passive Transport of *Aedes* (*Stegomyia*) *aegypti* in Eastern Argentina”. *PLoS neglected tropical diseases* 10.9, e0004839. DOI: 10.1371/journal.pntd.0004839.
- Dom, N. C., Faiz Madzlan, M., Nadira Yusoff, S. N., Hassan Ahmad, A., Ismail, R., and Nazrina Camalxaman, S. (2016). “Profile distribution of juvenile *Aedes* species in an urban area of Malaysia”. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 110.4, pp. 237–245. DOI: 10.1093/trstmh/trw015.
- Ebi, K. L. and Nealon, J. (2016). “Dengue in a changing climate”. *Environmental research* 151, pp. 115–123. DOI: 10.1016/j.envres.2016.07.026.
- Egid, B. R., Coulibaly, M., Dadzie, S. K., Kamgang, B., McCall, P. J., Sedda, L., Toe, K. H., and Wilson, A. L. (2022). “Review of the ecology and behaviour of *Aedes aegypti* and *Aedes albopictus* in Western Africa and implications for vector control”. *Current research in parasitology & vector-borne diseases* 2, p. 100074. DOI: 10.1016/j.crpvbd.2021.100074.
- Equihua, M., Ibáñez-Bernal, S., Benítez, G., Estrada-Contreras, I., Sandoval-Ruiz, C. A., and Mendoza-Palmero, F. S. (2017). “Establishment of *Aedes aegypti* (L.) in mountainous regions in Mexico: Increasing number of population at risk of mosquito-borne disease and future climate conditions”. *Acta tropica* 166, pp. 316–327. DOI: 10.1016/j.actatropica.2016.11.014.

- Eritja, R., Palmer, J. R. B., Roiz, D., Sanpera-Calbet, I., and Bartumeus, F. (2017). “Direct Evidence of Adult *Aedes albopictus* Dispersal by Car”. *Scientific reports* 7.1, p. 14399. DOI: 10.1038/s41598-017-12652-5.
- Espinosa, M. O., Polop, F., Rotela, C. H., Abril, M., and Scavuzzo, C. M. (2016). “Spatial pattern evolution of *Aedes aegypti* breeding sites in an Argentinean city without a dengue vector control programme”. *Geospatial health* 11.3, p. 471. DOI: 10.4081/gh.2016.471.
- Estallo, E. L., Sangermano, F., Grech, M., Ludueña-Almeida, F., Frías-Cespedes, M., Ainete, M., Almirón, W., and Livdahl, T. (2018). “Modelling the distribution of the vector *Aedes aegypti* in a central Argentine city”. *Medical and veterinary entomology* 32.4, pp. 451–461. DOI: 10.1111/mve.12323.
- Estallo, E. L., Lamfri, M. A., Scavuzzo, C. M., Almeida, F. F. L., Introini, M. V., Zaidenberg, M., and Almirón, W. R. (2008). “Models for predicting *Aedes aegypti* larval indices based on satellite images and climatic variables”. *Journal of the American Mosquito Control Association* 24.3, pp. 368–376. DOI: 10.2987/5705.1.
- Estallo, E. L., Ludueña-Almeida, F. F., Visintin, A. M., Scavuzzo, C. M., Lamfri, M. A., Introini, M. V., Zaidenberg, M., and Almirón, W. R. (2012). “Effectiveness of normalized difference water index in modelling *Aedes aegypti* house index”. *International Journal of Remote Sensing* 33.13, pp. 4254–4265. DOI: 10.1080/01431161.2011.640962.
- Flaibani, N., Pérez, A. A., Barbero, I. M., and Burrioni, N. E. (2020). “Different approaches to characterize artificial breeding sites of *Aedes aegypti* using generalized linear mixed models”. *Infectious diseases of poverty* 9.1, p. 107. DOI: 10.1186/s40249-020-00705-3.
- Fleischmann, M. (2019). “momepy: Urban Morphology Measuring Toolkit”. *Journal of Open Source Software* 4.43, p. 1807. DOI: 10.21105/joss.01807.
- Flores, H. A. and O’Neill, S. L. (2018). “Controlling vector-borne diseases by releasing modified mosquitoes”. *Nature reviews. Microbiology* 16.8, pp. 508–518. DOI: 10.1038/s41579-018-0025-0.
- Franco dos Santos, S. A., Karam, H. A., Filho, A. J. P., Da Silva, J. C. B., Rojas, J. L. F., Suazo, J. M. A., Panduro, I. L. V., and Peña, C. A. S. (2022). “Dengue Climate Variability in Rio de Janeiro City with Cross-Wavelet Transform”. *Journal of Environmental Protection* 13.03, pp. 261–276. DOI: 10.4236/jep.2022.133016.
- German, A., Espinosa, M. O., Abril, M., and Scavuzzo, C. M. (2018). “Exploring satellite based temporal forecast modelling of *Aedes aegypti* oviposition from an operational perspective”. *Remote Sensing Applications: Society and Environment* 11, pp. 231–240. DOI: 10.1016/j.rsase.2018.07.006.
- Getachew, D., Tekie, H., Gebre-Michael, T., Balkew, M., and Mesfin, A. (2015). “Breeding Sites of *Aedes aegypti*: Potential Dengue Vectors in Dire Dawa, East Ethiopia”. *Interdisciplinary perspectives on infectious diseases* 2015, p. 706276. DOI: 10.1155/2015/706276.
- Getis, A., Morrison, A. C., Gray, K., and Scott, T. W. (2003). “Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru”. *The American journal of tropical medicine and hygiene* 69.5, pp. 494–505.
- Gibson, G., Souza-Santos, R., Pedro, A. S., Honório, N. A., and Sá Carvalho, M. (2014). “Occurrence of severe dengue in Rio de Janeiro: an ecological study”. *Revista da Sociedade Brasileira de Medicina Tropical* 47.6, pp. 684–691. DOI: 10.1590/0037-8682-0223-2014.
- Gomes, H., Jesus, A. G. de, and Quaresma, J. A. S. (2023). “Identification of risk areas for arboviruses transmitted by *Aedes aegypti* in northern Brazil: A One Health

- analysis”. *One health (Amsterdam, Netherlands)* 16, p. 100499. DOI: 10.1016/j.onehlt.2023.100499.
- Guagliardo, S. A., Morrison, A. C., Barboza, J. L., Requena, E., Astete, H., Vazquez-Prokopec, G., and Kitron, U. (2015). “River boats contribute to the regional spread of the dengue vector *Aedes aegypti* in the Peruvian Amazon”. *PLoS neglected tropical diseases* 9.4, e0003648. DOI: 10.1371/journal.pntd.0003648.
- Hammond, S. N., Gordon, A. L., Del Lugo, E. C., Moreno, G., Kuan, G. M., López, M. M., López, J. D., Delgado, M. A., Valle, S. I., Espinoza, P. M., and Harris, E. (2007). “Characterization of *Aedes aegypti* (Diptera: Culicidae) Production Sites in Urban Nicaragua”. *Journal of medical entomology* 44.5, pp. 851–860. DOI: 10.1093/jmedent/44.5.851.
- Harrington, L. C., Scott T.W., Lerdthusnee, K., Coleman, R. C., Costero, A., Clark, G. G., Jones, J. J., Kitthawee, S., Kittayapong, P., Sithiprasasna, R., and Edman, J. D. (2005). “Dispersal of the dengue vector *Aedes aegypti* within and between rural communities”. *The American journal of tropical medicine and hygiene* 72.2, pp. 209–220. DOI: 10.4269/ajtmh.2005.72.209.
- Hilbe, J. M. (2012). *Negative Binomial Regression*. Cambridge University Press. DOI: 10.1017/CB09780511973420.
- Honório, N. A., Castro, M. G., Barros, F. S. M. de, Magalhães, M. d. A. F. M., and Sabroza, P. C. (2009). “The spatial distribution of *Aedes aegypti* and *Aedes albopictus* in a transition zone, Rio de Janeiro, Brazil”. *Cadernos de saude publica* 25.6, pp. 1203–1214. DOI: 10.1590/s0102-311x2009000600003.
- Honório, N. A., Da Silva, W. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003). “Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil”. *Memorias do Instituto Oswaldo Cruz* 98.2, pp. 191–198. DOI: 10.1590/s0074-02762003000200005.
- Hossain, M. S., Raihan, M. E., Hossain, M. S., Syeed, M. M. M., Rashid, H., and Reza, M. S. (2022). “*Aedes* Larva Detection Using Ensemble Learning to Prevent Dengue Endemic”. *BioMedInformatics* 2.3, pp. 405–423. DOI: 10.3390/biomedinformatics2030026.
- Hou, Y. and Biljecki, F. (2022). “A comprehensive framework for evaluating the quality of street view imagery”. *International Journal of Applied Earth Observation and Geoinformation* 115, p. 103094. DOI: 10.1016/j.jag.2022.103094.
- Humanitarian Data Exchange (2019). *Brazil: High Resolution Population Density Maps + Demographic Estimates: BRA\_southeast*.
- Humanitarian OpenStreetMap Team (2024). *HOT Export Tool*.
- Jesús Crespo, R. de and Rogers, R. E. (2021). “Habitat Segregation Patterns of Container Breeding Mosquitos: The Role of Urban Heat Islands, Vegetation Cover, and Income Disparity in Cemeteries of New Orleans”. *International journal of environmental research and public health* 19.1. DOI: 10.3390/ijerph19010245.
- Kache, P. A., Santos-Vega, M., Stewart-Ibarra, A. M., Cook, E. M., Seto, K. C., and Diuk-Wasser, M. A. (2022). “Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases”. *Nature ecology & evolution* 6.11, pp. 1601–1616. DOI: 10.1038/s41559-022-01876-y.
- Kamgang, B., Happi, J. Y., Boisier, P., Njiokou, F., Hervé, J.-P., Simard, F., and Paupy, C. (2010). “Geographic and ecological distribution of the dengue and chikungunya virus vectors *Aedes aegypti* and *Aedes albopictus* in three major Cameroonian towns”. *Medical and veterinary entomology* 24.2, pp. 132–141. DOI: 10.1111/j.1365-2915.2010.00869.x.

- Kaplan, L., Kendell, D., Robertson, D., Livdahl, T., and Khatchikian, C. (2010). “Aedes aegypti and Aedes albopictus in Bermuda: extinction, invasion, invasion and extinction”. *Biological invasions* 12.9, pp. 3277–3288. DOI: 10.1007/s10530-010-9721-z.
- Knoblauch, S. and Moritz, M. (2023). *Geospatial Innovation’s Potential for Addressing Mosquito-Borne Diseases*. Ed. by United Nations World Data Forum.
- Knoblauch, S., Su Yin, M., Chatrinan, K., Rocha, A. A. d. A., Haddawy, P., Biljecki, F., Lautenbach, S., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2024). “High-resolution mapping of urban Aedes aegypti immature abundance through breeding site detection based on satellite and street view imagery (under review)”. *Scientific Reports: Collection on Urbanization and communicable diseases*. DOI: 10.1038/s41598-024-67914-w.
- Knoblauch, S., Li, H., Lautenbach, S., Elshiaty, Y., Rocha, A. A. d. A., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2023). “Semi-supervised water tank detection to support vector control of emerging infectious diseases transmitted by Aedes Aegypti”. *International Journal of Applied Earth Observation and Geoinformation* 119, p. 103304. DOI: 10.1016/j.jag.2023.103304.
- Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., and Scott, T. W. (2011). “Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti”. *Proceedings of the National Academy of Sciences of the United States of America* 108.18, pp. 7460–7465. DOI: 10.1073/pnas.1101377108.
- Landau, K. I. and van Leeuwen, W. J. D. (2012). “Fine scale spatial urban land cover factors associated with adult mosquito abundance and risk in Tucson, Arizona”. *Journal of vector ecology : journal of the Society for Vector Ecology* 37.2, pp. 407–418. DOI: 10.1111/j.1948-7134.2012.00245.x.
- Lega, J., Brown, H. E., and Barrera, R. (2017). “Aedes aegypti (Diptera: Culicidae) Abundance Model Improved With Relative Humidity and Precipitation-Driven Egg Hatching”. *Journal of medical entomology* 54.5, pp. 1375–1384. DOI: 10.1093/jme/tjx077.
- Li, C. F., Lim, T. W., Han, L. L., and Fang, R. (1985). “Rainfall, abundance of Aedes aegypti and dengue infection in Selangor, Malaysia”. *The Southeast Asian journal of tropical medicine and public health* 16.4, pp. 560–568.
- Liew, C. and Curtis, C. F. (2004). “Horizontal and vertical dispersal of dengue vector mosquitoes, Aedes aegypti and Aedes albopictus, in Singapore”. *Medical and veterinary entomology* 18.4, pp. 351–360. DOI: 10.1111/j.0269-283X.2004.00517.x.
- Limkittikul, K., Brett, J., and L’Azou, M. (2014). “Epidemiological trends of dengue disease in Thailand (2000-2011): a systematic literature review”. *PLoS neglected tropical diseases* 8.11, e3241. DOI: 10.1371/journal.pntd.0003241.
- Lindgren, F., Rue, H., and Lindström, J. (2011). “An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73.4, pp. 423–498. DOI: 10.1111/j.1467-9868.2011.00777.x.
- Lindsay, S. W., Wilson, A., Golding, N., Scott, T. W., and Takken, W. (2017). “Improving the built environment in urban areas to control Aedes aegypti-borne diseases”. *Bulletin of the World Health Organization* 95.8, pp. 607–608. DOI: 10.2471/BLT.16.189688.
- Liu, T., Zhu, G., He, J., Song, T., Zhang, M., Lin, H., Xiao, J., Zeng, W., Li, X., Li, Z., Xie, R., Zhong, H., Wu, X., Hu, W., Zhang, Y., and Ma, W. (2017). “Early rigorous control interventions can largely reduce dengue outbreak magnitude: experience from Chaozhou, China”. *BMC public health* 18.1, p. 90. DOI: 10.1186/s12889-017-4616-x.

- Liu-Helmersson, J., Brännström, Å., Sewe, M. O., Semenza, J. C., and Rocklöv, J. (2019). “Estimating Past, Present, and Future Trends in the Global Distribution and Abundance of the Arbovirus Vector *Aedes aegypti* Under Climate Change Scenarios”. *Frontiers in public health* 7, p. 148. DOI: 10.3389/fpubh.2019.00148.
- Lorenz, C., Castro, M. C., Trindade, P. M. P., Nogueira, M. L., Oliveira Lage, M. de, Quintanilha, J. A., Parra, M. C., Dibo, M. R., Fávoro, E. A., Guirado, M. M., and Chiaravalloti-Neto, F. (2020a). “Predicting *Aedes aegypti* infestation using landscape and thermal features”. *Scientific reports* 10.1, p. 21688. DOI: 10.1038/s41598-020-78755-8.
- Lorenz, C., Chiaravalloti-Neto, F., Oliveira Lage, M. de, Quintanilha, J. A., Parra, M. C., Dibo, M. R., Fávoro, E. A., Guirado, M. M., and Nogueira, M. L. (2020b). “Remote sensing for risk mapping of *Aedes aegypti* infestations: Is this a practical task?” *Acta tropica* 205, p. 105398. DOI: 10.1016/j.actatropica.2020.105398.
- Louis, V. R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y., and Dambach, P. (2014). “Modeling tools for dengue risk mapping - a systematic review”. *International journal of health geographics* 13, p. 50. DOI: 10.1186/1476-072X-13-50.
- Lozano-Fuentes, S., Hayden, M. H., Welsh-Rodriguez, C., Ochoa-Martinez, C., Tapias-Santos, B., Kobylinski, K. C., Uejio, C. K., Zielinski-Gutierrez, E., Monache, L. D., Monaghan, A. J., Steinhoff, D. F., and Eisen, L. (2012). “The dengue virus mosquito vector *Aedes aegypti* at high elevation in Mexico”. *The American journal of tropical medicine and hygiene* 87.5, pp. 902–909. DOI: 10.4269/ajtmh.2012.12-0244.
- Lucena, A. J. de, Faria Peres, L. de, Correa Rotunno Filho, O., and Almeida Franca, J. R. de (2015). “Estimation of the urban heat island in the Metropolitan Area of Rio de Janeiro - Brazil”. *2015 Joint Urban Remote Sensing Event (JURSE)*. IEEE, pp. 1–4. DOI: 10.1109/JURSE.2015.7120540.
- Manrique-Saide, P., Davies, C. R., Coleman, P. G., Rebollar-Tellez, E., Che-Medoza, A., Dzul-Manzanilla, F., and Zapata-Peniche, A. (2008). “Pupal surveys for *Aedes aegypti* surveillance and potential targeted control in residential areas of Mérida, México”. *Journal of the American Mosquito Control Association* 24.2, pp. 289–298. DOI: 10.2987/5578.1.
- Mapillary (2023). *Mapillary API Documentation*.
- Maquart, P.-O., Froehlich, Y., and Boyer, S. (2022). “Plastic pollution and infectious diseases”. *The Lancet. Planetary health* 6.10, e842–e845. DOI: 10.1016/S2542-5196(22)00198-X.
- Martín, M. E., Alonso, A. C., Faraone, J., Stein, M., and Estallo, E. L. (2022a). *Aedes aegypti* and *Aedes albopictus* abundance, landscape coverage and spectral indices effects in a subtropical city of Argentina. DOI: 10.1101/2022.01.11.475665.
- Martín, M. E., Alonso, A. C., Faraone, J., Stein, M., and Estallo, E. L. (2022b). “Satellite observation to assess dengue risk due to *Aedes aegypti* and *Aedes albopictus* in a subtropical city of Argentina”. *Medical and veterinary entomology*. DOI: 10.1111/mve.12604.
- Martini, M., Triasputri, Y., Hestningsih, R., Yuliawati, S., and Purwantisasi, S. (2019). “Longevity and development of *Aedes aegypti* larvae to imago in domestic sewage water”. *Journal of the Medical Sciences (Berkala Ilmu Kedokteran)* 51.04. DOI: 10.19106/JMedSci005104201906.
- Medronho, R. A., Macrini, L., Novellino, D. M., Lagrotta, M. T. F., Câmara, V. M., and Pedreira, C. E. (2009). “*Aedes aegypti* immature forms distribution according to type of breeding site”. *The American journal of tropical medicine and hygiene* 80.3, pp. 401–404. DOI: 10.4269/ajtmh.2009.80.401.

- Menchaca-Armenta, I., Ocampo-Torres, M., Hernández-Gómez, A., and Zamora-Cerritos, K. (2018). “Risk perception and level of knowledge of diseases transmitted by *Aedes aegypti*”. *Revista do Instituto de Medicina Tropical de Sao Paulo* 60, e10. DOI: 10.1590/S1678-9946201860010.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. (2019). “The current and future global distribution and population at risk of dengue”. *Nature microbiology* 4.9, pp. 1508–1515. DOI: 10.1038/s41564-019-0476-8.
- Ministério da Saúde Brazil (2013). *Levantamento Rápido de Índices para Aedes aegypti*.
- Miranda, V. F. V. de, Faria Peres, L. de, José de Lucena, A., França, J. R. d. A., and Libonati, R. (2022). “Urbanization-induced impacts on heat-energy fluxes in tropical South America from 1984 to 2020: The Metropolitan Area of Rio de Janeiro/Brazil”. *Building and Environment* 216, p. 109008. DOI: 10.1016/j.buildenv.2022.109008.
- Misslin, R., Vaguet, Y., Vaguet, A., and Daudé, É. (2018). “Estimating air temperature using MODIS surface temperature images for assessing *Aedes aegypti* thermal niche in Bangkok, Thailand”. *Environmental monitoring and assessment* 190.9, p. 537. DOI: 10.1007/s10661-018-6875-0.
- Montagner, F. R. G., Silva, O. S., and Jahnke, S. M. (2018). “Mosquito species occurrence in association with landscape composition in green urban areas”. *Brazilian journal of biology = Revista brasleira de biologia* 78.2, pp. 233–239. DOI: 10.1590/1519-6984.04416.
- Moore, T. C. and Brown, H. E. (2022). “Estimating *Aedes aegypti* (Diptera: Culicidae) Flight Distance: Meta-Data Analysis”. *Journal of medical entomology* 59.4, pp. 1164–1170. DOI: 10.1093/jme/tjac070.
- Moraes Figueiredo, L. T. (2004). *Dengue in Brazil: Past, Present and Future Perspective - Dengue bulletin*. Ed. by WHO Regional Office for South-East Asia.
- Moreno-Madriñán, M., Crosson, W., Eisen, L., Estes, S., Estes Jr., M., Hayden, M., Hemmings, S., Irwin, D., Lozano-Fuentes, S., Monaghan, A., Quattrochi, D., Welsh-Rodriguez, C., and Zielinski-Gutierrez, E. (2014). “Correlating Remote Sensing Data with the Abundance of Pupae of the Dengue Virus Mosquito Vector, *Aedes aegypti*, in Central Mexico”. *ISPRS International Journal of Geo-Information* 3.2, pp. 732–749. DOI: 10.3390/ijgi3020732.
- Morrison, A. C., Astete, H., Chapilliquen, F., Ramirez-Prada, C., Diaz, G., Getis, A., Gray, K., and Scott, T. W. (2004). “Evaluation of a sampling methodology for rapid assessment of *Aedes aegypti* infestation levels in Iquitos, Peru”. *Journal of medical entomology* 41.3, pp. 502–510. DOI: 10.1603/0022-2585-41.3.502.
- Mukhtar, M. U., Han, Q., Liao, C., Haq, F., Arslan, A., and Bhatti, A. (2018). “Seasonal Distribution and Container Preference Ratio of the Dengue Fever Vector (*Aedes aegypti*, Diptera: Culicidae) in Rawalpindi, Pakistan”. *Journal of medical entomology* 55.4, pp. 1011–1015. DOI: 10.1093/jme/tjy010.
- Nagao, Y., Thavara, U., Chitnumsup, P., Tawatsin, A., Chansang, C., and Campbell-Lendrum, D. (2003). “Climatic and social risk factors for *Aedes* infestation in rural Thailand”. *Tropical medicine & international health : TM & IH* 8.7, pp. 650–659. DOI: 10.1046/j.1365-3156.2003.01075.x.
- Nascimento, K. L. C., Da Silva, J. F. M., Zequi, J. A. C., and Lopes, J. (2020). “Comparison Between Larval Survey Index and Positive Ovitrap Index in the Evaluation of Populations of *Aedes (Stegomyia) aegypti* (Linnaeus, 1762) North of Paraná,

- Brazil”. *Environmental health insights* 14, p. 1178630219886570. DOI: 10.1177/1178630219886570.
- Nasir, S., Jabeen, F., Abbas, S., Nasir, I., and Debboun, M. (2017). “Effect of Climatic Conditions and Water Bodies on Population Dynamics of the Dengue Vector, *Aedes aegypti* (Diptera: Culicidae)”. *Journal of Arthropod-Borne Diseases* 11.1, pp. 50–59.
- Novaes, C., Silva Pinto, F., and Marques, R. C. (2022). “*Aedes Aegypti*-Insights on the Impact of Water Services”. *GeoHealth* 6.11, e2022GH000653. DOI: 10.1029/2022GH000653.
- Oliveira Lemos, L. de, Oscar Júnior, A. C., and Assis Mendonça, F. de (2021). “Urban climate maps as a public health tool for urban planning: The case of dengue fever in Rio De Janeiro/Brazil”. *Urban Climate* 35, p. 100749. DOI: 10.1016/j.uclim.2020.100749.
- Pancetti, F. G. M., Honório, N. A., Urbinatti, P. R., and Lima-Camara, T. N. (2015). “Twenty-eight years of *Aedes albopictus* in Brazil: a rationale to maintain active entomological and epidemiological surveillance”. *Revista da Sociedade Brasileira de Medicina Tropical* 48.1, pp. 87–89. DOI: 10.1590/0037-8682-0155-2014.
- Paploski, I. A. D., Rodrigues, M. S., Mugabe, V. A., Kikuti, M., Tavares, A. S., Reis, M. G., Kitron, U., and Ribeiro, G. S. (2016). “Storm drains as larval development and adult resting sites for *Aedes aegypti* and *Aedes albopictus* in Salvador, Brazil”. *Parasites & vectors* 9.1, p. 419. DOI: 10.1186/s13071-016-1705-0.
- Parra, M. C. P., Lorenz, C., Dibo, M. R., Aguiar Milhim, B. H. G. de, Guirado, M. M., Nogueira, M. L., and Chiaravalloti-Neto, F. (2022). “Association between densities of adult and immature stages of *Aedes aegypti* mosquitoes in space and time: implications for vector surveillance”. *Parasites & vectors* 15.1, p. 133. DOI: 10.1186/s13071-022-05244-4.
- Peres, L. d. F., Lucena, A. J. de, Rotunno Filho, O. C., and França, J. R. d. A. (2018). “The urban heat island in Rio de Janeiro, Brazil, in the last 30 years using remote sensing data”. *International Journal of Applied Earth Observation and Geoinformation* 64, pp. 104–116. DOI: 10.1016/j.jag.2017.08.012.
- Portella Ornelas de Melo, Diogo, Scherrer, L. R., and Eiras, Á. E. (2012). “Dengue fever occurrence and vector detection by larval survey, ovitrap and MosquiTRAP: a space-time clusters analysis”. *PloS one* 7.7, e42125. DOI: 10.1371/journal.pone.0042125.
- PPGAU UFF (2023a). *Portal de dados urbanos: Building footprints*.
- PPGAU UFF (2023b). *Portal de dados urbanos: High resolution elevation data set for the city of Rio de Janeiro*.
- Prefeitura da Cidade do Rio de Janeiro (2023a). *Hidrografia*. Ed. by Data.Rio.
- Prefeitura da Cidade do Rio de Janeiro (2023b). *Land use*. Ed. by Data.Rio.
- Prefeitura da Cidade do Rio de Janeiro (2023c). *Vegetation Cover*. Ed. by Data.Rio.
- Prosdocimi, D. and Klima, K. (2020). “Health effects of heat vulnerability in Rio de Janeiro: a validation model for policy applications”. *SN Applied Sciences* 2.12. DOI: 10.1007/s42452-020-03750-7.
- QGIS Association (2024). *QGIS Geographic Information System*.
- Regilme, M. A. F., Carvajal, T. M., Honnen, A.-C., Amalin, D. M., and Watanabe, K. (2021). “The influence of roads on the fine-scale population genetic structure of the dengue vector *Aedes aegypti* (Linnaeus)”. *PLoS neglected tropical diseases* 15.2, e0009139. DOI: 10.1371/journal.pntd.0009139.
- Reiskind, M. H. and Lounibos, L. P. (2009). “Effects of intraspecific larval competition on adult longevity in the mosquitoes *Aedes aegypti* and *Aedes albopictus*”. *Medical*

- and veterinary entomology* 23.1, pp. 62–68. DOI: 10.1111/j.1365-2915.2008.00782.x.
- Reiter, P. (2007). “Oviposition, dispersal, and survival in *Aedes aegypti*: implications for the efficacy of control strategies”. *Vector borne and zoonotic diseases (Larchmont, N.Y.)* 7.2, pp. 261–273. DOI: 10.1089/vbz.2006.0630.
- Ribeiro, M. S., Ferreira, D. F., Azevedo, R. C., Santos, G. B. G. D., and Medronho, R. d. A. (2021). “Índices larvais de *Aedes aegypti* e incidência de dengue: um estudo ecológico no Estado do Rio de Janeiro, Brasil”. *Cadernos de saude publica* 37.7, e00263320. DOI: 10.1590/0102-311X00263320.
- Ricas Rezende, H., Malta Romano, C., Morales Claro, I., Santos Caleiro, G., Cerdeira Sabino, E., Felix, A. C., Bissoli, J., Hill, S., Rodrigues Faria, N., Da Cardoso Silva, T. C., Brioschi Santos, A. P., Cerutti Junior, C., and Vicente, C. R. (2020). “First report of *Aedes albopictus* infected by Dengue and Zika virus in a rural outbreak in Brazil”. *PLoS one* 15.3. DOI: 10.1371/journal.pone.0229847.
- Roslan, M. A., Ngui, R., Vythilingam, I., Chan, K. F., Ong, P. S., Low, C. K., Muhammed, N. H., and Wan Sulaiman, W. Y. (2022). “Surveillance of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in high-rise apartment buildings in Selangor, Malaysia”. *International Journal of Tropical Insect Science* 42.2, pp. 1959–1969. DOI: 10.1007/s42690-021-00725-y.
- Roslan, M. A., Shafie, A., Ngui, R., Lim, Y. A. L., and Sulaiman, W. Y. W. (2013). “Vertical infestation of the dengue vectors *Aedes aegypti* and *Aedes albopictus* in apartments in Kuala Lumpur, Malaysia”. *Journal of the American Mosquito Control Association* 29.4, pp. 328–336. DOI: 10.2987/13-6363.1.
- Rowley, W. A. and Graham, C. L. (1968). “The effect of temperature and relative humidity on the flight performance of female *Aedes aegypti*”. *Journal of insect physiology* 14.9, pp. 1251–1257. DOI: 10.1016/0022-1910(68)90018-8.
- Rue, H. and Lindgren, F. (2024). *R-INLA package*.
- Russell, B. M., Kay, B. H., and Shipton, W. (2001). “Survival of *Aedes aegypti* (Diptera: Culicidae) eggs in surface and subterranean breeding sites during the northern Queensland dry season”. *Journal of medical entomology* 38.3, pp. 441–445. DOI: 10.1603/0022-2585-38.3.441.
- Russell, B. M., McBride, W. J. J., Mullner, H., and Kay, B. H. (2002). “Epidemiological significance of subterranean *Aedes aegypti* (Diptera: Culicidae) breeding sites to dengue virus infection in Charters Towers, 1993”. *Journal of medical entomology* 39.1, pp. 143–145. DOI: 10.1603/0022-2585-39.1.143.
- Sallam, M. F., Fizer, C., Pilant, A. N., and Whung, P.-Y. (2017). “Systematic Review: Land Cover, Meteorological, and Socioeconomic Determinants of *Aedes* Mosquito Habitat for Risk Mapping”. *International journal of environmental research and public health* 14.10. DOI: 10.3390/ijerph14101230.
- Secretario Municipal de Saúde Rio de Janeiro (2024a). *Dengue: dados epidemiológicos*.
- Secretario Municipal de Saúde Rio de Janeiro (2024b). *Levantamento de Índice Rápido para *Aedes aegypti* (LIRAa)*.
- Semenza, J. C., Rocklöv, J., and Ebi, K. L. (2022). “Climate Change and Cascading Risks from Infectious Disease”. *Infectious diseases and therapy* 11.4, pp. 1371–1390. DOI: 10.1007/s40121-022-00647-3.
- Souza, S. S. de, Da Silva, I. G., and Da Silva, H. H. G. (2010). “Associação entre incidência de dengue, pluviosidade e densidade larvária de *Aedes aegypti*, no Estado de Goiás”. *Revista da Sociedade Brasileira de Medicina Tropical* 43.2, pp. 152–155. DOI: 10.1590/S0037-86822010000200009.
- Spiegel, J. M., Bonet, M., Ibarra, A.-M., Pagliccia, N., Ouellette, V., and Yassi, A. (2007). “Social and environmental determinants of *Aedes aegypti* infestation in

- Central Havana: results of a case-control study nested in an integrated dengue surveillance programme in Cuba”. *Tropical medicine & international health : TM & IH* 12.4, pp. 503–510. DOI: 10.1111/j.1365-3156.2007.01818.x.
- Stefopoulou, A., Balatsos, G., Petraki, A., LaDeau, S. L., Papachristos, D., and Michaelakis, A. (2018). “Reducing *Aedes albopictus* breeding sites through education: A study in urban area”. *PLoS one* 13.11, e0202451. DOI: 10.1371/journal.pone.0202451.
- Stewart Ibarra, A. M., Luzadis, V. A., Borbor Cordova, M. J., Silva, M., Ordoñez, T., Beltrán Ayala, E., and Ryan, S. J. (2014). “A social-ecological analysis of community perceptions of dengue fever and *Aedes aegypti* in Machala, Ecuador”. *BMC public health* 14, p. 1135. DOI: 10.1186/1471-2458-14-1135.
- Stewart Ibarra, A. M., Ryan, S. J., Beltrán, E., Mejía, R., Silva, M., and Muñoz, A. (2013). “Dengue vector dynamics (*Aedes aegypti*) influenced by climate and social factors in Ecuador: implications for targeted control”. *PLoS one* 8.11, e78263. DOI: 10.1371/journal.pone.0078263.
- Sun, H., Dickens, B. L., Richards, D., Ong, J., Rajarethinam, J., Hassim, M. E. E., Lim, J. T., Carrasco, L. R., Aik, J., Yap, G., Cook, A. R., and Ng, L. C. (2021). “Spatio-temporal analysis of the main dengue vector populations in Singapore”. *Parasites & vectors* 14.1, p. 41. DOI: 10.1186/s13071-020-04554-9.
- Tedjou, A. N., Kamgang, B., Yougang, A. P., Njiokou, F., and Wondji, C. S. (2019). “Update on the geographical distribution and prevalence of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae), two major arbovirus vectors in Cameroon”. *PLoS neglected tropical diseases* 13.3, e0007137. DOI: 10.1371/journal.pntd.0007137.
- Trewin, B. J., Parry, H. R., Pagendam, D. E., Devine, G. J., Zalucki, M. P., Darbro, J. M., Jansen, C. C., and Schellhorn, N. A. (2021). “Simulating an invasion: unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, *Aedes aegypti*, in Brisbane, Australia”. *Biological invasions* 23.12, pp. 3891–3906. DOI: 10.1007/s10530-021-02619-z.
- Troyo, A., Calderón-Arguedas, O., Fuller, D. O., Solano, M. E., Avendaño, A., Arheart, K. L., Chadee, D. D., and Beier, J. C. (2008). “Seasonal profiles of *Aedes aegypti* (Diptera: Culicidae) larval habitats in an urban area of Costa Rica with a history of mosquito control”. *Journal of Vector Ecology* 33.1, pp. 76–88. DOI: 10.3376/1081-1710(2008)33[76:SPOAAD]2.0.CO;2.
- Tsuda, Y. and Takagi, M. (2001). “Survival and Development of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) Larvae Under a Seasonally Changing Environment in Nagasaki, Japan”. *Environmental Entomology* 30.5, pp. 855–860. DOI: 10.1603/0046-225X-30.5.855.
- Valdez, L. D., Sibona, G. J., and Condat, C. A. (2018). “Impact of rainfall on *Aedes aegypti* populations”. *Ecological Modelling* 385, pp. 96–105. DOI: 10.1016/j.ecolmodel.2018.07.003.
- Valença, M. A., Marteis, L. S., Steffler, L. M., Silva, A. M., and Santos, R. L. C. (2013). “Dynamics and characterization of *Aedes aegypti* (L.) (Diptera: Culicidae) key breeding sites”. *Neotropical entomology* 42.3, pp. 311–316. DOI: 10.1007/s13744-013-0118-4.
- Valle, D. and Aguiar, R. (2023). “Is the mosquito the problem? *Aedes aegypti* and urban arboviruses - contradictions and reflections”. *Epidemiologia e serviços de saúde : revista do Sistema Único de Saúde do Brasil* 32.2, e2023538. DOI: 10.1590/S2237-96222023000200024.
- Vannavong, N., Seidu, R., Stenström, T.-A., Dada, N., and Overgaard, H. J. (2017). “Effects of socio-demographic characteristics and household water management on

- Aedes aegypti* production in suburban and rural villages in Laos and Thailand”. *Parasites & vectors* 10.1, p. 170. DOI: 10.1186/s13071-017-2107-7.
- Vanwambeke, S. O., Somboon, P., Harbach, R. E., Isenstadt, M., Lambin, E. F., Walton, C., and Butlin, R. K. (2007). “Landscape and Land Cover Factors Influence the Presence of *Aedes* and *Anopheles* Larvae”. *Journal of medical entomology* 44.1, pp. 133–144. DOI: 10.1093/jmedent/41.5.133.
- Vasconcelos, A. S. V., Lima, J. S., Cardoso, R. T. N., Acebal, J. L., and Loaiza, A. M. (2022). “Optimal control of *Aedes aegypti* using rainfall and temperature data”. *Computational and Applied Mathematics* 41.3. DOI: 10.1007/s40314-022-01804-7.
- Vasconcelos, D., Yin, M. S., Wetjen, F., Herbst, A., Ziemer, T., Förster, A., Barkowsky, T., Nunes, N., and Haddawy, P. (2021). “Counting Mosquitoes in the Wild”. *Proceedings of the Conference on Information Technology for Social Good*. New York, NY, USA: ACM, pp. 43–48. DOI: 10.1145/3462203.3475914.
- Vezzani, D. and Schweigmann, N. (2002). “Suitability of containers from different sources as breeding sites of *Aedes aegypti* (L.) in a cemetery of Buenos Aires City, Argentina”. *Memorias do Instituto Oswaldo Cruz* 97.6, pp. 789–792. DOI: 10.1590/S0074-02762002000600006.
- Vinhal Frutuoso, L. C. and Barbosa Duraes, S. M. (2023). *Information Note No. 37/2023 - CGARB/ DEDT/ SVSA/ MS*. Ed. by Ministry of Health - Department of Communicable Diseases - General Coordination of Arbovirus Surveillance. Unpublished confidential document.
- Westby, K. M., Adalsteinsson, S. A., Biro, E. G., Beckermann, A. J., and Medley, K. A. (2021). “*Aedes albopictus* Populations and Larval Habitat Characteristics across the Landscape: Significant Differences Exist between Urban and Rural Land Use Types”. *Insects* 12.3. DOI: 10.3390/insects12030196.
- Whelan, P. I., Kurucz, N., Pettit, W. J., and Krause, V. (2020). “Elimination of *Aedes aegypti* in northern Australia, 2004-2006”. *Journal of vector ecology : journal of the Society for Vector Ecology* 45.1, pp. 118–126. DOI: 10.1111/jvec.12379.
- Wilk-da-Silva, R., Souza Leal Diniz, M. M. C. de, Marrelli, M. T., and Wilke, A. B. B. (2018). “Wing morphometric variability in *Aedes aegypti* (Diptera: Culicidae) from different urban built environments”. *Parasites & vectors* 11.1, p. 561. DOI: 10.1186/s13071-018-3154-4.
- Wilson, A. L., Courtenay, O., Kelly-Hope, L. A., Scott, T. W., Takken, W., Torr, S. J., and Lindsay, S. W. (2020). “The importance of vector control for the control and elimination of vector-borne diseases”. *PLoS neglected tropical diseases* 14.1, e0007831. DOI: 10.1371/journal.pntd.0007831.
- Wilson, M. E. (2011). “Chapter 4 Megacities and Emerging Infections: Case Study of Rio de Janeiro, Brazil”. *Megacities & Global Health*. Ed. by O. A. Khan and G. Pappas. American Public Health Association. DOI: 10.2105/9780875530031ch04.
- Wilson-Bahun, T. A., Kamgang, B., Lenga, A., and Wondji, C. S. (2020). “Larval ecology and infestation indices of two major arbovirus vectors, *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae), in Brazzaville, the capital city of the Republic of the Congo”. *Parasites & vectors* 13.1, p. 492. DOI: 10.1186/s13071-020-04374-x.
- Wong, G. K. and Jim, C. Y. (2017). “Urban-microclimate effect on vector mosquito abundance of tropical green roofs”. *Building and Environment* 112, pp. 63–76. DOI: 10.1016/j.buildenv.2016.11.028.
- World Mosquito Program (2023).
- Young, K. I., Mundis, S., Widen, S. G., Wood, T. G., Tesh, R. B., Cardoso, J., Vasilakis, N., Perera, D., and Hanley, K. A. (2017). “Abundance and distribution of sylvatic

- dengue virus vectors in three different land cover types in Sarawak, Malaysian Borneo”. *Parasites & vectors* 10.1, p. 406. DOI: 10.1186/s13071-017-2341-z.
- Zahouli, J. B. Z., Koudou, B. G., Müller, P., Malone, D., Tano, Y., and Utzinger, J. (2017). “Effect of land-use changes on the abundance, distribution, and host-seeking behavior of *Aedes* arbovirus vectors in oil palm-dominated landscapes, southeastern Côte d’Ivoire”. *PloS one* 12.12, e0189082. DOI: 10.1371/journal.pone.0189082.
- Zainon, N., Mohd Rahim, F. A., Roslan, D., and Abd Samat, A. H. (2016). “PREVENTION OF Aedes BREEDING HABITATS FOR URBAN HIGH-RISE BUILDING IN MALAYSIA”. *PLANNING MALAYSIA* 5. DOI: 10.21837/pm.v14i5.197.
- Zuur, A. F., Ieno, E. N., and Saveliev, A. A. (2017). “Beginner’s Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA”.



## Publication IV:

# Long-term Validation of Inner-Urban Mobility Metrics Derived From Twitter/X

**Authors.** Steffen Knoblauch, Simon Gross, Sven Lautenbach, Antonio Rocha, Marta C. Gonzalez, Bernd Resch, Dorian Arifi, Thomas Jänisch, Ivonne Morales, Alexander Zipf

**Outlet.** *Environment and Planning B: Urban Analytics and City Science*

**Status.** Published on 2024-10-14 under CC-BY 4.0

### Abstract.

Urban mobility analysis using Twitter as a proxy has gained significant attention in various application fields; however, long-term validation studies are scarce. This paper addresses this gap by assessing the reliability of Twitter data for modeling inner-urban mobility dynamics over a 27-month period in the metropolitan area of Rio de Janeiro, Brazil. The evaluation involves the validation of Twitter-derived mobility estimates at both temporal and spatial scales, employing over  $1.6 \times 10^{11}$  mobile phone records of around three million users during the non-stationary mobility period from April 2020 to June 2022, which coincided with the COVID-19 pandemic. The results highlight the need for caution when using Twitter for short-term modeling of urban mobility flows. Short-term inference can be influenced by Twitter policy changes and the availability of publicly accessible tweets. On the other hand, this long-term study demonstrates that employing multiple mobility metrics simultaneously, analyzing dynamic and static mobility changes concurrently, and employing robust preprocessing techniques such as rolling window downsampling can enhance the inference capabilities of Twitter data. These novel insights gained from a long-term perspective are vital, as Twitter - rebranded to X in 2023 - is extensively used by researchers worldwide to infer human movement patterns. Since conclusions drawn from studies using Twitter could be used to inform public policy, emergency response, and urban planning, evaluating the reliability of this data is of utmost importance.

**Keywords.** Mobile phone records · Twitter · Human Mobility · Urban · Rio de Janeiro · COVID-19

## Introduction

The substantial increase in the volume of geodata collected worldwide on human mobility behavior has the potential to yield valuable insights about various application domains, including urban transportation planning and epidemiology (Barbosa et al., 2018). By leveraging information on human trajectories, urban planners and policymakers can create more livable, sustainable, and responsive cities that cater to the needs of their inhabitants. This ranges from optimizing traffic flow and more efficient resource allocation to understanding infectious disease dynamics (Ruan et al., 2020; Wang et al., 2021a; Wang et al., 2021b). However, the availability of freely-accessible mobility data sources with high spatiotemporal resolution is limited, which often hampers quantitative research on unexplained phenomena associated with human movement patterns. Consequently, researchers have commonly resorted to open-access and georefer-

enced Twitter data as a proxy for inferring human mobility patterns. Twitter, a social media platform named X since 2023, enables users to tag their online posts with geocoordinates. Inferring mobility patterns from this data involves tracking the successive tweet locations of individuals over time. These locations typically do not represent trajectories in the conventional sense of semi-continuous paths but rather a random collection of locations with temporal references. Nonetheless, given that not all individuals use Twitter and not all content is posted with geocoordinates, there exists a concern regarding potential biases in this data and its inference capabilities for mobility patterns of the general population (Tsou et al., 2017; Zhao et al., 2021).

In literature, there is a paucity of studies that justify and validate the use of Twitter as a reliable proxy for mobility patterns, particularly on a small spatial scale where Twitter data may be extremely sparse. The interest in using Twitter data for mobility-related urban phenomena, however, is increasingly high, encompassing real-time event monitoring for example of traffic congestion and accidents (Bao et al., 2017; Zia et al., 2022), disaster relief to improve coordination of rescue efforts (Reynard and Shirgaokar, 2019; Wang and Taylor, 2018), social sensing of urban land use (Soliman et al., 2017), urban planning (Milusheva et al., 2021), as well as the early detection and analysis of disease outbreaks (Bisanzio et al., 2020a; Bisanzio et al., 2020b; Huang et al., 2020a; Huang et al., 2020b). Validation studies that exist on larger scales have employed survey data (Terroso-Saenz et al., 2022a; Terroso-Saenz et al., 2022b), census tracts (Petutschnig et al., 2022), or tourism statistics (Hawelka et al., 2014; Provenzano et al., 2018) for evaluation purposes. At the urban scale similar data sources have been utilized, but only five validation studies have been conducted to the best of our knowledge. Kurkcu A. et al. (2016) compared Twitter data with regional household travel surveys by calculating various mobility metrics, such as the radius of gyration and origin-destination flows, for New York City. However, this study did not examine temporal mobility trends over longer time periods. Lenormand et al. (2014) performed a comparison of Twitter data, mobile phone records, and census statistics, assessing spatiotemporal mobility metrics for Barcelona and Madrid. This study compared datasets from two different time frames, which we consider to have limited validity particularly during non-stationary periods like pandemics. The same limitation applies to the studies conducted by Qian et al. (2018), Steiger et al. (2015), and Osorio-Arjona and García-Palomares (2019), as they used either survey or census data from earlier years than when the Twitter data was collected.

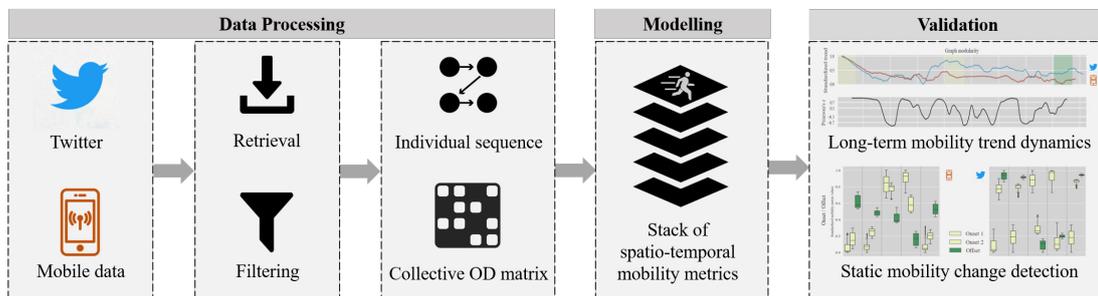
Consolidating aforementioned findings highlights the research gap concerning long-term validation studies pertaining to inner-urban mobility metrics extracted from Twitter data. More specifically, this relates to employing a time-overlapping validation set to assess the accuracy and reliability of Twitter-derived mobility estimates for urban areas over an extended time period. Given these limitations, this paper introduces a novel urban validation study comparing long-term mobility dynamics extracted from geolocated tweets with mobile phone records covering a time frame of 27 months. The research was carried out in the city of Rio de Janeiro during and after COVID-19-related lockdowns, specifically from April 6<sup>th</sup>, 2020 to June 30<sup>th</sup>, 2022. The second-largest city in Brazil was chosen due to the availability for mobile phone records and the extended use of Twitter

in the country, which ranks fourth globally in terms of Twitter usage (Statista, 2023). Furthermore, the metropolitan area of Rio de Janeiro, with its nearly 14 million inhabitants, provided a suitable urban landscape to address research objectives related to urban science. More specifically, we addressed the following two research questions (RQs):

- **RQ1:** To what extent can the method of rolling window downsampling assist in counteracting the scarcity of daily-geocoded tweet sequences in cities?
- **RQ2:** How similar are urban mobility patterns derived from Twitter to long-term spatiotemporal mobility metrics derived from mobile phone data?

## 2 Materials and Methods

In order to answer the derived research questions, we propose a consecutive framework of data processing, modelling, and validation (cf. Figure 1). The processing part describes the retrieval and filtering of applied datasets as well as the generation of individual movement trajectories and collective origin-destination (OD) matrices. In the modelling section, a stack of five representative spatiotemporal mobility metrics were calculated. The validation part was divided into two studies: (1) a dynamic assessment of long-term mobility trends and (2) a static validation of mobility change detection capabilities.



**Figure 1:** Workflow for the comparison of inner-urban mobility metrics derived from Twitter and mobile phone records. The quantitative validation study is divided into two parts: (1) a long-term trend analysis and (2) a validation of Twitter’s capability for static mobility change detection.

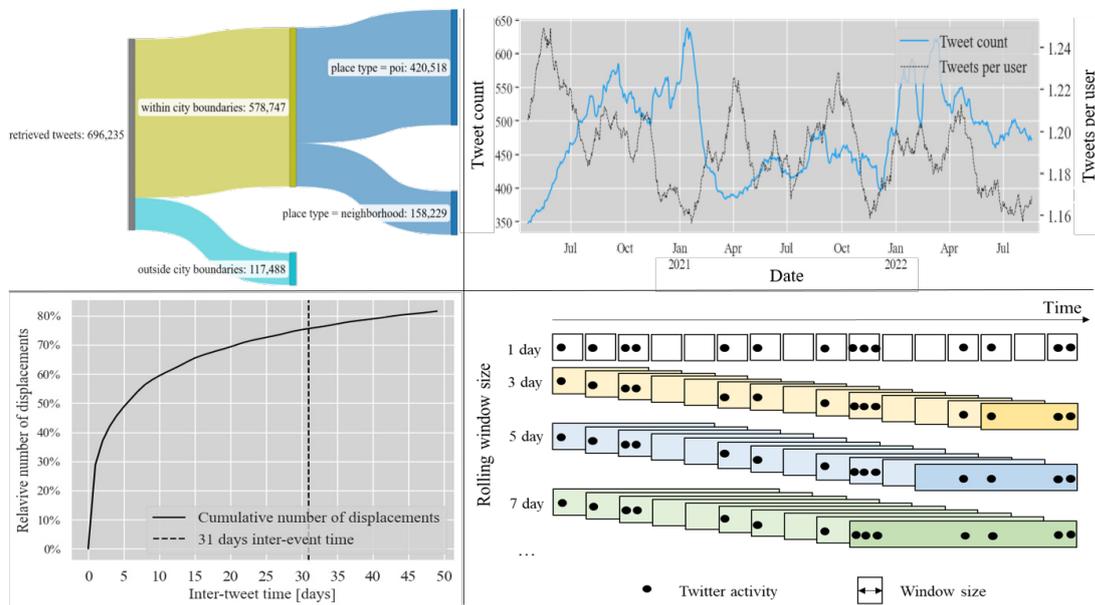
### 2.1 Data Processing

#### 2.1.1 Twitter data

Twitter data was derived from the publicly available Twitter API v2 (Twitter, Inc., 2023b) with special terms of use for academic research (Twitter, Inc., 2023a). The research licence supported the collection of more precise, complete, and unbiased datasets than the publicly available API for commercial use. API access policies and privacy concerns can undergo constant change. We treated all data in accordance with stringent privacy by design guidelines published in Kounadi et al. (2018) and Kounadi and Resch (2018). During the API request we specified an API token, start and end timestamps for the period of analysis,

details regarding the case study region presented as rectangular bounding boxes, and two parameters to filter retweeted content and tweets lacking geolocation. Twitter stores geotags implicitly via place IDs. A place ID can be either a point of interest (POI) such as a bus stop close to a user location, a neighborhood, a city or a country name. As we are only interested in inner-urban movement patterns for the city of Rio de Janeiro, tweets with any ‘place\_type’ larger than a city were excluded from the API request. This selection resulted in 696235 tweets for the whole study period of 27 months.

After this initial data retrieval, tweets which were located inside the rectangular bounding boxes but outside of the city boundaries of Rio de Janeiro were removed. Since we encountered issues with the shape and naming of city districts within Twitter, tweets with the ‘place\_type’ tag ‘neighborhood’ were additionally filtered out. This filtering step was applied to prevent possible distortion of the Twitter data in space and, consequently also resulted in a higher resolution of geocoded tweets. The final analysis was therefore conducted on tweets of ‘place\_type = poi’ only. The ratio of tweets per user exhibited notable heterogeneity over time (cf. Figure 2 - top), as indicated by a standard deviation of 20.96. To address this imbalance, we employed supplementary bot filtering technique. Through a comprehensive examination of tweet distributions across all users, we identified and filtered out tweets originating from potential bot accounts by implementing a maximum daily tweet threshold of 50 and a maximum daily tweet share threshold of one percent. This was found to be consistent with methodologies employed in other twitter studies (Osorio-Arjona and García-Palomares, 2019; Terroso-Saenz et al., 2022b).



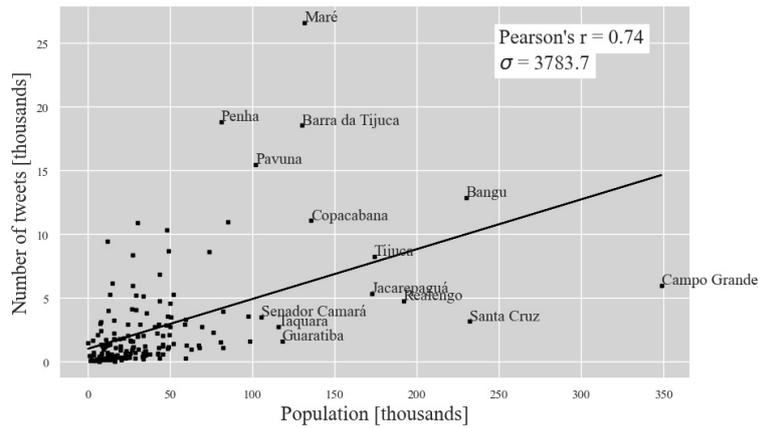
**Figure 2:** Sankey diagram of Twitter filtering process (top left); daily geolocated Twitter stream used for long-term validation study with corresponding number of unique users and tweet amount per user in the city of Rio de Janeiro (top right); histogram of inter-tweet time for study period in the city of Rio de Janeiro (bottom left); schematic rolling window downsampling concept for temporal Twitter signals (bottom right).

After data cleaning 420518 geolocated tweets from 107500 unique users were used to build individual user movement sequences with the scikit-mobility python

library (Pappalardo et al., 2022). To address the daily scarcity of geolocated tweet sequences from individual users (cf. Figure 2 - bottom left), a rolling window downsampling approach was implemented. This method, contingent upon the chosen window size, can increase data volume, enabling the calculation of individual movement trajectories by enhancing the length of tweet sequences from unique Twitter users (Li, 2008). This method enables the calculation of daily mobility metrics while effectively smoothing out short-term fluctuations and outliers, thereby preserving the temporal trend within the dataset. This collective functionality renders it a suitable approach for deriving daily mobility trend signals from limited datasets, such as daily geo-tagged tweets from urban areas, aligning with the necessary objectives of the study. From a practical standpoint, this method involves aggregating and sequencing tweets accumulated over multiple days to compute mobility metrics specifically for a single day positioned at the center of the aggregation window (cf. Figure 2 - bottom right).

However, it is crucial to acknowledge that this method may also entail certain adverse consequences, such as diminished granularity or analytical precision. To address this concern, we applied a rolling window downsampling approach using a grid search across various, uneven window sizes spanning from three to 31 days, resulting in 15 distinct temporal signals. The optimal choice of these window sizes to calculate urban mobility metrics was evaluated as described in section 3.2.1. The selection of the range of window sizes employed for the grid search was predicated on the objective of encompassing around 75 percent of individual displacements identified in the Twitter data (cf. Figure 2 - bottom left).

Individual human movement trajectories were retrieved from the list of temporally-ordered tweet locations of single users. The collection of these sequences over all users contributed to the generation of collective OD matrices. During this process each tweet location was matched to one of the 163 neighborhoods present in the city of Rio de Janeiro, characterized by a different ratio of tweets per capita for the residential population (cf. Figure 3). The scale of neighborhoods was chosen to align with many census statistics, which could be potentially relevant for follow-up studies. The geographic matching process resulted in daily OD matrices of shape 163x163 used for the subsequent calculation of spatiotemporal mobility metrics as explained in section 2.2. To enhance comparison capabilities with mobile phone data, OD matrix entries on the diagonal were set to zero and normalized by the amount of measured movements, which is equal to the remaining sum of OD matrix entries.

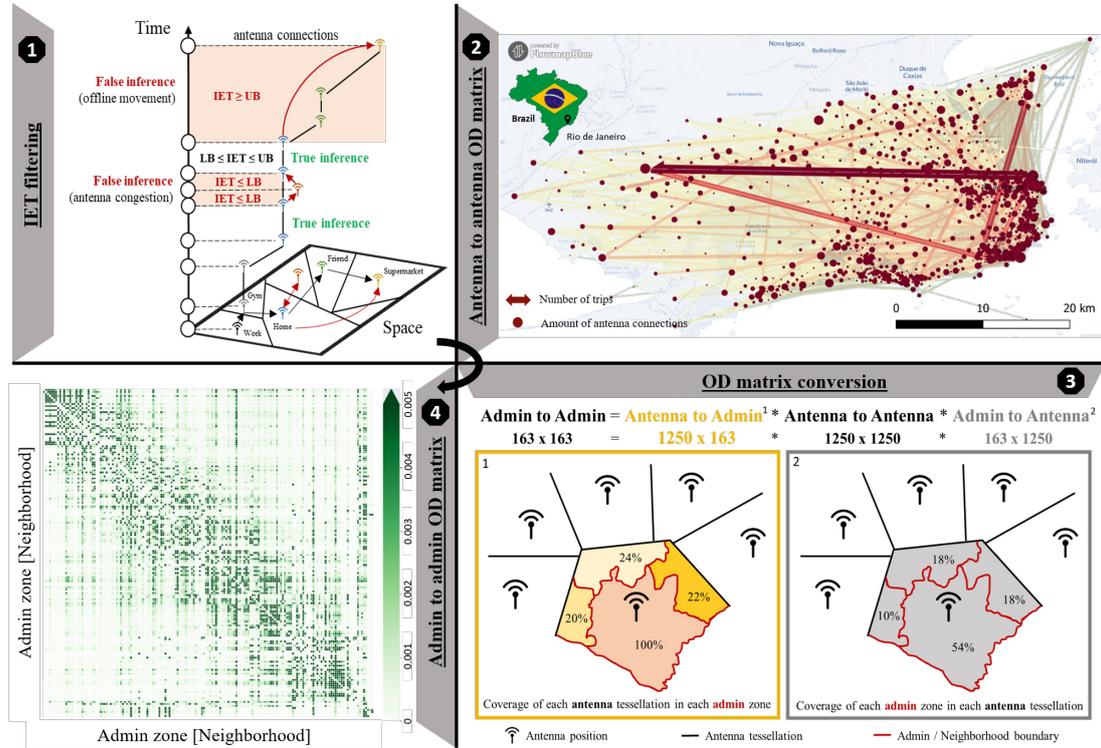


**Figure 3:** Ratio of tweets per capita for the residential neighborhood population in the city of Rio de Janeiro.

### 2.1.2 Mobile phone data

As a validation set we used anonymized mobile phone records provided by a large Brazilian telecommunications company. The dataset included individual antenna connections from approximately three million unique users over a time period of 27 months. This is equal to an approximated penetration rate of around 45 percent for the population of the city of Rio de Janeiro. The temporal resolution of the raw data was five minutes. The data was provided at the level of the antennas (cf. Figure 4 - top right). The mobile phone user is typically connected to the closest antenna, which is used as a proxy for the position of the user at this point in time. The number of antennas in our data set varied daily between 1200 and 1250 due to technical failures of some antennas. An antenna connection from an user was recorded when sending a text message, using mobile internet data, or making a call. We retrieved and processed the data of 164250 million mobile phone records via the distributed computing tool Apache Spark as well as the GPU-accelerated parallel computing framework Dask using the mobilkit python library (Ubaldi et al., 2021). As a first cleaning step, we dropped connections with antennas outside the city boundaries. In order to derive human movement patterns, we generated a sequence of antenna connections for each user over the whole time period using a machine with 7 TB of local scratch. To increase the informative power of successive antenna connections for inferring human movement patterns, we introduced a lower bound (LB) and upper bound (UB) as filters for the inter event time (IET) between sequential antenna connections from a single user as proposed by Zhao et al. (2019). As a result, successive antenna connections between which less than 15 minutes (LB) or more than four hours (UB) elapsed were not counted as movements (cf. Figure 4 - top left). The introduction of a LB was justified by the fact that antenna congestion can cause the user to jump back and forth between antennas without physical moving. A UB was introduced to avoid the counting of movements that are not necessarily made in a direct way. The choice of the lower threshold was selected based on Zhao et al. (2019) and Schlosser et al. (2020). The choice of the upper threshold was inspired by Barboza et al. (2021). OD matrices were created based on IET-filtered daily user sequences. The entries in the diagonal of daily OD matrices were set equal to zero. To ensure comparability with the OD matrices of Twitter data, the

OD matrices were normalized by the overall amount of movement activity before being converted from antenna format (1250x1250) to district format (163x163) using methods from Fabrikant (2017) (cf. Figure 4 - bottom right).



**Figure 4:** Schematic time line of recorded antenna connections with IET-filtering (top left); resulting antenna to antenna OD matrix flows for the first day of analysis derived from mobile phone data, where darker color represents larger movements (top right); formula for OD matrix conversion from antenna to admin level with schematic illustrations of “antenna to admin” (orange) and “admin to antenna” (grey) matrix calculation (bottom right); resulting admin to admin OD matrix heatmap for the first day of analysis using mobile phone data, where darker shade of green describes a higher percentage of measured movement in the city (bottom left).

## 2.2 Mobility metrics

In order to answer our second research question, whether Twitter is a good proxy for modeling inner-urban human movement patterns, we calculated five spatiotemporal mobility metrics. These include the (i) total number of movements, (ii) the average movement distance of individuals, (iii) land use activity metrics, (iv) graph modularity, and (v) the radius of gyration (cf. Figure A. 1).

$$M_t = \sum_{i=1}^{163} \sum_{j=1}^{163} a_{i,j} \quad (1)$$

$$OD_t = \begin{bmatrix} a_{1,1} & \dots & a_{1,163} \\ \vdots & \ddots & \vdots \\ a_{163,1} & \dots & a_{163,163} \end{bmatrix} \quad (2)$$

$$\bar{D}_t = \frac{1}{M_t} \sum_{u=1}^U \left( \sum_{i=1}^{n_u-1} \sqrt{(x_{u,i+1} - x_{u,i})^2 + (y_{u,i+1} - y_{u,i})^2} \right) \quad (3)$$

Inspired by previous research on human movement patterns (Aletta et al., 2020; Haas et al., 2020; Hensher et al., 2021; Li et al., 2021; Mützel and Scheiner, 2022; Schlosser et al., 2020), the total number of all movements, denoted as  $M_t$ , (cf. Formula 1), was calculated using daily OD matrices, where  $a_{i,j} = 0$  for  $i = j$  (cf. Formula 2). The daily average travel distance over all users  $U$ , denoted as  $\bar{D}_t$ , was derived from the number of visited locations  $n_u$  in the IET-filtered user sequences for each user  $u$  (cf. Formula 3). For each movement from the  $i$ -th location visited by user  $u$  on day  $t$  to the  $(i + 1)$ -th location, the Euclidean distance between each consecutive pair of visited locations by user  $u$  ( $x_{u,i}, y_{u,i}$ ) was calculated. The geolocations of sequential tweets were used as location coordinates for Twitter data. For mobile data, the distance between antennas was utilized. The sum of all tracked paths was then divided by the total number of considered movements,  $M_t$ , to calculate the average travel distance in kilometers, following the precedent set by other research papers (Abdullah et al., 2021; Abdullah et al., 2020; Engle et al., 2020; Fatmi, 2020; Gao et al., 2020a; Gao et al., 2020b; Pardo et al., 2021; Park et al., 2022).

$$\% \text{ activity in residential area}_t = \frac{\text{Number of Tweets or Mobile Activity in Residential Areas on day } t}{\text{Total Number of Tweets or Mobile Activity on day } t} \times 100\% \quad (4)$$

We calculated land use-dependent activity metrics using land use land cover maps from the DATA.RIO portal (Municipality of Rio de Janeiro, 2022). These metrics can provide information about the percentage of Twitter or mobile activity that can be assigned to a certain land use structure (Aktay et al., 2020; Da Cavalcante Silva et al., 2021; Hakim et al., 2021; Nanda et al., 2022; Ossimetha et al., 2021; Paez, 2020; Saha et al., 2020; Saha et al., 2021; Shumway-Cook et al., 2005; Sulyok and Walker, 2020; Zhu et al., 2020). In our analysis, we measured the percentage of activity for six types of typical urban land cover categories (residential, public, leisure, industry, education, commerce) present in the city of Rio de Janeiro. For the validation of Twitter, only the percentage of activity in residential areas was used as a representative of this mobility metric type (cf. Formula 4). The inclusion of all land use-dependent activity metrics was rejected to improve clarity and diversify the analyzed mobility metrics in this study. Several metrics of land use-dependent activity were considered redundant. Residential areas were chosen as the land class of highest interest as they promised the highest variability related to lockdown style policies. For calculating land use-dependent mobility metrics, tweet POI and antenna location were used correspondingly.

The graph modularity, a measure indicating the extent of links within communities compared to links between communities (cf. Figure 5), was calculated using the Louvain algorithm. Modularity, denoted by  $Q$ , is computed as the difference between the observed fraction of intra-community edges and the expected fraction if edges were distributed randomly (Blondel et al., 2008). It is defined as follows:

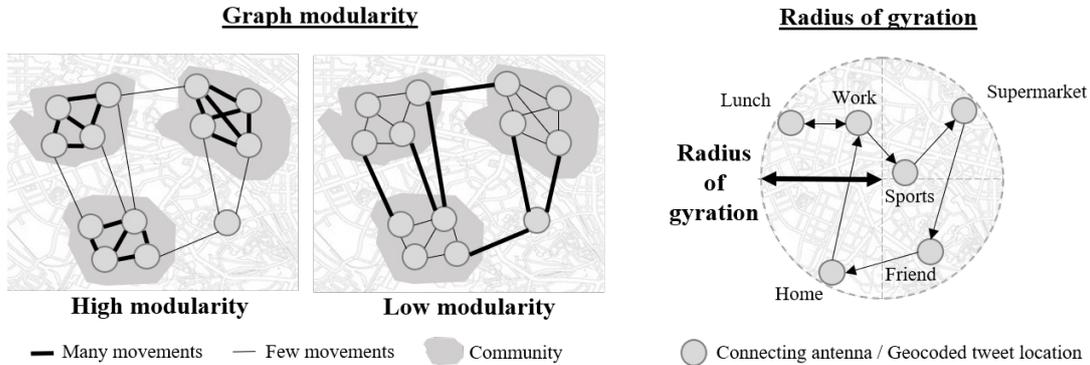
$$Q_t = \frac{1}{2m_t} \sum_{ij} \left( A_{ij,t} - \frac{k_{i,t}k_{j,t}}{2m_t} \right) \delta(c_{i,t}, c_{j,t}) \quad (5)$$

where  $A_{ij}$  is the element in the adjacency matrix representing the connection between nodes  $i$  and  $j$ ,  $k_i$  and  $k_j$  are the degrees of nodes  $i$  and  $j$  respectively,  $m$  is the total number of edges, and  $\delta(c_i, c_j)$  is 1 if nodes  $i$  and  $j$  belong to the same

community and 0 otherwise. In our context, nodes represent neighborhoods, and edge weights represent the sum of traced movements between neighborhoods. The Louvain algorithm operates on an undirected graph constructed using the origin-destination (OD) matrices specified beforehand in Formula 2, which were initially directed to represent one-way movements between origins and destinations and made undirected by multiplying them with their transposes. This process ensures that each element  $A_{ij}$  of the adjacency matrix represents the total movements between nodes  $i$  and  $j$ , accounting for both directions, in contrast to  $a_{i,j}$  which represents one-way movements. The Louvain modularity value ranges from -0.5 to 1, where higher values indicate mobility networks with more inner-community movements than outer-community movements (Heiler et al., 2020; Newman, 2006; Yildirimoglu and Kim, 2018).

$$R_{g,t} = \frac{1}{M_t} \sum_{u=1}^U \sqrt{\frac{1}{n_{u,t}} \sum_{i=1}^{n_{u,t}} (x_{u,i,t} - \bar{x}_{u,t})^2 + (y_{u,i,t} - \bar{y}_{u,t})^2} \quad (6)$$

The radius of gyration  $R_g$  indicates the average radius of movement of a single user  $u$  (cf. Figure 5). We averaged this value over all recorded users  $U$  and calculated it on a daily basis. Analogous to the methodology applied in computing the average movement distance, we conducted distance calculations between the Twitter POIs and the antenna location, respectively. Both calculations were run on the basis of the IET-filtered user sequences (Hernando et al., 2021; Kishore et al., 2020; Liu et al., 2018; Wang and Taylor, 2014). The variables  $\bar{x}_u$  and  $\bar{y}_u$  correspond to the mean of the  $x$ -coordinates or  $y$ -coordinates of user's visited locations on day  $t$ .

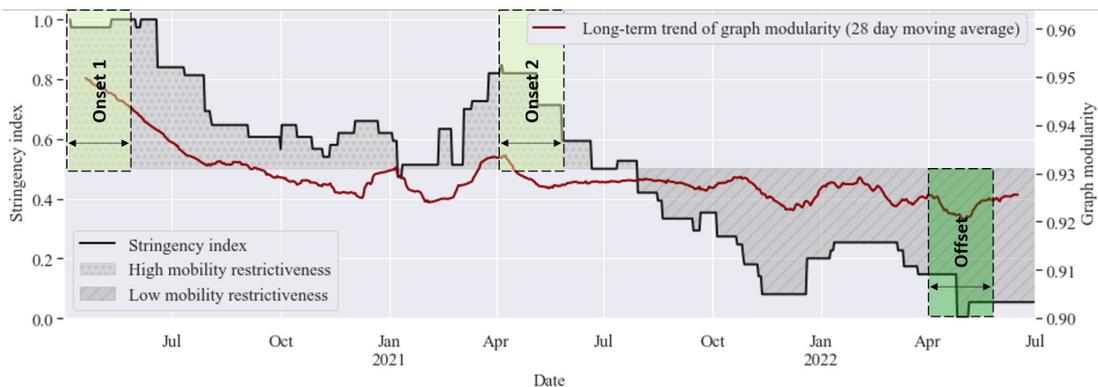


**Figure 5:** Schematic concept of graph modularity and radius of gyration. Graph modularity is a measure for the strength of the division of a graph into communities, based on the density of connections within communities compared to connections between communities. The radius of gyration refers to the average travel distance of an individual measured from the center of its movement circle, representing the overall distribution of visited places.

### 2.3 Long-term validation of urban mobility patterns derived from Twitter

The long-term validation of urban mobility metrics derived from Twitter was conducted over a non-stationary mobility period of 817 days. This time period covers the major peaks of the COVID-19 pandemic including subsequent months

with high to low mobility restrictions implemented by the local state and municipal government of Rio de Janeiro (Mathieu et al., 2020). The capability of Twitter as a data source to detect long-term mobility change in urban environments was evaluated using mobile phone records. To justify the utilization of mobile phone records as a 'ground-truth' validation set in our case study, we previously tested spatiotemporal mobility metrics derived from mobile phone data as valid evaluation sets for modeling real-world human movement behavior at an urban scale. For this evaluation, we obtained the stringency index for the city of Rio de Janeiro (cf. Figure 6), which is a globally-standardized indicator of politically-implemented mobility restrictions affecting human movement behaviour (Mathieu et al., 2020). It is a widely-used indicator derived from ordinal measurements for containment, closure policies, and public information campaigns. For the whole study time period during and after the COVID-19 pandemic, we calculated an average absolute Pearson correlation coefficient of 0.7 between all mobility metrics and the stringency index. The graph modularity mobility metric showed the highest overall Pearson correlation coefficient of 0.77. The main advantage of mobile phone records over the stringency index as an assessment dataset for this case study was the high temporal resolution of mobility measurements on a daily basis.



**Figure 6:** Stringency index recording the strictness of lockdown style policies in the city of Rio de Janeiro and graph modularity measurements derived from mobile phone data (red). On- and offset time periods indicate manually selected time frames of high to low mobility restrictions defined for static mobility change detection analysis.

The quantitative assessment involved the computation of moving window synchrony among long-term mobility trend signals indicating individual and collective mobility metrics derived from Twitter and mobile phone data as outlined in section 2.2. Time series synchrony denotes the extent to which time series exhibit similar patterns across multiple time steps. Unlike correlation, which quantifies the strength and direction of the linear relationship between time series, synchrony characterizes the temporal alignment and similarity in temporal patterns. We approximated the moving window synchrony by calculating the daily Pearson's correlation coefficients applying a window size of 60-days.

Long-term trend signals of calculated daily mobility metrics were generated by applying a moving average of 28 days and MinMax-Standardization considering the whole time frame of analysis. Moving average size for trend decomposition was selected based on visual diagnostics to remove weekly oscillations and outliers that appear due to technical antenna failures (cf. Figure A. 1). The moving average

size of 28 days seemed to generate a plausible trade-off signal between long-term trend and short-term mobility changes. Absolute moving window synchrony surpassing values of 0.7 was classified as indicating a high level of alignment, while values below 0.3 were considered to signify a weak tendency to exhibit similar temporal pattern. Intermediate moving window synchrony values ranging from 0.3 to 0.7 represented moderate alignment of events and changes in our study.

Three on- and offset periods were defined based on the stringency index to evaluate the capability of Twitter to additionally measure static change detection (cf. Figure 6). Considering implemented mobility restrictions in the city of Rio de Janeiro, we classified the two-month time periods from April 6<sup>th</sup> to June 6<sup>th</sup> in 2020 and 2021 as lockdown style periods (onset) and the time frame from April 6<sup>th</sup> to June 6<sup>th</sup> in 2022 as post-lockdown period (offset). With this selection, our goal was to include time intervals that exhibit diverse levels of human mobility, independent of potential seasonal fluctuations, throughout the three years of analysis. We selected a two-month interval period starting at the beginning of our analysis to capture both static mobility circumstances and their associated changes. The outcomes of the static mobility change detection were displayed through boxplots and compared with weekday/weekend onsets and offsets extracted from the entire analysis time period. To provide statistical quantification for static urban mobility changes, we conducted Mann-Whitney U tests between on- and offset periods, applying a confidence threshold of 0.05.

## 3 Results and Discussion

### 3.1 RQ1: Evaluation of rolling window downsampling

Examining the initial time period of analysis spanning from April 2020 to September 2020 (cf. Figure 7), all computed mobility metrics derived from Twitter exhibited discernible patterns that aligned with our expectations based on the implemented lockdown measures in the city of Rio de Janeiro. Notably, while the long-term trend of the graph modularity metrics and the percentage of activity in residential areas decreased, the long-term trends of average movement distance, overall movement volume, and the radius of gyration increased.

During the subsequent time period from September 2020 to May 2021, all mobility metrics derived from Twitter, except the percentage of activity in residential areas, displayed unexpected changes. They all showed a rapid shift starting in February 2021 dis-aligning our assumptions on more or less constant mobility behaviour in that time period. Coinciding with this period, there was a sharp decline in the number of geolocated tweets collected via the public Twitter API (cf. Figure 2). We hypothesize that this decline was attributed to changes in the terms of use implemented by Twitter. However, official evidence of regulatory changes during that specific time period has not been found. Additional experiments using a constant amount of tweets per day, derived by the 98<sup>th</sup> percentile of tweet volume in the corresponding rolling window subset, showed a similar shift in mobility metrics (cf. Figure A. 3). This highlights the robustness of calculated mobility metrics in the face of daily fluctuations in the number of tweets. For the analysis period subsequent to May 2021, the calculated mobility metrics once again aligned with our expectations and confirmed our knowledge of fewer mobility restrictions implemented in the city of Rio de Janeiro following the COVID-19

pandemic.



**Figure 7:** Standardized inner-urban mobility metrics derived from daily tweet sequences applying rolling window downsampling (RWDS). Results of 7-day and 27-day rolling windows (dark blue) are compared with the daily raw and trend signal of Twitter mobility metrics without applying RWDS (light blue). The trend signals are calculated using a moving average of 28 days. Non-standardized mobility metrics derived from Twitter for a 11-day rolling window size are visualized in Figure A. 2.

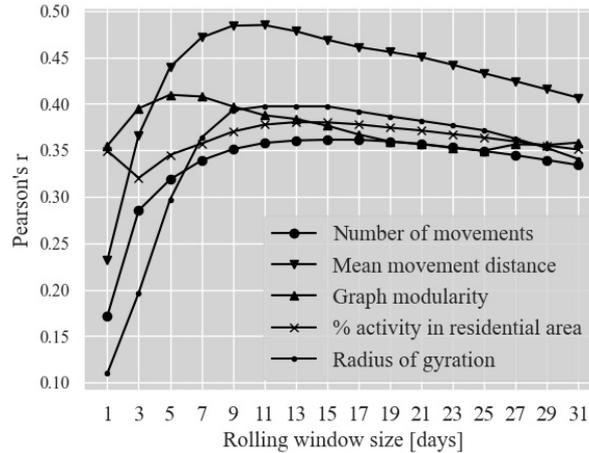
The results also demonstrate that, while a moving average can effectively eliminate weekly fluctuations and data noise, it does not suffice for generating accurate long-term trends for all considered mobility metrics in this analysis. However, when combined with the specifically designed rolling window down-sampling (RWDS) approach, more precise long-term mobility trends can be derived. This effect becomes particularly evident when examining the calculated graph modularity metrics in our case study, as the modularity values between the one-day window size signals and the seven- or 27-day rolling window size signals exhibit larger differences. In contrast, for other calculated mobility metrics, the impact of RWDS appears to have relatively low significance and yields effects comparable to those obtained by calculating a one-day window trend signal. Supplementary materials provide corresponding results of daily mobility metrics calculated without applying a moving average (cf. Table A. 3). The influence of different rolling window sizes is more extensively investigated in the subsequent section in conjunction with long-term trends derived from mobile phone data.

### **3.2 RQ2: Validation of long-term urban mobility patterns derived from Twitter**

Long-term validations of urban mobility metrics derived from Twitter are infrequent, despite the well-established usage of Twitter applications in various research domains worldwide. However, the outcomes of our comprehensive long-term validation study emphasize the need for caution when utilizing Twitter data for urban studies within restricted time frames. Although urban mobility metrics derived from Twitter may exhibit high correlation values with mobility metrics computed from mobile phone data during short time periods, long-term validation with mobile phone data reveals fluctuating deviations (cf. Figure 9). This phenomenon can potentially give rise to erroneous assumptions when relying solely on Twitter as a reliable source for modeling human movement patterns.

#### **3.2.1 Sensitivity of rolling window size**

The results presented earlier in section 3.1 demonstrate that the RWDS method is a valuable tool for addressing the data scarcity challenge associated with urban Twitter data and deriving more precise long-term mobility trends. However, additional findings highlight the significant dependence of these findings on the chosen rolling window size (cf. Figure 8). In our experiments we observed the highest average correlation value between mobility metrics from Twitter and mobile data when using an 11-day rolling window size. Increasing the window size from one day to three days had the most pronounced effect on the calculated Pearson correlation values. For window sizes exceeding 11 days, the correlation values remained consistently high but showed a slight flattening. This can be attributed to the loss of high-resolution information resulting from the application of larger window sizes beyond 11 days. These findings align with our expectations regarding the functionality of the RWDS method described in section 2.1.1. The mean movement distance index yielded the highest average Pearson correlation coefficient among all considered mobility metrics, achieving its peak of 0.48 at the 11-day rolling window downsampling size (cf. Table A. 1).



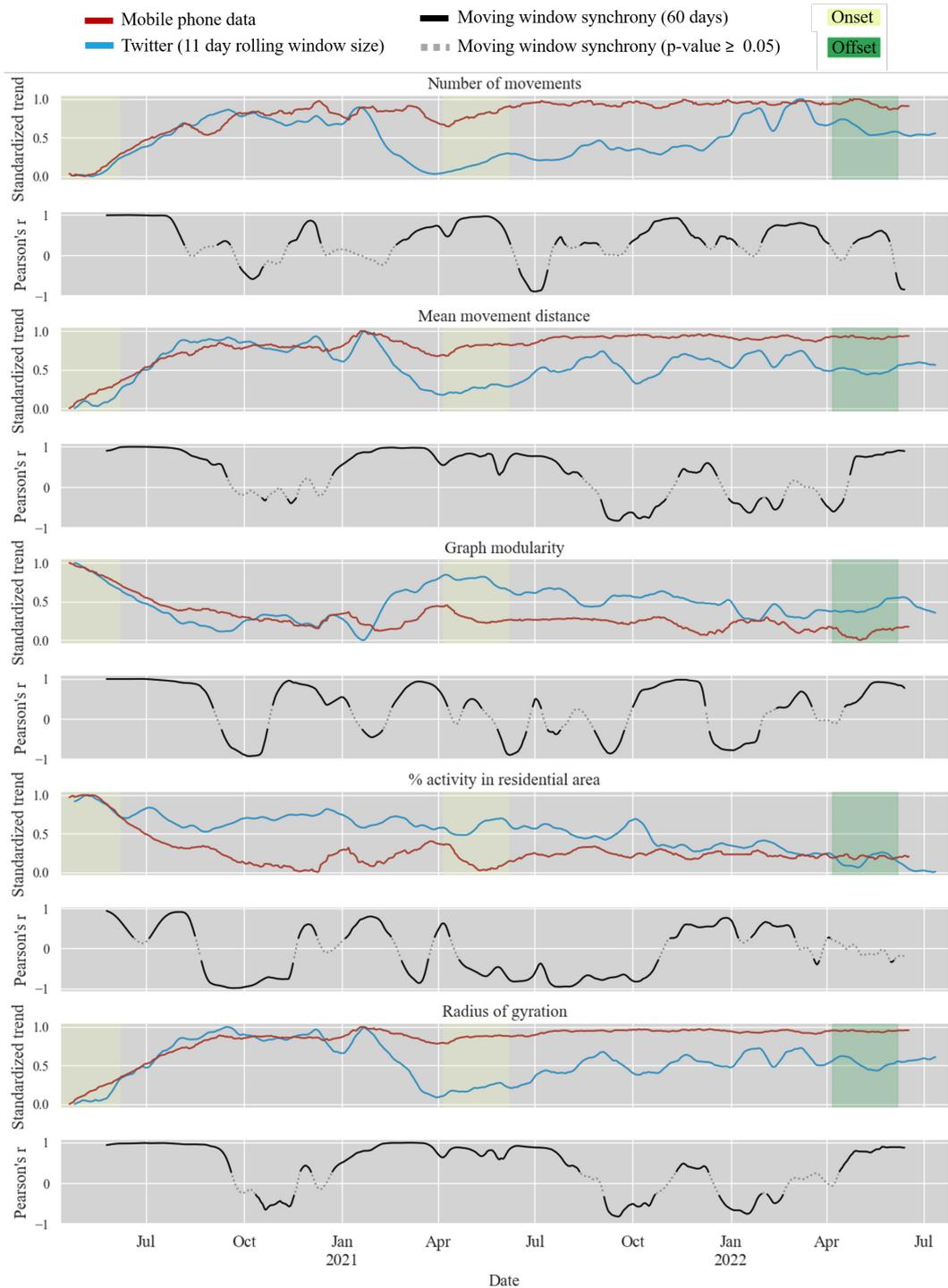
**Figure 8:** Mean Pearson’s correlation coefficients calculated over the whole time period of analysis between mobility metrics derived from Twitter and mobile phone data considering varying window sizes for RWDS.

### 3.2.2 Long-term mobility trend:

During the dynamic analysis of the long-term trend of calculated mobility metrics using moving window synchrony, it becomes evident that the Pearson’s correlation coefficients exhibit substantial variations over time for all the calculated mobility signals (cf. Figure 9). We observed the occurrence of short time periods characterized by both extremely high and extremely low correlation values. These findings indicate that the informative capacity of mobility metrics derived from Twitter exhibits temporal variability and is strongly contingent upon the chosen time frame for analysis. During the initial phase of the study period, when the most stringent mobility restrictions were implemented (cf. Figure 6), we observed high positive correlation values across all metrics simultaneously. Conversely, we did not observe similar prolonged time periods characterized by a weak alignment, as indicated by low Pearson’s correlation coefficients around zero. Notably, higher moving window correlation values exhibited greater statistical significance than lower values.

To eliminate the possibility of spurious correlations, all time series were examined for unit roots using the appropriate version of the Dickey-Fuller test before calculating Pearson correlation coefficients. The test results indicated that seven out of ten time series were stationary, allowing for the calculation of Pearson correlation coefficients. However, the time series for “Number of movements”, “Graph modularity”, and “% activity in residential areas” measured based on Twitter data, remained non-stationary. Following the “Standard sequence of steps for dealing with non-stationary time series” as outlined by (Studenmund, 2017), we tested the pairs of Twitter data and mobile phone time series for the metrics “Number of movements”, “Graph modularity”, “% activity in residential areas” for cointegration using the Engle-Granger test. The Engle-Granger test results indicated that the time series for the metrics “Number of movements” and “% activity in residential areas” were cointegrated at a confidence level of 95%, while the time series for the metric “Graph modularity” were cointegrated at a confidence level of 90%. According to (Studenmund, 2017), if the variables have unit roots and are also cointegrated, this allows for the calculation of the Pear-

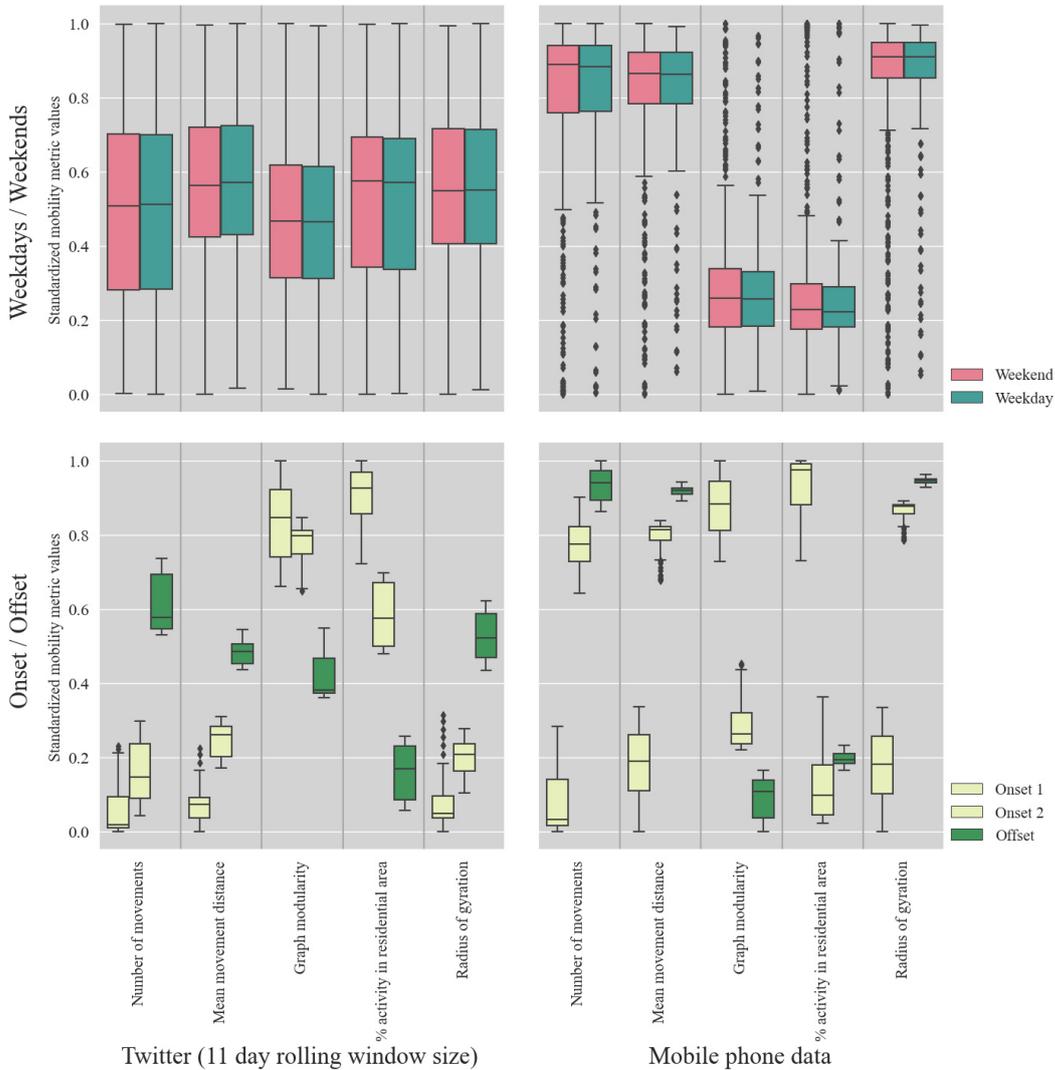
son correlation coefficient using the original units, thereby ruling out spurious correlations.



**Figure 9:** Long-term mobility metrics derived from Twitter (blue) applying rolling windows size of 11 days and mobile phone data (red) including moving window synchrony of 60 days (black), where on- and offset represent time periods of high and low mobility restrictions. The moving window correlations exhibited statistical significance, except for transitional phases between positive and negative synchrony. Non-standardized mobility metrics are visualized in Figure A. 1 and Figure A. 2. (cf. Figure A. 4 for more detailed visualization).

### 3.2.3 Static mobility change detection:

Additional findings from a static change detection analysis reinforce the results of our long-term trend analysis (cf. Figure 10). While it is evident that Twitter data does not always accurately capture long-term mobility trends, it does have the potential to detect significant (cf. Table A. 2) inner-urban mobility changes measured by mobile phone data and indicate the correct direction of the shift. In our case study, this holds true for all the measured variables except for the percentage of activity in residential areas during the time period of the second onset. In summary, we conclude that both the Twitter and mobile phone datasets synchronously detected the shift in inner-urban human movement behavior between the years 2020, 2021, and 2022, attributable to COVID-19 lockdown policies. Static mobility changes between weekdays and weekends were not detected to be significant (cf. Table A. 2) when testing both datasets, leading to the conclusion that Twitter can be a useful substitute for mobile phone records when trying to derive the direction of static inner-urban mobility shifts.

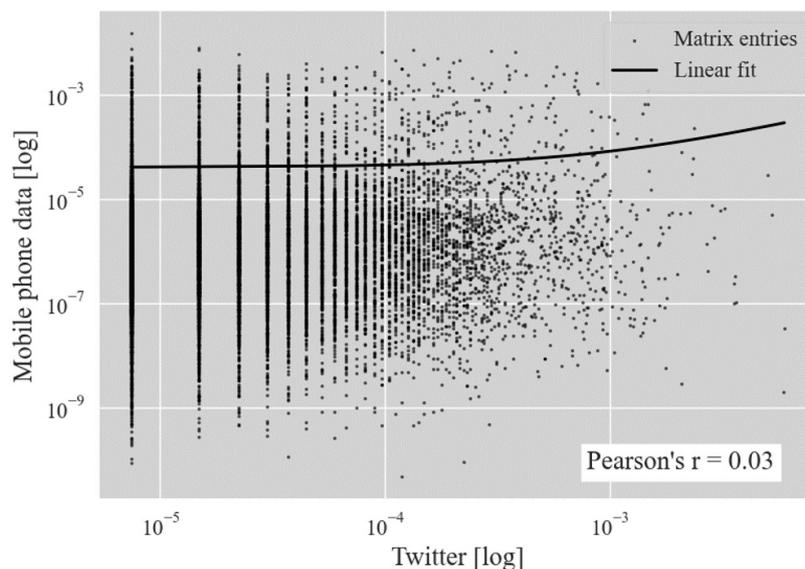


**Figure 10:** Static urban mobility change detection applying Twitter and mobile phone data. Comparison between weekday/weekend (top row) and lockdown style/post lockdown style time periods (bottom row). P-values of applied Mann-Whitney U tests for static urban mobility change detection are listed in Table A. 2.

### 3.3 Limitations

We performed a sensitivity analysis of various window sizes for RWDS. Thereby, we employed a combination of different modeling techniques. This included a dynamic mobility trend analysis and a static mobility change detection. In addition, we considered a set of five distinct mobility metrics. However, our findings show certain limitations, primarily stemming from the choice of a 28-day moving average for trend calculation, a 60-day window synchrony time frame for analyzing dynamic alignment of trend signals, and the temporal selection of on- and offsets for static change detection analysis. Furthermore, our results may be subject to potential biases due to the uneven distribution of Twitter user groups within the overall population (Li et al., 2013; Malik et al., 2015). We did not account for the spatial distribution of inferential uncertainty in our analysis either, although districts with fewer geocoded tweets can be expected to exhibit a higher degree of uncertainty (Huang and Carley, 2019; Huang and Wong, 2015). This particularly affects the graph modularity metrics calculated based on daily OD matrices. The spatial distortion in the applied datasets is supported by the low correlation of non-zero OD matrix entries aggregated over the entire analysis period (cf. Figure 11). Additional results from spatial data exploration, which highlight these issues, are provided in the supplementary GitHub repository (Knoblauch and Gross, 2023).

To address these limitations, several approaches might be applicable: Recent studies on semantic analysis (Hu et al., 2023; Serere et al., 2023) demonstrate promising results in deriving geolocalized information from tweet texts of non-geolocated tweets, which could enhance the Twitter dataset with supplemental geoinformation. Another approach involves utilizing the locations provided in user profiles as a further source of geoinformation. However, it should be noted that these techniques have limited applicability in the context of inner-urban mobility studies (Nguyen et al., 2022).



**Figure 11:** Comparison of temporally aggregated OD matrix entries from Twitter and mobile phone data without considering zero values. Here an OD matrix entry represents a movement between two distinct neighborhoods.

Another aspect of discussion in our long-term validation study pertains to the disparate spatial and temporal resolutions of the employed datasets. Additionally, the raw Twitter data utilized represents less than one percent of the total mobile phone records used in this validation study, leading to a substantial imbalance with potential implications on our validation outcomes (Zhao et al., 2021). Furthermore, certain assumptions were made during the pre-processing stage to facilitate the generation of our validation signal. These assumptions include the selection of lower and upper bounds for IET filtering and the assumption of a uniform distribution of cellular activity in space when converting antenna-based OD matrices into neighborhood-based mobility flows. Additionally, we assumed that the sequential activities of individual users directly represent movements, disregarding the possibility of detours which may introduce a bias in our results. However, we believe that the overall impact of these constraints is relatively minor. We anticipate that conducting supplementary sensitivity analyses on the model parameters would not alter the main findings of this novel long-term validation study, primarily because all parameters and steps were carefully chosen and justified, as described in section 2.

## 4 Conclusion

Our findings demonstrate the effectiveness of employing rolling window down-sampling as a viable strategy to address the limited availability of geolocated tweets in urban areas (cf. Figure 7). Our results indicate that Twitter has the potential to capture short-term changes in mobility at an inner-urban scale (cf. Figure 10), although long-term disparities were observed when compared to mobility metrics derived from mobile phone data in our case study (cf. Figure 9). To enhance the reliability of short-term inference from Twitter data on inner-urban human movement patterns, we propose a combination of multiple analysis techniques, including dynamic and static mobility change detection, simultaneous consideration of various human movement metrics, and sensitivity analysis for modeling parameters. Implementing these approaches can significantly mitigate the risk of false inference in diverse application domains where Twitter is commonly utilized as an open-source proxy for deriving human movement patterns.

Considering the increasingly stringent open-access limitations to Twitter data, this long-term study establishes a foundation for assessing the validity of also upcoming social media platforms. Voluntarily shared geo-social media data can be a powerful and promising tool, especially in locations where other mobility data sources are not openly-accessible or too costly to generate. Since the availability of data sources significantly impacts applications, future research should encompass not only data performance metrics for delineating mobility patterns but also sustainability in terms of long-lasting and openly accessible APIs. Another research option could involve the fusion of data from multiple sources such as Waze, GDELT, Facebook, Instagram, Reddit, Telegram, YouTube, or Weibo. The methods developed in this paper could then be transferred to other geo-social media platforms. Besides that developed methods and generated insights could always be applied with payment plans for API access offered by Twitter.

By conducting this study, our aim was not only to support researchers in

effectively utilizing social media data for modeling human movement patterns but also to gain valuable insights into human mobility within the city of Rio de Janeiro, Brazil. These findings open up new avenues for future research on unexplained mobility-driven phenomena in urban science, such as the location of informal economy (López-García, 2023), accessibility impacts of transport policy (Pereira, 2019), and inner-urban transmission processes of mosquito-borne diseases (Ramadona et al., 2019).

## Data statement

All digitally shareable materials necessary to reproduce the reported methodology have been made available in a public, open-access repository (<https://doi.org/10.5281/zenodo.8305678>).

## Acknowledgements

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the German Research Foundation (DFG) [grant number 451956976].

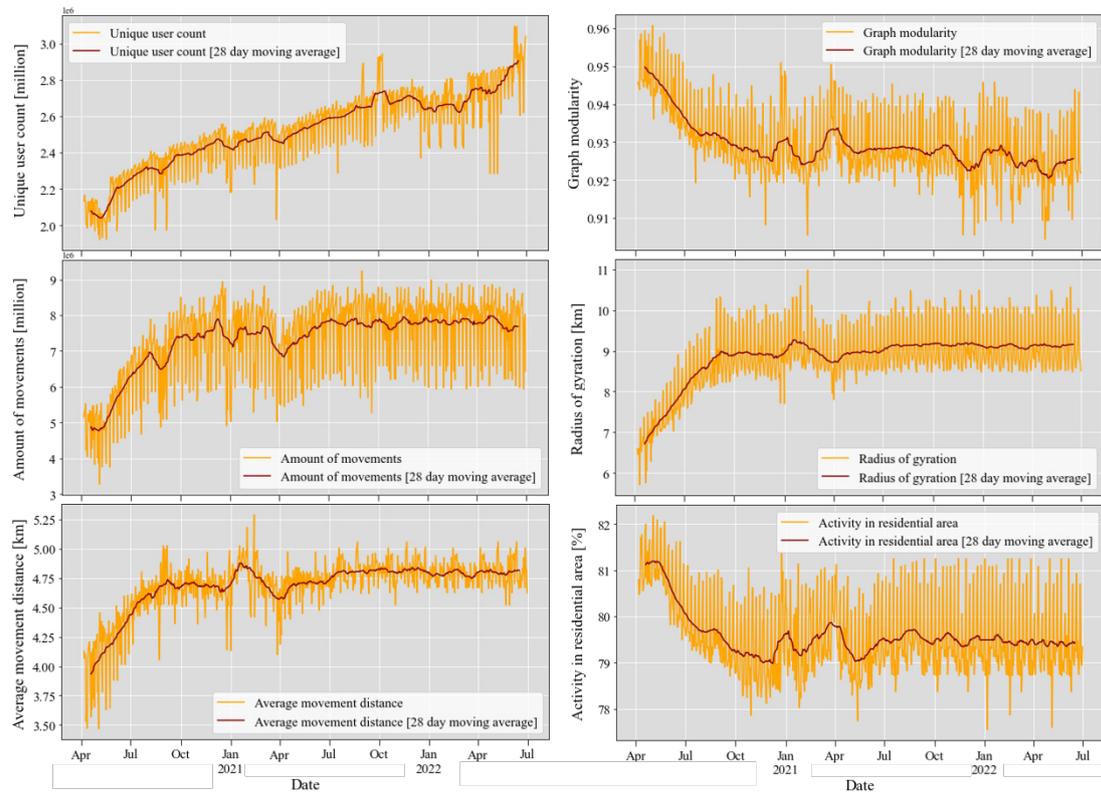
## Declaration of Competing Interest

The authors declare no conflict of interest.

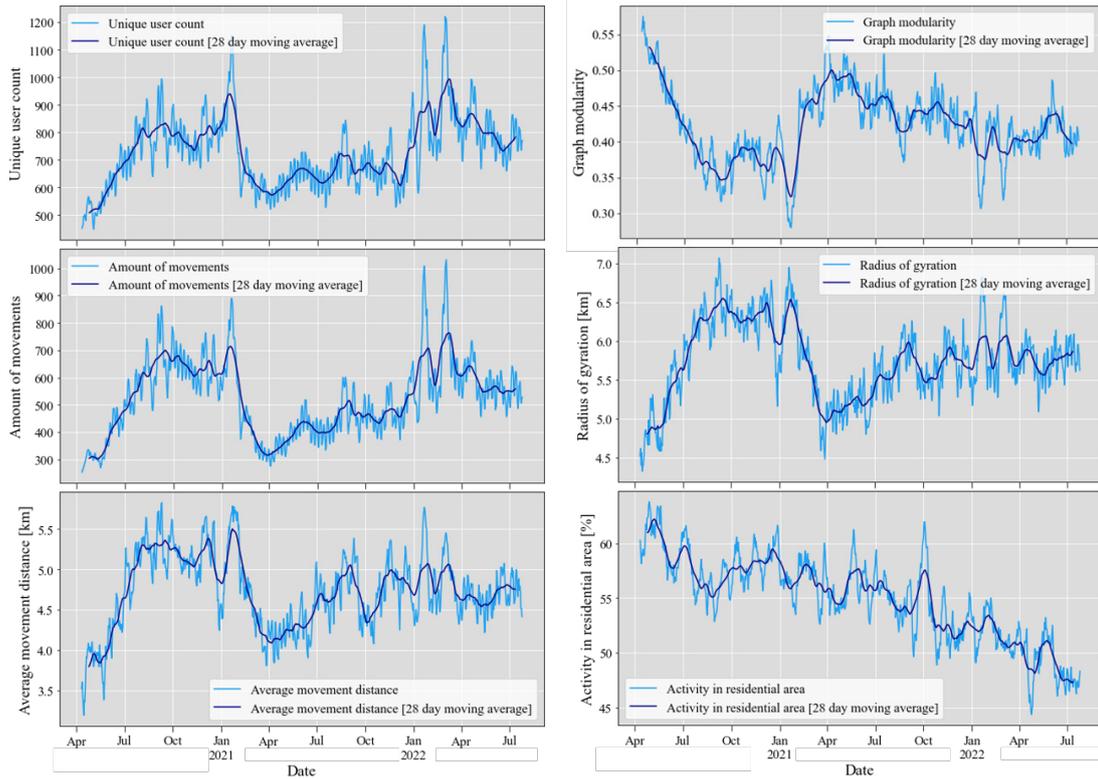
## Credit authorship contribution statement

SK conceptualized the study, conducted the experiment, and drafted the manuscript. SK and SG worked together on the software and both validated the results. AR provided resources. SL, MG, BR, DA, TJ, IM, and AZ reviewed and edited the manuscript. AZ, BR, SL, and TJ supported acquisition of funding. AZ and SL supervised the study. All authors read and approved the manuscript and its subsequent revisions prior to submission.

## Appendix



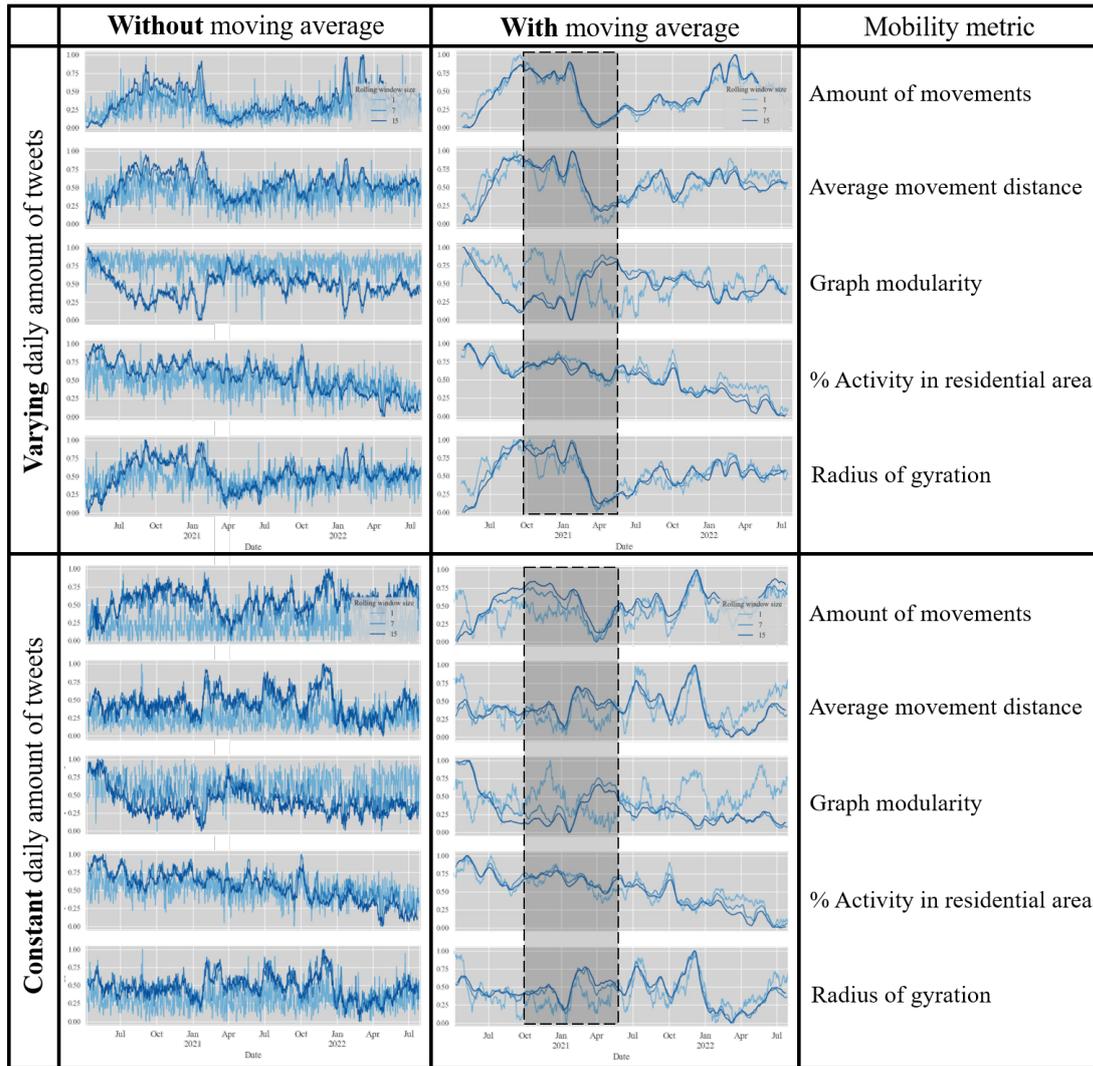
**Figure A. 1:** Not-standardized daily mobility metrics derived from mobile phone data for the city of Rio de Janeiro.



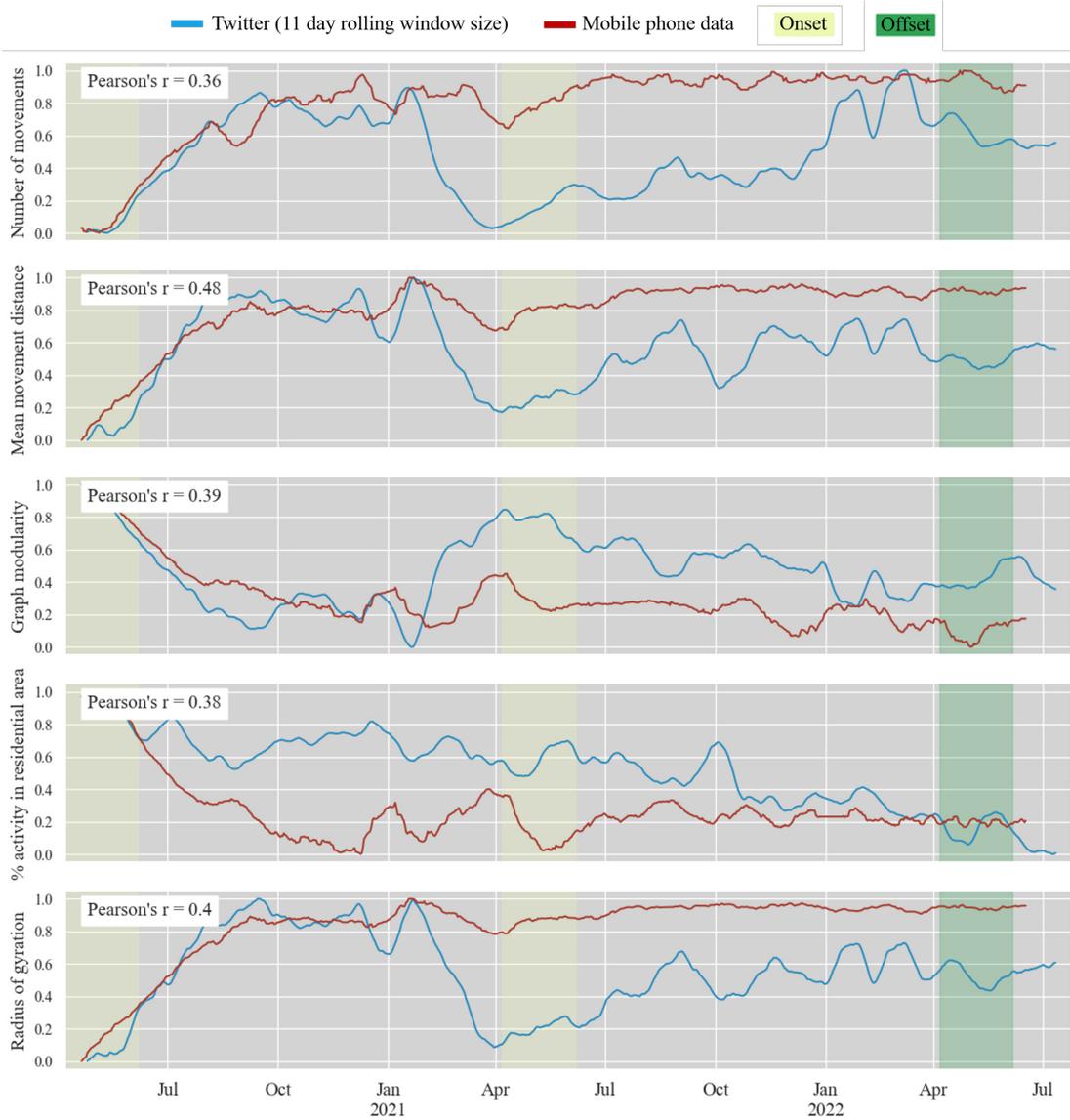
**Figure A. 2:** Not-standardized daily mobility metrics derived from Twitter data for the city of Rio de Janeiro applying a 11-day rolling window size.

**Table A. 1:** Pearson’s correlation coefficients of calculated mobility metric time signals derived from Twitter and mobile phone data applying a moving average of 28 days. Values for 11 day rolling window size resulted in the highest average correlation value over all calculated mobility metrics and are thus highlighted in bold.

	Rolling window size															
	1	3	5	7	9	<b>11</b>	13	15	17	19	21	23	25	27	29	31
Number of movements	0.17	0.29	0.32	0.34	0.35	<b>0.36</b>	0.36	0.36	0.36	0.36	0.36	0.35	0.35	0.34	0.34	0.33
Mean movement distance	0.23	0.37	0.44	0.47	0.48	<b>0.48</b>	0.48	0.47	0.46	0.46	0.45	0.44	0.43	0.42	0.42	0.41
Graph modularity	0.36	0.40	0.41	0.41	0.40	<b>0.39</b>	0.38	0.38	0.37	0.36	0.36	0.35	0.35	0.36	0.36	0.36
% activity in residential area	0.35	0.32	0.35	0.36	0.37	<b>0.38</b>	0.38	0.38	0.38	0.37	0.37	0.37	0.36	0.36	0.36	0.35
Radius of gyration	0.11	0.20	0.30	0.36	0.39	<b>0.40</b>	0.40	0.40	0.39	0.39	0.38	0.38	0.37	0.36	0.35	0.34
Mean over mobility metrics	0.24	0.31	0.36	0.39	0.40	<b>0.41</b>	0.40	0.40	0.39	0.39	0.38	0.38	0.37	0.37	0.36	0.36



**Figure A. 3:** Comparison of mobility metrics calculated with varying and constant amount of tweets per day. Visualization with and without moving average. The constant amount of tweets was calculated using the 98<sup>th</sup> percentile of tweet volume for each window size (one day window: 61 tweets; seven day rolling window: 1115 tweets, 15 day rolling window: 2931 tweets). The grey box highlights the time period of unexpected shifts in calculated mobility metrics for the city of Rio de Janeiro.



**Figure A. 4:** Comparison of daily mobility metrics derived from Twitter and mobile phone data for the city of Rio de Janeiro.

**Table A. 2:** P-values of Mann-Whitney U tests for static urban mobility change detection.

	Twitter				Mobile phone data			
	Weekday-Weekend	Onset1-Onset2	Onset1-Offset1	Onset2-Offset1	Weekday-Weekend	Onset1-Onset2	Onset1-Offset1	Onset2-Offset1
Number of movements	0.96	$4.08e^{-09}$	$3.85e^{-18}$	$7.80e^{-22}$	0.74	$3.09e^{-19}$	$3.07e^{-19}$	$1.01e^{-19}$
Mean movement distance	0.96	$4.36e^{-17}$	$3.85e^{-18}$	$7.80e^{-22}$	0.64	$3.09e^{-19}$	$3.08e^{-19}$	$7.77e^{-22}$
Graph modularity	0.95	$2.22e^{-03}$	$3.85e^{-18}$	$7.80e^{-22}$	0.83	$3.09e^{-19}$	$3.09e^{-19}$	$7.80e^{-22}$
% activity in residential area	0.92	$3.85e^{-18}$	$3.85e^{-18}$	$7.80e^{-22}$	0.71	$3.09e^{-19}$	$3.09e^{-19}$	$5.44e^{-07}$
Radius of gyration	0.92	$1.98e^{-10}$	$3.85e^{-18}$	$7.80e^{-22}$	0.64	$3.09e^{-19}$	$3.08e^{-19}$	$7.77e^{-22}$



# Bibliography

- Abdullah, M., Ali, N., Hussain, S. A., Aslam, A. B., and Javid, M. A. (2021). “Measuring changes in travel behavior pattern due to COVID-19 in a developing country: A case study of Pakistan”. *Transport Policy* 108, pp. 21–33. DOI: 10.1016/j.tranpol.2021.04.023.
- Abdullah, M., Dias, C., Muley, D., and Shahin, M. (2020). “Exploring the impacts of COVID-19 on travel behavior and mode preferences”. *Transportation research interdisciplinary perspectives* 8, p. 100255. DOI: 10.1016/j.trip.2020.100255.
- Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., Kamath, C., Kansal, M., Lange, A., Mandayam, C., Oplinger, A., Pluntke, C., Roessler, T., Schlosberg, A., Shekel, T., Vispute, S., Vu, M., Wellenius, G., Williams, B., and Wilson, R. J. (2020). *Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.1)*. DOI: 10.48550/arXiv.2004.04145.
- Aletta, F., Brinchi, S., Carrese, S., Gemma, A., Guattari, C., Mannini, L., and Patella, S. M. (2020). “Analysing urban traffic volumes and mapping noise emissions in Rome (Italy) in the context of containment measures for the COVID-19 disease”. *Noise Mapping* 7.1, pp. 114–122. DOI: 10.1515/noise-2020-0010.
- Bao, J., Liu, P., Yu, H., and Xu, C. (2017). “Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas”. *Accident; analysis and prevention* 106, pp. 358–369. DOI: 10.1016/j.aap.2017.06.012.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). “Human mobility: Models and applications”. *Physics Reports* 734, pp. 1–74. DOI: 10.1016/j.physrep.2018.01.001.
- Barboza, M. H. C., Alencar, R. d. S., Chaves, J. C., Silva, M. A. H. B., Orrico, R. D., and Evsukoff, A. G. (2021). “Identifying Human Mobility Patterns in the Rio de Janeiro Metropolitan Area using Call Detail Records”. *Transportation Research Record: Journal of the Transportation Research Board* 2675.4, pp. 213–221. DOI: 10.1177/0361198120977655.
- Bisanzio, D., Kraemer, M. U. G., Bogoch, I. I., Brewer, T., Brownstein, J. S., and Reithinger, R. (2020a). “Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale”. *Geospatial health* 15.1. DOI: 10.4081/gh.2020.882.
- Bisanzio, D., Kraemer, M. U. G., Brewer, T., Brownstein, J. S., and Reithinger, R. (2020b). “Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2”. *Journal of travel medicine* 27.5. DOI: 10.1093/jtm/taaa120.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). “Fast unfolding of communities in large networks”. *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008.
- Da Cavalcante Silva, G., Monteiro de Almeida, F., Oliveira, S., Wanner, E. F., Bezerra, L. C., Takahashi, R. H., and Lima, L. (2021). “Comparing community mobility re-

- duction between first and second COVID-19 waves”. *Transport Policy* 112, pp. 114–124. DOI: 10.1016/j.tranpol.2021.08.004.
- Engle, S., Stromme, J., and Zhou, A. (2020). “Staying at Home: Mobility Effects of COVID-19”. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3565703.
- Fabrikant, M. (2017). “A guide to creating a mobility matrix: Cell towers, chiefdoms, and anonymized call detail records”. *Medium*.
- Fatmi, M. R. (2020). “COVID-19 impact on urban mobility”. *Journal of Urban Management* 9.3, pp. 270–275. DOI: 10.1016/j.jum.2020.08.002.
- Gao, S., Rao, J., Kang, Y., Liang, Y., and Kruse, J. (2020a). “Mapping county-level mobility pattern changes in the United States in response to COVID-19”. *SIGSPATIAL Special* 12.1, pp. 16–26. DOI: 10.1145/3404820.3404824.
- Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A. K., Mandujano Reyes, J. F., Yandell, B. S., and Patz, J. A. (2020b). “Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US”. *JAMA network open* 3.9, e2020485. DOI: 10.1001/jamanetworkopen.2020.20485.
- Haas, M. de, Faber, R., and Hamersma, M. (2020). “How COVID-19 and the Dutch ‘intelligent lockdown’ change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands”. *Transportation research interdisciplinary perspectives* 6, p. 100150. DOI: 10.1016/j.trip.2020.100150.
- Hakim, A. J., Victory, K. R., Chevinsky, J. R., Hast, M. A., Weikum, D., Kazazian, L., Mirza, S., Bhatkoti, R., Schmitz, M. M., Lynch, M., and Marston, B. J. (2021). “Mitigation policies, community mobility, and COVID-19 case counts in Australia, Japan, Hong Kong, and Singapore”. *Public health* 194, pp. 238–244. DOI: 10.1016/j.puhe.2021.02.001.
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., and Ratti, C. (2014). “Geo-located Twitter as proxy for global mobility patterns”. *Cartography and geographic information science* 41.3, pp. 260–271. DOI: 10.1080/15230406.2014.890072.
- Heiler, G., Reisch, T., Hurt, J., Forghani, M., Omani, A., Hanbury, A., and Karimipour, F. (2020). “Country-wide Mobility Changes Observed Using Mobile Phone Data During COVID-19 Pandemic”. *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 3123–3132. DOI: 10.1109/BigData50022.2020.9378374.
- Hensher, D. A., Beck, M. J., and Wei, E. (2021). “Working from home and its implications for strategic transport modelling based on the early days of the COVID-19 pandemic”. *Transportation research. Part A, Policy and practice* 148, pp. 64–78. DOI: 10.1016/j.tra.2021.03.027.
- Hernando, A., Mateo, D., Bayer, J., and Barrios, I. (2021). *Radius of Gyration as predictor of COVID-19 deaths trend with three-weeks offset*. DOI: 10.1101/2021.01.30.21250708.
- Hu, X., Hu, Y., Resch, B., and Kersten, J. (2023). “Geographic Information Extraction from Texts (GeoExT)”. *Advances in Information Retrieval*. Ed. by J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo. Vol. 13982. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 398–404. DOI: 10.1007/978-3-031-28241-6{\textunderscore}44.
- Huang, B. and Carley, K. M. (2019). “A large-scale empirical study of geotagging behavior on Twitter”. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Ed. by F. Spezzano, W. Chen, and X. Xiao. New York, NY, USA: ACM, pp. 365–373. DOI: 10.1145/3341161.3342870.

- Huang, Q. and Wong, D. W. S. (2015). “Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data”. *Annals of the Association of American Geographers* 105.6, pp. 1179–1197. DOI: 10.1080/00045608.2015.1081120.
- Huang, X., Li, Z., Jiang, Y., Li, X., and Porter, D. (2020a). “Twitter reveals human mobility dynamics during the COVID-19 pandemic”. *PloS one* 15.11, e0241957. DOI: 10.1371/journal.pone.0241957.
- Huang, X., Li, Z., Jiang, Y., Li, X., and Porter, D. (2020b). “Twitter, human mobility, and COVID-19”. DOI: 10.48550/arXiv.2007.01100.
- Kishore, N., Kiang, M. V., Engø-Monsen, K., Vembar, N., Schroeder, A., Balsari, S., and Buckee, C. O. (2020). “Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis”. *The Lancet. Digital health* 2.11, e622–e628. DOI: 10.1016/S2589-7500(20)30193-X.
- Knoblauch, S. and Gross, S. (2023). *GitHub repository for this manuscript called Long-term validation of inner-urban mobility metrics derived from Twitter*. DOI: 10.5281/zenodo.8304597.
- Kounadi, O. and Resch, B. (2018). “A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data”. *Journal of Empirical Research on Human Research Ethics* 13.3, pp. 203–222. DOI: 10.1177/1556264618759877.
- Kounadi, O., Resch, B., and Petutschnig, A. (2018). “Privacy Threats and Protection Recommendations for the Use of Geosocial Network Data in Research”. *Social Sciences* 7.10, p. 191. DOI: 10.3390/socsci7100191.
- Kurkcü A., Ozbay K., and Morgül E. (2016). “Evaluating the Usability of Geo-Located Twitter as a Tool for Human Activity and Mobility Patterns: A Case Study for New York City”. *Transportation research board’s 95th annual meeting*, pp 1–20.
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frías-Martínez, E., and Ramasco, J. J. (2014). “Cross-checking different sources of mobility information”. *PloS one* 9.8, e105184. DOI: 10.1371/journal.pone.0105184.
- Li, A., Zhao, P., Haitao, H., Mansourian, A., and Axhausen, K. W. (2021). “How did micro-mobility change in response to COVID-19 pandemic? A case study based on spatial-temporal-semantic analytics”. *Computers, environment and urban systems* 90, p. 101703. DOI: 10.1016/j.compenvurbsys.2021.101703.
- Li, L., Goodchild, M. F., and Xu, B. (2013). “Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr”. *Cartography and geographic information science* 40.2, pp. 61–77. DOI: 10.1080/15230406.2013.777139.
- Li, T. (2008). *Upsampling and Downsampling*. Ed. by EE Times.
- Liu, T., Yang, Z., Zhao, Y., Wu, C., Zhou, Z., and Liu, Y. (2018). “Temporal understanding of human mobility: A multi-time scale analysis”. *PloS one* 13.11, e0207697. DOI: 10.1371/journal.pone.0207697.
- López-García, D. (2023). *Worker mobility and urban policy in latin america: Policy interactions and urban Outcomes in mexico city*. 1 Edition. Routledge advances in regional economics, science and policy. New York, NY: Routledge.
- Malik, M., Lamba, H., Nakos, C., and Pfeffer, J. (2015). “Population Bias in Geotagged Tweets”. *Proceedings of the International AAAI Conference on Web and Social Media* 9.4, pp. 18–27. DOI: 10.1609/icwsm.v9i4.14688.
- Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., and Roser, M. (2020). *Coronavirus Pandemic (COVID-19)*. Our World in Data.
- Milusheva, S., Marty, R., Bedoya, G., Williams, S., Resor, E., and Legovini, A. (2021). “Applying machine learning and geolocation techniques to social media data (Twi-

- ter) to develop a resource for urban planning”. *PloS one* 16.2, e0244317. DOI: 10.1371/journal.pone.0244317.
- Municipality of Rio de Janeiro (2022). *Land use land cover (LULC) data*.
- Mützel, C. M. and Scheiner, J. (2022). “Investigating spatio-temporal mobility patterns and changes in metro usage under the impact of COVID-19 using Taipei Metro smart card data”. *Public Transport* 14.2, pp. 343–366. DOI: 10.1007/s12469-021-00280-2.
- Nanda, R. O., Nursetyo, A. A., Ramadona, A. L., Imron, M. A., Fuad, A., Setyawan, A., and Ahmad, R. A. (2022). “Community Mobility and COVID-19 Dynamics in Jakarta, Indonesia”. *International journal of environmental research and public health* 19.11. DOI: 10.3390/ijerph19116671.
- Newman, M. E. J. (2006). “Modularity and community structure in networks”. *Proceedings of the National Academy of Sciences of the United States of America* 103.23, pp. 8577–8582. DOI: 10.1073/pnas.0601602103.
- Nguyen, H. L., Tsolak, D., Karmann, A., Knauff, S., and Kühne, S. (2022). “Efficient and Reliable Geocoding of German Twitter Data to Enable Spatial Data Linkage to Official Statistics and Other Data Sources”. *Frontiers in sociology* 7, p. 910111. DOI: 10.3389/fsoc.2022.910111.
- Osorio-Arjona, J. and García-Palomares, J. C. (2019). “Social media and urban mobility: Using twitter to calculate home-work travel matrices”. *Cities* 89, pp. 268–280. DOI: 10.1016/j.cities.2019.03.006.
- Ossimetha, A., Ossimetha, A., Kosar, C. M., and Rahman, M. (2021). “Socioeconomic Disparities in Community Mobility Reduction and COVID-19 Growth”. *Mayo Clinic proceedings* 96.1, pp. 78–85. DOI: 10.1016/j.mayocp.2020.10.019.
- Paez, A. (2020). “Using Google Community Mobility Reports to investigate the incidence of COVID-19 in the United States”. *Findings*. DOI: 10.32866/001c.12976.
- Pappalardo, L., Simini, F., Barlacchi, G., and Pellungrini, R. (2022). “scikit-mobility : A Python Library for the Analysis, Generation, and Risk Assessment of Mobility Data”. *Journal of Statistical Software* 103.4. DOI: 10.18637/jss.v103.i04.
- Pardo, C. F., Zapata-Bedoya, S., Ramirez-Varela, A., Ramirez-Corrales, D., Espinosa-Oviedo, J.-J., Hidalgo, D., Rojas, N., González-Uribe, C., García, J. D., and Cucunubá, Z. M. (2021). “COVID-19 and public transport: an overview and recommendations applicable to Latin America”. *Infectio* 25.3, p. 182. DOI: 10.22354/in.v25i3.944.
- Park, S., Kim, Y. R., and Ho, C. S. T. (2022). “Analysis of travel mobility under Covid-19: application of network science”. *Journal of Travel & Tourism Marketing* 39.3, pp. 335–352. DOI: 10.1080/10548408.2022.2089954.
- Pereira, R. H. (2019). “Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro”. *Journal of Transport Geography* 74, pp. 321–332. DOI: 10.1016/j.jtrangeo.2018.12.005.
- Petutschnig, A., Albrecht, J., Resch, B., Ramasubramanian, L., and Wright, A. (2022). “Commuter Mobility Patterns in Social Media: Correlating Twitter and LODES Data”. *ISPRS International Journal of Geo-Information* 11.1, p. 15. DOI: 10.3390/ijgi11010015.
- Provenzano, D., Hawelka, B., and Baggio, R. (2018). “The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data”. *Tourism Review* 73.1, pp. 28–43. DOI: 10.1108/TR-03-2017-0052.
- Qian, C., Kats, P., Malinchik, S., Hoffman, M., Kettler, B., Kontokosta, C., and Sobolevsky, S. (2018). “Geo-Tagged Social Media Data as a Proxy for Urban Mobility”. *Advances in Cross-Cultural Decision Making*. Ed. by M. Hoffman. Vol. 610.

- Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, pp. 29–40. DOI: 10.1007/978-3-319-60747-4{\textunderscore}4.
- Ramadona, A. L., Tozan, Y., Lazuardi, L., and Rocklöv, J. (2019). “A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia”. *PLoS neglected tropical diseases* 13.4, e0007298. DOI: 10.1371/journal.pntd.0007298.
- Reynard, D. and Shirgaokar, M. (2019). “Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?” *Transportation Research Part D: Transport and Environment* 77, pp. 449–463. DOI: 10.1016/j.trd.2019.03.002.
- Ruan, S., Bao, J., Liang, Y., Li, R., He, T., Meng, C., Li, Y., Wu, Y., and Zheng, Y. (2020). “Dynamic Public Resource Allocation Based on Human Mobility Prediction”. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.1, pp. 1–22. DOI: 10.1145/3380986.
- Saha, J., Barman, B., and Chouhan, P. (2020). “Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020”. *Children and youth services review* 116, p. 105160. DOI: 10.1016/j.childyouth.2020.105160.
- Saha, J., Mondal, S., and Chouhan, P. (2021). “Spatial-temporal variations in community mobility during lockdown, unlock, and the second wave of COVID-19 in India: A data-based analysis using Google’s community mobility reports”. *Spatial and spatio-temporal epidemiology* 39, p. 100442. DOI: 10.1016/j.sste.2021.100442.
- Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., and Brockmann, D. (2020). “COVID-19 lockdown induces disease-mitigating structural changes in mobility networks”. *Proceedings of the National Academy of Sciences of the United States of America* 117.52, pp. 32883–32890. DOI: 10.1073/pnas.2012326117.
- Serere, H. N., Resch, B., and Havas, C. R. (2023). “Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection”. *PloS one* 18.3, e0282942. DOI: 10.1371/journal.pone.0282942.
- Shumway-Cook, A., Patla, A., Stewart, A. L., Ferrucci, L., Ciol, M. A., and Guralnik, J. M. (2005). “Assessing environmentally determined mobility disability: self-report versus observed community mobility”. *Journal of the American Geriatrics Society* 53.4, pp. 700–704. DOI: 10.1111/j.1532-5415.2005.53222.x.
- Soliman, A., Soltani, K., Yin, J., Padmanabhan, A., and Wang, S. (2017). “Social sensing of urban land use based on analysis of Twitter users’ mobility patterns”. *PloS one* 12.7, e0181657. DOI: 10.1371/journal.pone.0181657.
- Statista (2023). *Number of active Twitter users in selected countries*.
- Steiger, E., Westerholt, R., Resch, B., and Zipf, A. (2015). “Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data”. *Computers, environment and urban systems* 54, pp. 255–265. DOI: 10.1016/j.compenvurbsys.2015.09.007.
- Studenmund, A. H. (2017). *A practical guide to using econometrics*. Seventh edition, global edition. Harlow, England: Pearson.
- Sulyok, M. and Walker, M. (2020). “Community movement and COVID-19: a global study using Google’s Community Mobility Reports”. *Epidemiology and infection* 148, e284. DOI: 10.1017/S0950268820002757.
- Terroso-Saenz, F., Muñoz, A., Arcas, F., and Curado, M. (2022a). “An analysis of twitter as a relevant human mobility proxy: A comparative approach in Spain during the COVID-19 pandemic”. *GeoInformatica*, pp. 1–30. DOI: 10.1007/s10707-021-00460-z.

- Terroso-Saenz, F., Muñoz, A., Arcas, F., and Curado, M. (2022b). “Can Twitter be a Reliable Proxy to Characterize Nation-wide Human Mobility? A Case Study of Spain”. *Social Science Computer Review*, p. 089443932110710. DOI: 10.1177/08944393211071071.
- Tsou, M.-H., Zhang, H., and Jung, C.-T. (2017). *Identifying Data Noises, User Biases, and System Errors in Geo-tagged Twitter Messages (Tweets)*. DOI: 10.48550/arXiv.1712.02433.
- Twitter, Inc. (2023a). *Twitter API for Academic Research*. Ed. by Twitter, Inc.
- Twitter, Inc. (2023b). *Twitter API Documentation*. Ed. by Twitter, Inc.
- Ubaldi, E., Yabe, T., Jones, N. K. W., Khan, M. F., Ukkusuri, S. V., Di Clemente, R., and Strano, E. (2021). “Mobilkit: A Python Toolkit for Urban Resilience and Disaster Risk Management Analytics using High Frequency Human Mobility Data”. DOI: 10.48550/arXiv.2107.14297.
- Wang, A., Zhang, A., Chan, E. H. W., Shi, W., Zhou, X., and Liu, Z. (2021a). “A Review of Human Mobility Research Based on Big Data and Its Implication for Smart City Development”. *ISPRS International Journal of Geo-Information* 10.1, p. 13. DOI: 10.3390/ijgi10010013.
- Wang, P., Lai, J., Huang, Z., Tan, Q., and Lin, T. (2021b). “Estimating Traffic Flow in Large Road Networks Based on Multi-Source Traffic Data”. *IEEE Transactions on Intelligent Transportation Systems* 22.9, pp. 5672–5683. DOI: 10.1109/TITS.2020.2988801.
- Wang, Q. and Taylor, J. E. (2014). “Quantifying human mobility perturbation and resilience in Hurricane Sandy”. *PloS one* 9.11, e112608. DOI: 10.1371/journal.pone.0112608.
- Wang, Y. and Taylor, J. E. (2018). “Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake”. *Natural Hazards* 92.2, pp. 907–925. DOI: 10.1007/s11069-018-3231-1.
- Yildirimoglu, M. and Kim, J. (2018). “Identification of communities in urban mobility networks using multi-layer graphs of network traffic”. *Transportation Research Part C: Emerging Technologies* 89, pp. 254–267. DOI: 10.1016/j.trc.2018.02.015.
- Zhao, Y., Yin, P., Li, Y., He, X., Du, J., Tao, C., Guo, Y., Prospero, M., Veltri, P., Yang, X., Wu, Y., and Bian, J. (2021). “Data and Model Biases in Social Media Analyses: A Case Study of COVID-19 Tweets”. *AMIA ... Annual Symposium proceedings. AMIA Symposium 2021*, pp. 1264–1273.
- Zhao, Z., Shaw, S.-L., Yin, L., Fang, Z., Yang, X., Zhang, F., and Wu, S. (2019). “The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data”. *International Journal of Geographical Information Science* 33.7, pp. 1471–1495. DOI: 10.1080/13658816.2019.1584805.
- Zhu, D., Mishra, S. R., Han, X., and Santo, K. (2020). “Social distancing in Latin America during the COVID-19 pandemic: an analysis using the Stringency Index and Google Community Mobility Reports”. *Journal of travel medicine* 27.8. DOI: 10.1093/jtm/taaa125.
- Zia, M., Fürle, J., Ludwig, C., Lautenbach, S., Gumbrich, S., and Zipf, A. (2022). “SocialMedia2Traffic: Derivation of Traffic Information from Social Media Data”. *ISPRS International Journal of Geo-Information* 11.9, p. 482. DOI: 10.3390/ijgi11090482.

## Publication V:

# Assessing Dengue Risk in Urban Areas: The Role of Daytime *Aedes*-Human Interaction

**Authors.** Steffen Knoblauch, Julian Heidecke; Marcel Reinmuth, Sven Lautenbach, Paulo Filemon Paolucci Pimenta, Joacim Rocklöv, Oliver J. Brady, Antonio Rocha, Thomas Jänisch, Bernd Resch, Filip Biljecki, Annelies Wilder-Smith, Till Bärnighausen, Alexander Zipf

**Outlet.** *The Lancet Planetary Health*

**Status.** Submitted on 2024-07-29 and under review.

### Abstract.

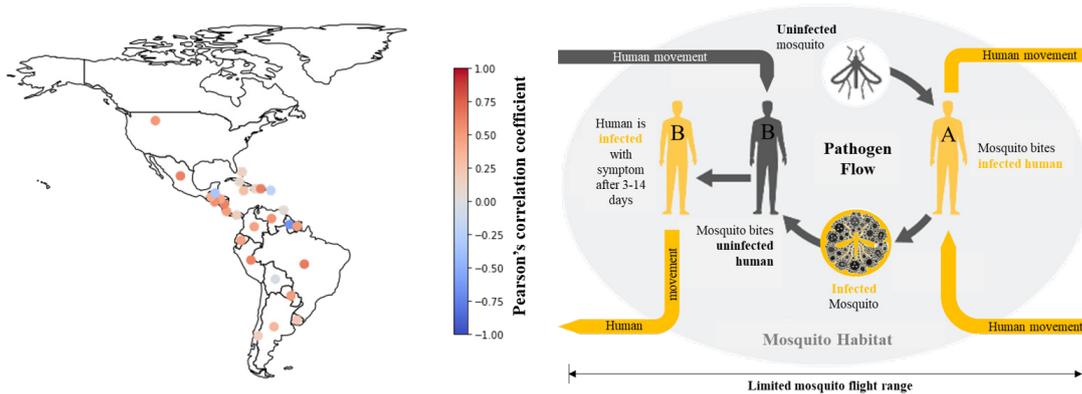
Cities are the hot spots for global dengue transmission. The increasing availability of human movement data obtained from mobile devices presents a substantial opportunity to address this prevailing public health challenge. Leveraging mobile phone data to guide vector control can be relevant for numerous mosquito-borne diseases, where the influence of human commuting patterns impacts not only the dissemination of pathogens but also the daytime exposure to vectors. This study utilizes hourly mobile phone records of approximately 3 million urban residents and daily dengue case counts at the address level, spanning 8 years (2015-2022), to evaluate the importance of modeling human-mosquito interactions at an hourly resolution in elucidating sub-neighborhood dengue occurrence in the municipality of Rio de Janeiro. The findings of this urban study demonstrate that integrating knowledge of *Aedes* biting behavior with human movement patterns can significantly improve inferences on urban dengue occurrence. The inclusion of spatial eigenvectors and vulnerability indicators such as healthcare access, urban centrality measures, and estimates for immunization as predictors, allowed a further fine-tuning of the spatial model. The proposed concept enabled the explanation of 77% of the deviance in sub-neighborhood DENV infections. The transfer of these results to optimize vector control in urban settings bears significant epidemiological implications, presumably leading to lower infection rates of *Aedes*-borne diseases in the future. It highlights how increasingly collected human movement patterns can be utilized to locate zones of potential DENV transmission, identified not only by mosquito abundance but also connectivity to high incidence areas considering *Aedes* peak biting hours. These findings hold particular significance given the ongoing projection of global dengue incidence and urban sprawl.

**Keywords.** Urban dengue transmission · Daytime exposure · *Aedes* biting rates · Human movement · Urban mobility · Spatial eigenvector mapping

## 1 Introduction

The increasing amount of worldwide collected human movement data has a large potential to address current public health challenges (Althouse et al., 2015; Kraemer et al., 2016; Sattenspiel, Lisa and Lloyd, Alun, 2009). Human mobility patterns, derived from a variety of data sources, such as mobile phone networks or social media platforms (Lenormand et al., 2014; Yuan and Raubal, 2012), offer not only the ability to predict the spatial occurrence of infectious diseases

(Finger et al., 2016; Funk et al., 2010; Kogan et al., 2021; Panigutti et al., 2017) but also to assess the effectiveness of control interventions (Chowell and Nishiura, 2014; Cohen et al., 2017). This is of particular interest for many vector-borne diseases, for which labor-intensive vector control remains the most efficacious countermeasure (Hladish et al., 2020; Lobo et al., 2018; Wilson et al., 2020). Among them, the mosquito-borne disease dengue fever is the most important, with a 30-fold increase in incidence over the last 50 years, causing approximately 400 million infections each year (Bhatt et al., 2013).



**Figure 1:** Pearson's correlation coefficients between yearly dengue incidence (Dengue explorer, 2023) and percentage share of population in urban areas (World Bank, 2023) for PAHO (Pan American Health Organization) countries between 1960 and 2021. This analysis explores the potential association between urban growth and the increase in global DENV occurrence, recognizing that correlation does not necessarily imply causation (left). Urban cycle of DENV transmission, highlighting the role of human movement and limited mosquito flight range for disease occurrence (right).

Urban growth, climate change, and international travel are known key drivers for this global incline in the dengue virus (DENV) occurrence (cf. Figure 1) (Messina et al., 2019). DENV transmission dynamics are highly determined by the interplay of mosquito abundance and connectivity, as defined by human movements (Stoddard et al., 2013; Stoddard et al., 2009). A precise understanding of these risk factors, especially their spatial variation and interaction, is essential for an efficient allocation of vector control resources and the prevention of DENV outbreaks world wide (Perkins et al., 2013; Salje et al., 2017; Vanlerberghe et al., 2017). Modeling human-mosquito interactions however can be a challenging task, especially at urban scale, where most DENV infections occur and precise knowledge about mosquito abundance as well as hourly human movement patterns is often missing (Knoblauch et al., 2024b). This research work is part of the 'Lancet Commission on Dengue' and aims to study the phenomenon of urban areas as global hotspots for dengue transmission and prevention.

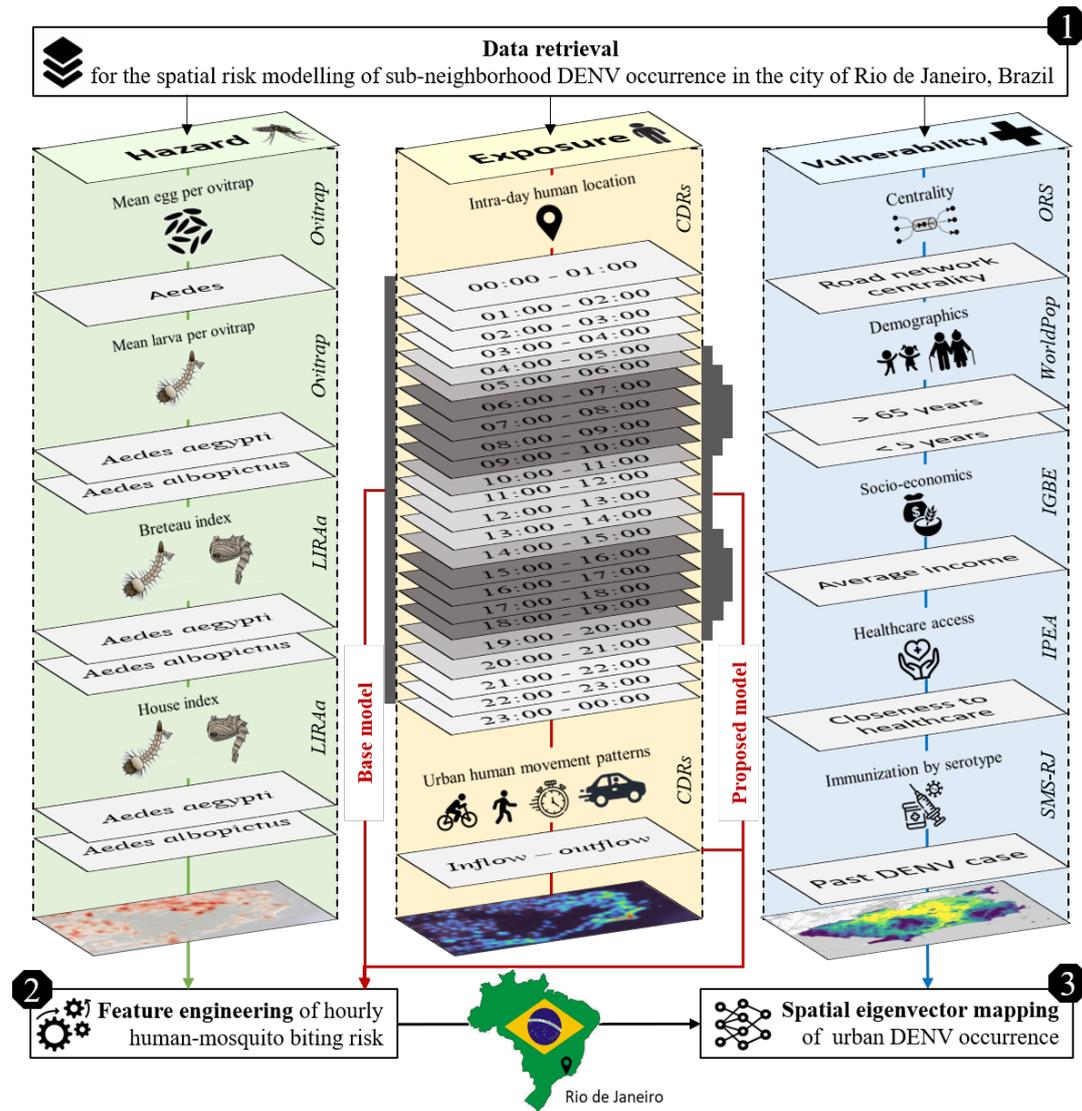
Another challenging aspect of modeling human-mosquito interaction involves considering the ecological characteristics of the vector. DENV is a flavivirus transmitted primarily by female mosquitoes of the species *Ae. aegypti* and *Ae. albopictus* (Ferreira-de-Lima and Lima-Camara, 2018; Simmons et al., 2012). Both mosquito species tend to breed in small, artificial water containers often found in close proximity to human settlements (Banerjee et al., 2015; Getachew et al., 2015; Morrison et al., 2004; Paploski et al., 2016; Trewin et al., 2021; Vezzani,

2007; World Health Organization, 2012). Additionally, they exhibit a limited flight range, which is estimated to be below 1000 m (Bomfim et al., 2020; Harrington et al., 2005; Honório et al., 2003; Massad et al., 2017), and a diurnal biting behavior that mainly covers evening and morning twilights. Incorporating these ecological vector characteristics into the modeling of urban dengue outbreaks is imperative in the pursuit of alleviating the global burden of dengue fever (Kraemer et al., 2018).

In this study, our primary objective is to assess the importance of ecological vector characteristics in explaining the spatial distribution of urban dengue infections. We propose that incorporating human-mosquito interactions on an hourly basis, while considering the diurnal biting behavior of mosquitoes and daytime commuting patterns of humans at a sub-neighborhood scale, may significantly impact our understanding of inner-urban dengue dynamics. To achieve this, we analyze the sensitivity of inferences related to various assumptions about hourly human-mosquito interactions. We develop two distinct modeling scenarios: one that neglects existing knowledge about *Aedes* mosquito twilight activity, and another that incorporates this knowledge through feature engineering, allowing for a more comprehensive analysis of the intricate dynamics of urban dengue infections. A low sensitivity to these inferences would suggest that ecological vector characteristics play a minor role in urban dengue outbreaks. Conversely, a high sensitivity would underscore the need to carefully account for diurnal biting behavior and daytime human movements when modeling DENV infections, especially at a fine-grained urban scale. In order to carry out this investigation we integrate data from previous research on high-resolution urban mosquito mapping (Knoblauch et al., 2023) and inner-urban human mobility patterns (Knoblauch et al., 2024a), thereby creating hourly transmission risk maps. Our study focuses on the municipality of Rio de Janeiro, Brazil, an urban area endemic for *Aedes* mosquitoes and experiencing numerous dengue cases annually (Secretaria Municipal de Saude Rio de Janeiro, 2022). The findings from this research could significantly enhance our understanding of urban dengue transmission dynamics and potentially contribute to the development of more effective control strategies for this disease. More specifically, our investigation focuses on evaluating the impact of two key factors on model enhancement: (i) the feature engineering of hourly human-mosquito biting risk and (ii) the incorporation of spatial eigenvector mapping and vulnerability indicators.

## 2 Materials and Methods

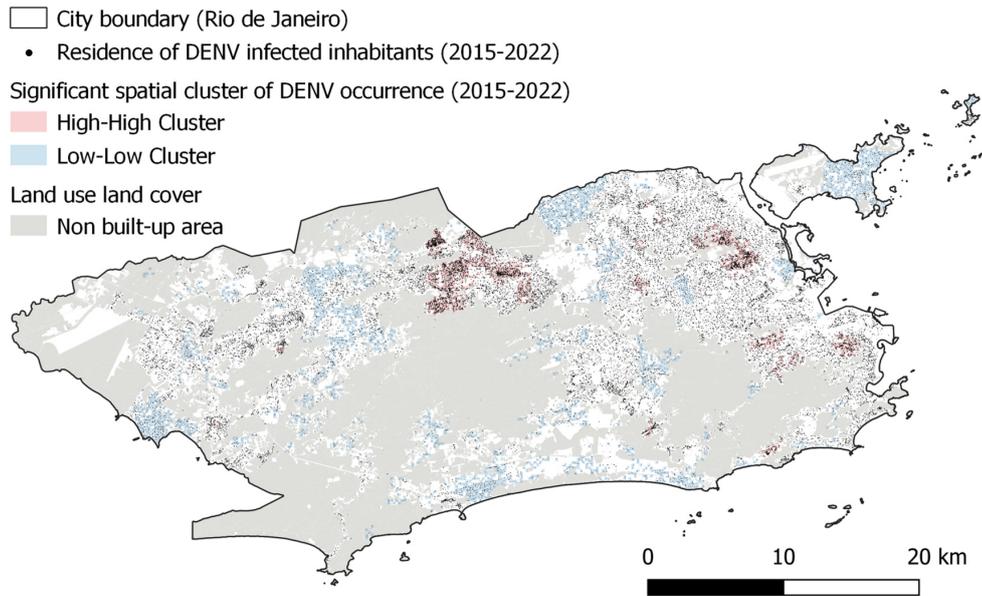
Here, we propose a novel risk modeling framework integrating ecological characteristics of *Ae. aegypti* and *Ae. albopictus* with data-driven insights on inner-urban human movement flows. This framework consists of three main parts (cf. Figure 2): i) the retrieval of DENV-related proxies capturing the three risk components of hazard, exposure, and vulnerability, ii) the daytime feature engineering of human-mosquito biting risk integrating human movement data with knowledge of *Aedes* biting behavior, and iii) the spatial eigenvector mapping (Griffith, 2019) of urban DENV occurrence using vulnerability indicators.



**Figure 2:** Workflow for the sub-neighborhood spatial eigenvector mapping of urban DENV occurrence applying entomological surveillance (left) and call detail records (middle) to model daytime human-mosquito biting risk for the municipality of Rio de Janeiro in Brazil on an hourly basis. Voronoi tessellations based on mobile phone antenna locations were employed as the spatial unit for analysis. In the feature engineering process, the base model assumed a constant human-mosquito interaction throughout the day, while the proposed model accounted for the fluctuating exposure of humans to mosquito bites, considering the twilight biting activity of *Aedes* mosquitoes and the hourly commuting patterns of humans. (CDRs: Call detail records; ORS: OpenRouteService; IGBE: Brazilian Institute of Geography and Statistics; IPEA: Institute of Applied Economic Research; SMS-RJ: Municipal Health Ministry of Rio de Janeiro).

## 2.1 Data

All employed datasets, including their sources, spatial resolutions, and pre-processing procedures, are listed in Table A. 1. In order to evaluate our approach, we acquired daily counts of DENV cases from 2015 to 2022 with geographical coordinates corresponding to residential addresses (cf. Figure 3). In adherence to ethical considerations and following approval by the Research Ethics Committee (CEP), this dataset underwent anonymization and was made accessible upon formal request by the municipal health ministry of Rio de Janeiro.



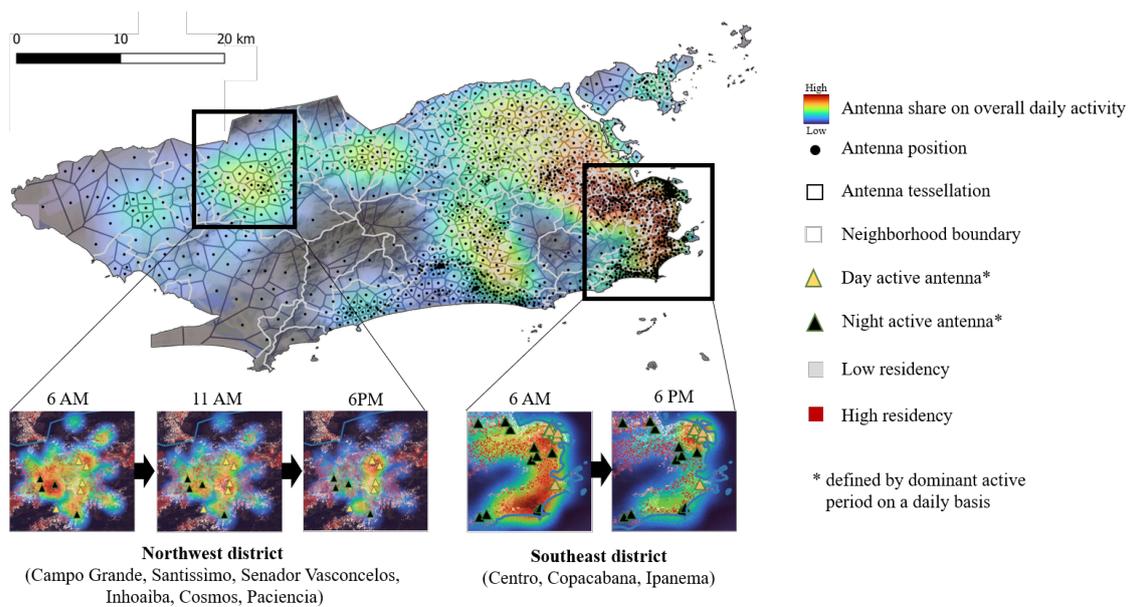
**Figure 3:** A 200 m grid displaying statistically significant hotspots, cold spots, and spatial outliers derived from daily DENV health records collected for the municipality of Rio de Janeiro between January 2015 and December 2022. Spatial autocorrelation and the identification of clusters with similar or dissimilar values were assessed using the Anselin Local Moran’s I statistic. In this context, ‘High-High’ clusters represent areas with high DENV occurrence surrounded by neighboring areas with similarly high occurrence, and ‘Low-Low’ clusters indicate areas with low occurrence surrounded by low-occurrence neighbors. Areas colored white indicate the absence of significant spatial autocorrelation in dengue occurrence.

### 2.1.1 Hazard

The hazard risk components in this study were modeled using a variety of entomological surveillance data sources, focusing on different stages of the mosquito lifecycle. These data included mosquito egg and larva counts (cf. Figure A. 1) as well as indices representing larva and pupa infestation (cf. Figure A. 2). To model the abundance of immature *Ae. aegypti* and *Ae. albopictus* in urban areas, we used egg and larval counts from the year 2019, collected from 2698 ovitraps distributed across the study area. However, since ovitrap counts have limited spatial validity due to heterogeneous urban landscapes and the restricted flight range of *Aedes* mosquitoes, a previous study enhanced this dataset by incorporating high-resolution urban suitability indicators (Knoblauch et al., 2023). This approach allowed for a continuous approximation of seasonal urban mosquito suitability, considering a limited flight range of 200 m. For modeling larvae and pupae infestation, data from the Larval Infestation Rapid Assay (LIRA) were applied. These datasets covering the years 2015 to 2022 included seasonal Breteau and House indices, which were collected for 256 homogeneous street blocks defined by entomologists during survey design. The House index (HI) gauged the number of infested houses relative to the total visited buildings during LIRA, while the Breteau index (BI) represented the number of positive containers per 100 houses inspected. All entomological surveillance datasets used in this research were obtained from the municipal health department of the municipality of Rio de Janeiro upon request, exclusively for the purpose of this study.

### 2.1.2 Exposure

To model the exposure components, we obtained hourly origin-destination (OD) matrices and corresponding population density maps using anonymized call detail records provided by a large Brazilian telecommunications company. This raw dataset comprised individual antenna connections from approximately three million individual users, representing an estimated 45 percent of the population of the municipality of Rio de Janeiro. The raw data had a temporal resolution of five minutes, capturing user connections to 1 359 antennas distributed across 163 neighborhoods. In this study, the collective OD matrices for Voronoi tessellations, delineated by the locations of antennas, were generated based on the temporal sequences of individual antenna connections spanning from July 2021 to July 2022, encompassing a complete annual cycle of human mobility patterns. A more extensive description of the applied methods was given by a previous study (Knoblauch et al., 2024a). Figure 4 illustrates the fluctuations in human population density throughout the day due to commuting dynamics within the municipality of Rio de Janeiro, where day and night active antennas are defined by their dominant active periods on a daily basis.



**Figure 4:** Daytime human population density in the municipality of Rio de Janeiro, estimated by using mobile phone data. Hourly changes in antenna activity behave differently in various zones of the case study region, as shown for two selected subregions. While the dominant mobility motif in the northwest district involved movement between three locations, the southeast district exhibited a dominant mobility motif characterized by movement between two locations.

### 2.1.3 Vulnerability

We hypothesized that the likelihood of an infected individual appearing in official health registries is influenced not only by the human-mosquito biting risk but also by other factors, which were defined as vulnerability indicators and considered to refine the precision of estimating the spatial distribution of dengue cases, especially within an urban setting. The utilized indicators can be classified into five subgroups: centrality, accessibility, socio-economics, demographics, and level

of immunization. Centrality indicators were derived from OpenStreetMap (OSM) using the OpenRouteService API (HeiGIT gGmbH, 2023). Mean road centrality by average travel time was calculated on a 200m grid using cars as a transportation medium (cf. Figure A. 3). These measures were hypothesized to estimate human closeness and interaction, as tested in a previous study (Mahabir et al., 2012). Accessibility to job opportunities and travel time to the closest healthcare facilities using active transportation as well as public schools were retrieved from the Institute for Applied Economic Research (IPEA) on a hexagonal grid of 0.11  $km^2$  (cf. Figure A. 4) (Pereira and Herszenhut, 2023). The same data source was applied to download a cumulative opportunity measure of the whole population, indicating the number of opportunities that can be reached within 60 minutes of travel time, and the socio-economic indicator of average household income per capita. We hypothesized that all these accessibility and socio-economic indicators influence the appearance of dengue infections in the official health database. For instance, we assumed that inhabitants of favelas in Rio de Janeiro with lower average income and lower accessibility measure are less likely to visit a doctor with dengue symptoms compared to people of higher social class (Lai et al., 2020; McMaughan et al., 2020).

Information about the most vulnerable age groups for DENV infections was included by using population estimates for children below five years and elderly individuals above 60 years from the Humanitarian Data Exchange (Humanitarian Data Exchange, 2019). These indicators were included to estimate the severity of symptoms (Yang et al., 2023) and, thus also the likelihood of visiting a doctor and being registered in an official health database (cf. Figure A. 5). Additionally, we retrieved the locations of past DENV infections, including all four DENV serotypes, hypothesizing that past dengue epidemics serve as a reliable indicator for modeling immunization levels at the population level (Thai et al., 2011). However, this immunological vulnerability effect would likely be complex as it is dependent on the sequence of DENV serotypes causing DENV cases over time as well as the time intervals between them (Katzelnick et al., 2015). All mentioned vulnerability components were combined with the daytime models for human-mosquito biting risk to facilitate spatial eigenvector mapping of urban DENV occurrence, which is described in the following subsection.

## 2.2 Methods

### 2.2.1 Feature engineering of hourly human-mosquito biting risk

In this study, we propose a novel method to model the spatial distribution of human-mosquito biting risk in urban areas by incorporating ecological characteristics of mosquitoes, specifically focusing on the diurnal biting behavior of *Ae. aegypti* and *Ae. albopictus*. Our method involved integrating local estimates of mosquito abundance  $M_i$  with knowledge about hourly human distribution over the considered area to derive an aggregate measure of mosquito biting risk  $B_i$  for residents of cell  $C_i$ :

$$B_i = \left[ \sum_{h=1}^{24} \omega(h) \left( \sum_{j=1}^N \chi_{i,j}(h) M_j \right) \right] \quad (1)$$

Equation 1 aims to more accurately reflect mosquito bite risk than estimates solely based on local mosquito abundance by incorporating two key principles. First, due to human movement, individual hosts are exposed to different mosquito populations throughout the day. To capture this for each hour of the day  $h$ , we calculated a weighted sum approximating the contribution of mosquito populations  $M_j$  from all cells  $C_j$  to the biting risk of people resident in cell  $C_i$ . This sum reflects the extent to which the hourly mosquito biting risk originating from the mosquito population  $M_j$  in cell  $C_j$  affects individuals residing in cell  $C_i$ . To this end, we estimated  $\chi_{i,j}(h)$ , representing the fraction of people present in antenna tessellation cell  $C_j$  during hour  $h$ , relative to the total number of residents in antenna tessellation cell  $C_i$ . The calculation of  $\chi_{i,j}(h)$  utilized hourly OD matrices, indicating collective human mobility from cell  $C_i$  to cell  $C_j$ . Secondly, considering the daytime variation in mosquito biting behavior, we introduced the hourly weighting function, denoted as  $w(h)$  in our model (cf. Equation 2). It is well-documented that *Ae. aegypti* and *Ae. albopictus* biting behavior occurs exclusively during daylight hours, with heightened activity observed during twilight (Muhammad et al., 2020; Mutebi et al., 2022; Zahid et al., 2023). As such, we assumed a decrease in mosquito biting activity during midday hours. However, we posited that this behavior might persist in shaded regions characterized by elevated humidity and other environmental factors favoring mosquito activity (Baik et al., 2020; Egid et al., 2022; Wei et al., 2023). Notably, mosquito biting activity during the night was excluded from our proposed model.

$$\omega(h) = \begin{cases} 3, & \text{if } h \in \{6, 7, 8, 9, 15, 16, 17, 18\} \\ 2, & \text{if } h \in \{5, 10, 14, 19\} \\ 1, & \text{if } h \in \{4, 11, 12, 13, 20\} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The proposed feature engineering underwent evaluation employing a quasi-Poisson generalized linear model (QP-GLM), wherein the target variable  $D_i$  was defined by overdispersed official dengue case counts aggregated on 1 359 antenna tessellations between the years 2015 and 2022 (cf. Equation 3). For evaluation, we calculated Cohen's pseudo- $R^2$  (cf. Equation 4). The explained deviance for this regression model was compared to the pseudo- $R^2$  of a base model that did not consider assumptions related to diurnal *Aedes* mosquito biting behavior and hourly human movement (cf. Figure 2). In contrast to the proposed model, the base model was implemented utilizing identity OD matrices for  $\chi_{i,j}(h)$ .

$$\begin{aligned} D_i &\sim \text{quasi-Poisson}(\hat{\mu}_i, \hat{\theta}) \\ \mathbb{E}(D_i) &= \hat{\mu}_i \\ \text{Var}(D_i) &= \hat{\mu}_i * \hat{\theta}, \text{ with } \hat{\theta} \neq 1 \\ \log(\hat{\mu}_i) &= \log(H_i) + \hat{\beta}_0 + \hat{\beta}_1 * B_i \end{aligned} \quad (3)$$

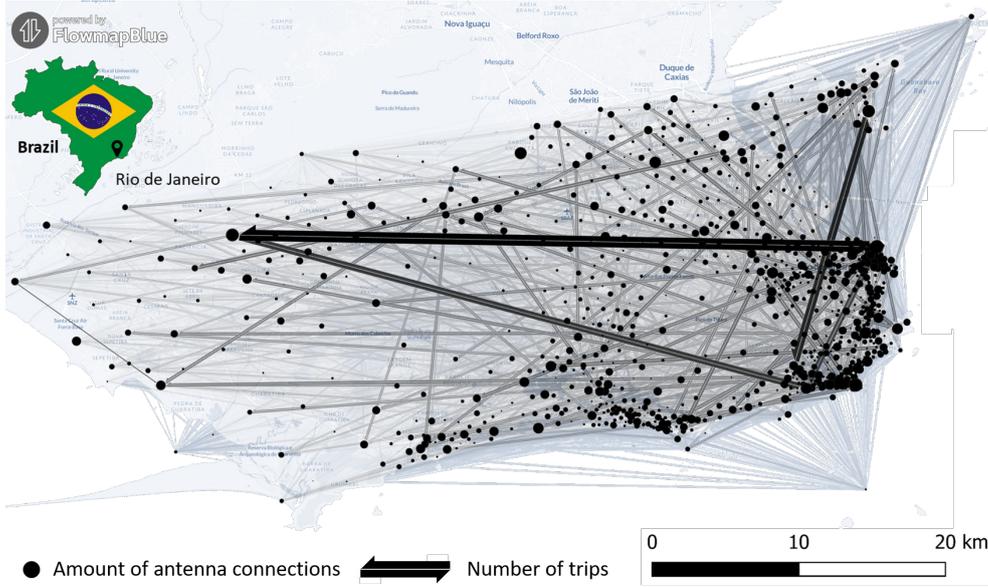
$$\text{Cohen's pseudo } R^2 = 1 - \frac{\text{model deviance}}{\text{null model deviance}} \quad (4)$$

### 2.2.2 Spatial eigenvector mapping incorporating selected vulnerability indicators

After evaluating the feature engineering of hourly human-mosquito biting risk, we expanded our QP-GLM in two aspects: (i) by incorporating vulnerability indicators to model the likelihood of an infected individual being registered in official health registries, geolocated by residency, and (ii) by integrating spatial eigenvectors to address spatial autocorrelation of residuals. To mitigate multicollinearity among covariates, we selected vulnerability indicators with low intercorrelation ( $\leq 0.7$ ). These two model extensions led to a more comprehensive model for sub-neighborhood DENV occurrence, considering daytime human-mosquito biting risk, as explored in our initial research question.

Our second research objective focuses on the enhancement of spatial estimates of sub-neighborhood DENV occurrence by incorporating spatial eigenvectors and selected vulnerability indicators. By addressing this question, we aim to assess the extent to which these additional variables improve the predictive capability and understanding of DENV transmission dynamics within the urban environment. Here, vulnerability features were defined as variables that influence the appearance and collection process of DENV infections at the urban scale, but not the human-mosquito biting risk itself. This strategic inclusion allows us to dissect the nuanced factors contributing to DENV occurrence, beyond solely focusing on the dynamics of human-mosquito interactions. In this study, these factors included the location of vulnerable age groups, accessibility to health care facilities, road network centrality, the socio-economic factor of average income, and estimates on immunization levels derived from past DENV infections. In contrast to the first model defined in Equation 3, the year 2022 was selected as the reference year for predictions, coinciding with the occurrence of the last major DENV outbreak in the municipality of Rio de Janeiro (cf. Figure B. 1). Consequently, immunization levels were estimated based on the spatial distribution of past infections recorded between 2015 and 2021.

The applied spatial eigenvector mapping, originally proposed by Griffith et al. (2019) (Griffith, 2019), involved the incorporation of additional covariates to absorb spatial autocorrelation. This ensures unbiased estimators for other predictors. These covariates, derived from the eigenfunction decomposition of the spatial weight matrix  $W$ , are called spatial eigenvectors. They represent orthogonal components that effectively separate and capture information on spatial autocorrelation, similar to principal component analysis. In our study, we employed daily aggregated OD matrices from July 2021 to July 2022 to illustrate human connectivity between antenna tessellations, serving as a spatial weight matrix (cf. Figure 5). This led to the generation of 1 359 spatial eigenvectors, out of which the ME function from the 'spatialreg' R package facilitated the identification of a specific subset applying brute-force search (Bivand, 2023; Griffith, 2000) under consideration of an alpha threshold of 0.05 to mitigate residual autocorrelation. This selected subset of eigenvectors was integrated as additional covariates into the QP-GLM (cf. Equation 3).



**Figure 5:** Human movement patterns used for spatial eigenvector mapping of DENV occurrence in the municipality of Rio de Janeiro. Spatial weights were estimated applying mobile phone records from July 2021 to July 2022. Thick dark black edges represent high human connectivity between antenna locations, whereas thin and bright black stripes indicate a lower amount of human movements.

### 3 Results

#### 3.1 Evaluating the feature engineering of hourly human-mosquito biting risk

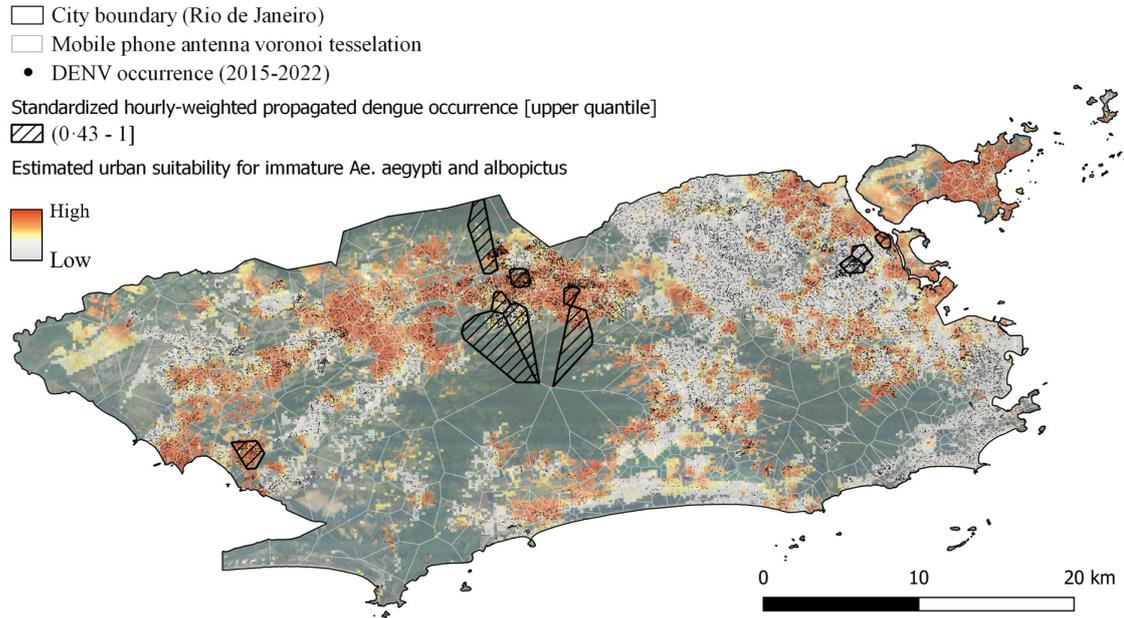
The results in Table 1 demonstrate how considering mosquito biting hours and human movement corridors can enhance the accuracy of spatial estimates for urban DENV occurrence. The proposed feature engineering method outperforms the baseline model, which does not consider the daylight activity of *Aedes* mosquitoes, and demonstrates a 13.5% increase in the explained deviance within the response of the QP-GLM. Both models yielded positive and significant estimates for their hazard and exposure combined covariate of human-mosquito biting risk  $B_i$ . The computed global Moran's I value for the residuals was 0.59.

**Table 1:** Coefficients, standard errors, and p-values for QP-GLMs applying two different model scenarios for human-mosquito interaction, where the base model did not consider any temporal variation in human-mosquito biting risk and the proposed model incorporated hourly adapting mosquito activity and human population densities. Regression coefficients and standard errors are reported at the link scale. The limited explained deviance in both models hints at the presence of missing latent covariates.

QP-GLM	Intercept		Human-mosquito biting risk ( $B_i$ )			Cohen's explained deviance pseudo- $R^2$
	$\hat{\beta}_0$	$\hat{\sigma}_{\hat{\beta}_0}$	$\hat{\beta}_1$	$\hat{\sigma}_{\hat{\beta}_1}$	$Pr(>  z )$	
Base model (BM)	-0.7914	0.2856	2.3228	0.4591	$4.78e^{-7}$	0.0395
Proposed model (PM)	0.4273	0.0532	4.2866	0.2230	$< 2e^{-16}$	0.1750

Considering the aforementioned results, it implies that integrating knowledge of *Aedes* biting behavior with human movement patterns can also facilitate the inference of probable transmission sites for reported dengue cases. If this holds

true, increased mosquito control interventions in these locations would have the potential to combat *Aedes*-borne diseases more effectively.



**Figure 6:** Novel vector control planning map considering daytime mosquito activity and human movement flows for the municipality of Rio de Janeiro. The figure illustrates the discrepancy between DENV occurrence and estimated mosquito abundance at an urban scale. Areas of dark red color represent target effectiveness zones measured by entomological surveillance. The black-striped Voronoi tessellations highlight potential danger areas for transmission that might be underestimated when relying solely on entomological surveillance. The identification of these zones relied on hourly-weighted propagated dengue occurrence  $HP - DENV_i$ , weighted by biting activity, to reflect the locations of infected persons during the days denoted as  $HP - DENV_i = \sum_{h=1}^{24} w(h) \cdot (DENV \cdot \prod_{j=1}^h OD_j)$ . Within the black-striped Voronoi tessellations, sub-regions with high mosquito suitability are particularly relevant to guide interventions.

Figure 6 presents, as a highlight of this work, the practical implications of these research findings for the municipality of Rio de Janeiro. A novel mapping approach for vector control intervention was developed, incorporating (i) the spatial distribution of mosquitoes, as indicated by temporally aggregated entomological surveillance data, (ii) the spatial dispersal of dengue occurrence, and (iii) the most likely transmission locations for reported dengue cases, taking into account daytime *Aedes* biting behavior. This target effectiveness map marks regions that were potentially underestimated for vector control planning using entomological datasets only, while at the same time emphasizing the enduring importance of areas with high mosquito abundance.

### 3.2 The role of vulnerability indicators and spatial eigenvector mapping in model enhancement

We hypothesized that incorporating vulnerability indicators and spatial eigenvectors would further enhance the proposed QP-GLM (cf. Equation 3), which considers *Aedes*-human interactions for predicting the spatial occurrence of dengue in the municipality of Rio de Janeiro. The Cohen's pseudo-R<sup>2</sup> of the

more extensive QP-GLM considering hourly human-mosquito biting risk was determined to be 0.77, indicating that the extended model was capable of explaining up to 77 percent of the deviance in dengue occurrence on the sub-neighborhood level for the municipality of Rio de Janeiro in the year 2022. The computed global Moran's I value for the residuals was 0.07, indicating low spatial autocorrelation. A QP-GLM with human-mosquito biting risk and vulnerability indicators but without spatial eigenvectors was not considered, as it yielded a higher overdispersion value of 26.85 and a higher global Moran's I of 0.2, despite having a Cohen's pseudo-R<sup>2</sup> of 0.83. This underscores the importance of vulnerability indicators and spatial eigenvector mapping in improving spatial predictions of sub-neighborhood dengue occurrences, which are georeferenced based on residency. Additional result on the applied vulnerability indicators and spatial eigenvectors are listed in the Appendix B (Table B. 1, Figure B. 2).

## 4 Discussion

In this study, we analyzed the impact of modeling urban DENV occurrence under consideration of daytime mosquito activity and human movement patterns. The analysis has shown how urban areas exhibit spatial heterogeneity in numerous factors relevant to infectious disease transmission. The findings contribute to the understanding of infectious disease dynamics at a sub-neighborhood scale by highlighting the important role played by daytime mosquito activity and human movement flows in linking observed patterns of DENV incidence to inferred patterns of disease transmission. The inferred degree of spatial variation in urban DENV occurrence was sensitive to assumptions about daytime mosquito activity. Spatial discrepancy existed between the dominant location of mosquitoes, the spatial patterns of human-mosquito interaction points, and disease occurrence collected by residency. Taking these findings into account, one can conclude that methodologies that presume consistent human exposure to mosquito bites throughout the day potentially yield exaggerated and biologically inadequate interpretations regarding the patterns of disease transmission.

### 4.1 Challenges and opportunities

Additional knowledge about pathogen penetration rates in host and vector populations would potentially enhance prediction capabilities for urban DENV occurrence. However, the practical challenges associated with establishing such virus penetration measurements, e.g. within entomological surveillance systems, pose significant obstacles, especially due to the need for appropriate laboratory infrastructure. The utilization of mobile phone data as a proxy for human movement in the present study could have resulted in additional inherent constraints. Despite the high penetration rate of the mobile phone provider, mobility estimates may have been biased due to the exclusion of individuals without mobile phones or those using different services. To counteract this factor, an improvement strategy could involve integrating social media streams. Higher-order descriptions of movement, such as social network structure, have been shown to affect transmission dynamics in urban environments (Reiner et al., 2014; Stoddard et al., 2013; Vazquez-Prokopec et al., 2013). The consideration of the interplay among disease symptoms, infectiousness, and the mobility of individuals infected

with DENV seems additionally promising in this context (Perkins et al., 2015; Perkins et al., 2016; Schaber et al., 2019). This complicates the assumption that the movement patterns of apparently healthy individuals can adequately represent the mobility patterns of those involved in transmission (Wesolowski et al., 2016).

## 4.2 Future directions

Follow-up activities could connect by examining cross-boundary human movements to out-of-city regions and potentially model on individual human scale instead of aggregating risk value into areal units. The fixed scenario employed to model the daytime risk of human-mosquito biting could also be extended to consider seasonal fluctuations in sunset and sunrise times. We showed that approximating resident's exposure to different mosquito populations throughout the city, leveraging human movement flows derived from mobile phone data, results in higher predictive power than focal mosquito abundance alone. This highlights the importance of considering mechanisms driving human-mosquito interactions for understanding of mosquito-borne disease occurrence at urban scale. We employed a relatively basic and temporally static statistical model. Future studies could leverage similar data for building spatial process-based models of intra-city transmission dynamics (Kache et al., 2022; Wu et al., 2023). Such models, although computationally and conceptually challenging, could effectively incorporate mosquito behavior as well as ecology and capture feedback processes, such as immunity dynamics and transmission cycles, amongst others. Furthermore, daytime variation of human host density across the city could be incorporated, potentially impacting local vectorial capacity by modulating mosquito biting behavior and mosquito-to-host ratio. Implementing this modeling approach would yield additional insights on the efficacy of prevention and control strategies, thereby enhancing our understanding and management of mosquito-borne diseases in urban environments.

## 4.3 Conclusion

We reported that modeling hourly human-mosquito interaction, integrating knowledge of *Aedes* biting behavior with human commuting patterns, can improve inference on dengue occurrence in the city of Rio de Janeiro by 13.5% compared to a model neglecting daytime variation in biting activity and human movement. Additionally, the inclusion of vulnerability indicators and spatial eigenvectors improved the accuracy of our model at the sub-neighborhood scale by up to 77%.

These findings hold significance for all mosquito-borne diseases, underscoring the critical role of accumulating data on human mobility. Such data not only serves as a cornerstone for pathogen dissemination modeling but also facilitates a deeper understanding of exposure dynamics to vectors characterized by diurnal activity fluctuations. Future studies could incorporate these results into vector control planning workflows, for example by identifying zones of potential pathogen transmission potentially underestimated by conventionally applied mosquito abundance maps for targeted control measures. The efficacy of such targeted interventions in mitigating dengue infections warrants evaluation through field studies, extending beyond the specific case of Rio de Janeiro.

## Data statement

The materials and datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Restrictions apply only to the sharing of entomological surveillance data and health data collected by the Municipal Health Ministry of Rio de Janeiro, for which access should be granted directly from there via ethical approval. Additionally, all products generated from the mobile phone dataset can only be shared after approval from the provider.

## Acknowledgements

This work was funded by the German Research Foundation (DFG) [grant number 451956976].

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Credit authorship contribution statement

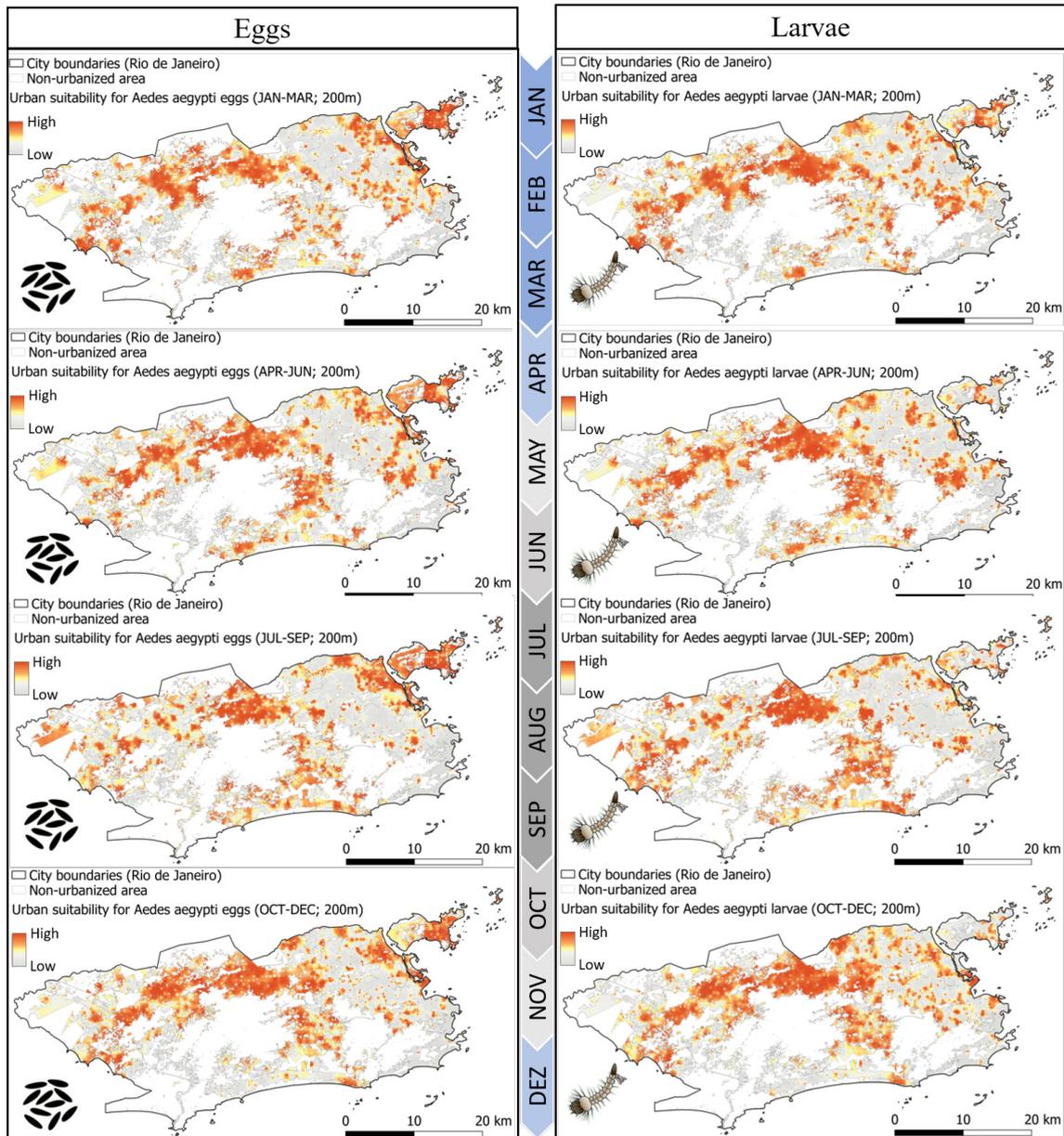
S.K. conceptualized the study and conducted the experiment. S.K. and J.H. validated the results. S.K. and A.R. acquired entomological surveillance data and anonymized health data via an approval of the ethical committee. S.K. and A.R. acquired access to anonymized mobile phone data. M.R. provided resources for the road network centrality measures. S.K. drafted the manuscript. A.Z., A.W.S., B.R., J.R., M.R., O.B., P.P., S.K., T.B., S.L., F.B., and T.J. reviewed and edited the manuscript. A.Z., B.R., S.L., and T.J. supported acquisition of funding. A.Z. and S.L. supervised the study. All authors read and approved the manuscript and its subsequent revisions prior to submission. PG wrote the first draft of the manuscript.

# Appendix

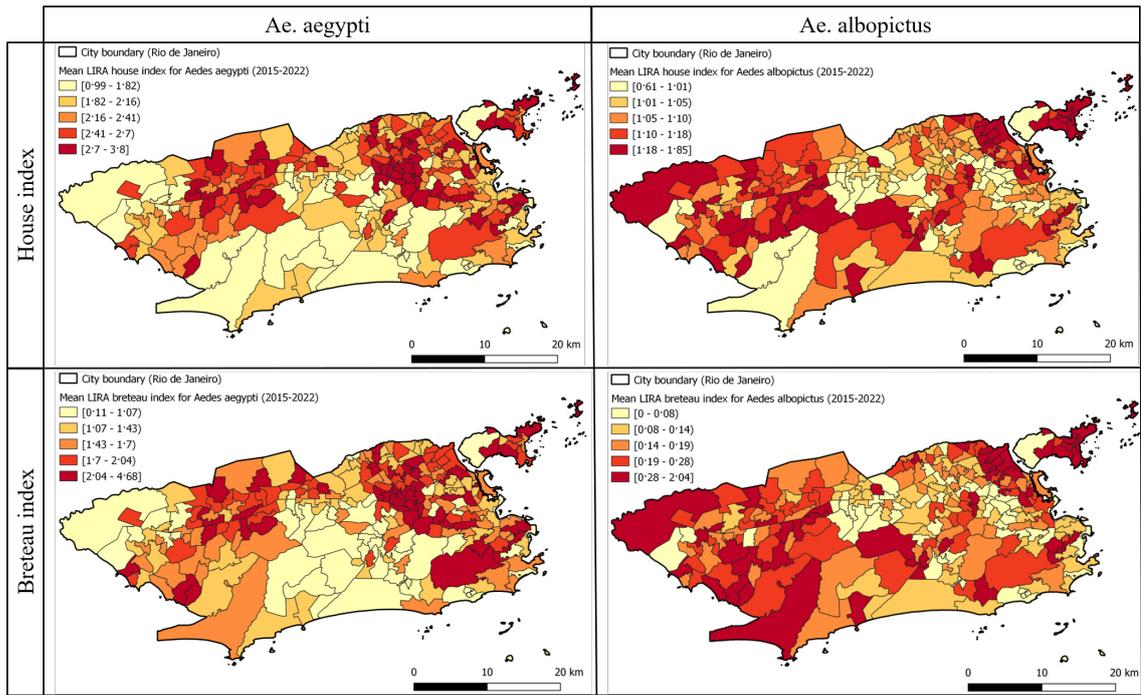
## Appendix A

**Table A. 1:** List of retrieved covariates for the modeling of urban DENV occurrence in the municipality of Rio de Janeiro, including information on the spatial resolution, data source and the required pre-processing step before running zonal statistics on antenna tessellation level.

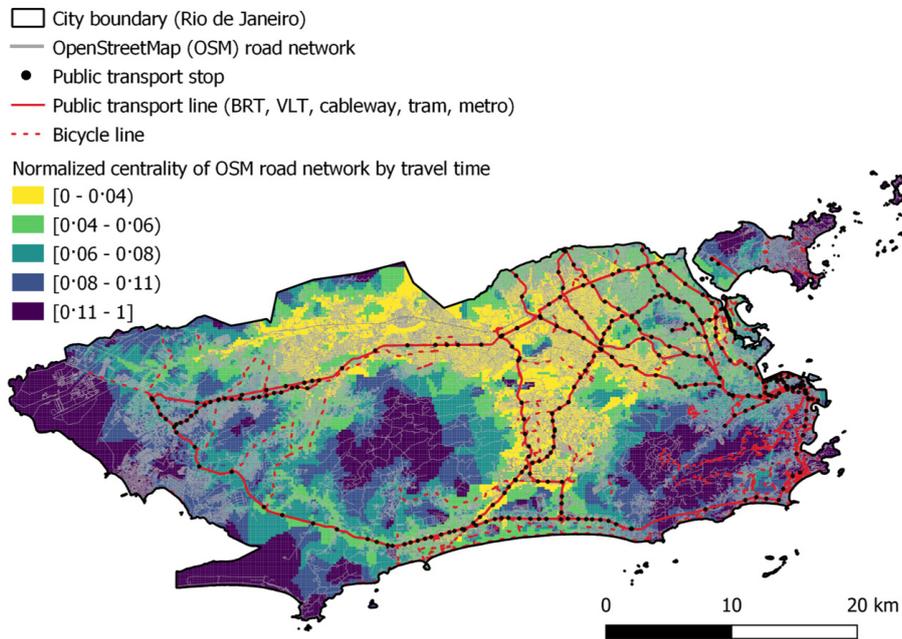
Category	Subcategory	Proxy	Spatial resolution	Source	Required pre-processing step
Hazard	Egg	Estimated <i>Aedes</i> mean egg count per trap (MET)	200m	Ovitrapp	Interpolation (Knoblauch et al., 2023)
	Larva	Estimated <i>Ae. aegypti</i> mean larva count per trap (MLT)	200m	Ovitrapp	Interpolation (Knoblauch et al., 2023)
	Larva and pupa	Estimated <i>Ae. albopictus</i> mean larva count per trap (MLT)	200m	Ovitrapp	Interpolation (Knoblauch et al., 2023)
		Mean <i>Ae. aegypti</i> larva pre-ent index (B1)	Survey strata	LIRA	Rasterization
Vulnerability		Mean <i>Ae. albopictus</i> larva pre-ent index (B1)	Survey strata	LIRA	Rasterization
		Mean <i>Ae. aegypti</i> larva house index (HI)	Survey strata	LIRA	Rasterization
		Mean <i>Ae. albopictus</i> larva house index (HI)	Survey strata	LIRA	Rasterization
		Travel time to closest healthcare facility (TMIST)	H3 spatial grid 0.11 km2	IPEA	-
		Total number of healthcare facilities (S001)	H3 spatial grid 0.11 km2	IPEA	-
		Travel time to closest public school (TMLET)	H3 spatial grid 0.11 km2	IPEA	-
		Total number of public schools (E001)	H3 spatial grid 0.11 km2	IPEA	-
		Cumulative opportunity measure to access jobs in 60 minutes (CMATT60)	H3 spatial grid 0.11 km2	IPEA	-
		Total number of formal jobs (T001)	H3 spatial grid 0.11 km2	IPEA	-
		OSM road network centrality by averaged travel time in car	200m	OpenStreetMap	OpenRouteService (HeGITT GmbH, 2023)
Centrality		Count of public transport stations	< 1m	Data.Rio	-
		Number of residents between 0 and 5 years old	30m	Humanitarian Data Exchange	-
		Number of residents older than 60	30m	Humanitarian Data Exchange	-
Demographics		Average household income per capita (R001)	H3 spatial grid 0.11 km2	IPEA	-
		Socio-economic	< 1m	Municipal health ministry (MIRJ)	Spatiotemporal aggregation
Immunization level		DENV occurrence	< 1m	Municipal health ministry (MIRJ)	Spatiotemporal aggregation
		Human movement flows	30m	Antenna tessellation	Ping extraction (Knoblauch et al., 2024a)
Other		Residential population density	30m	Humanitarian Data Exchange	-



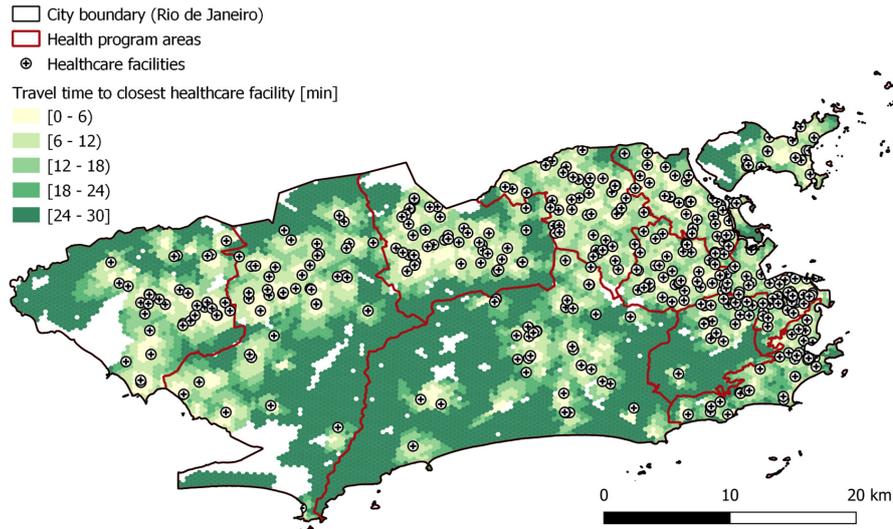
**Figure A. 1:** Seasonal urban suitability for *Ae. aegypti* eggs (left) and larvae (right) at a 200-meter resolution within the municipality of Rio de Janeiro for the year 2019, generated in a prior study (Knoblauch et al., 2023). The analysis employed a complementary approach integrating entomological surveillance data from ovitraps, ecological knowledge concerning limited mosquito flight range, and urban landscape indicators relevant to infer immature *Ae. aegypti* suitability. The blue timescale on the left indicates the wet and dry season.



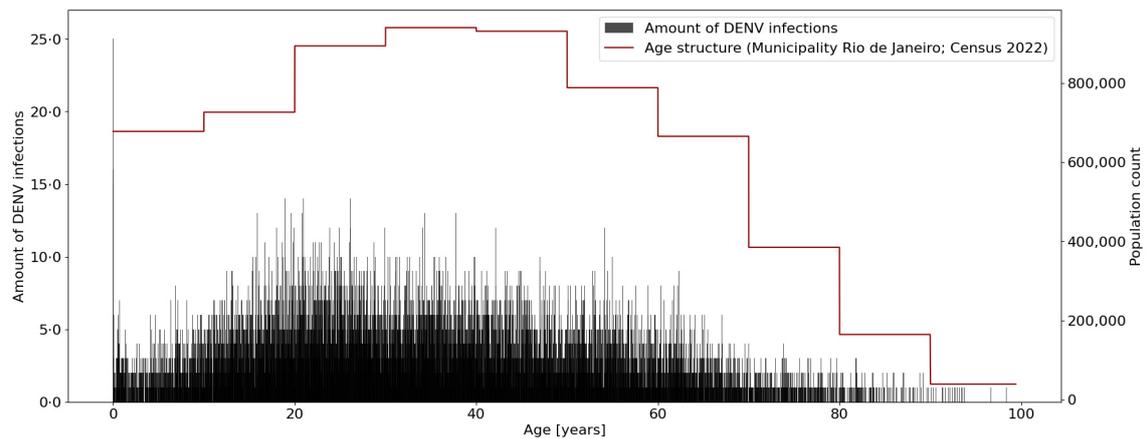
**Figure A. 2:** Entomological surveillance data collected via household survey called LIRA.Ministério da Saúde Brazil, 2013 Maps display the house index (left) and breteau index (right) for *Ae. aegypti* (top) and *Ae. albopictus* (bottom) averaged over 48 seasonal LIRA surveys between 2015 and 2022.



**Figure A. 3:** Normalized centrality of OSM road network measured by travel time in car for the municipality of Rio de Janeiro.



**Figure A. 4:** Healthcare accessibility measured by travel time using an equal modal split of car, public transport, bicycle and walk for the municipality of Rio de Janeiro.



**Figure A. 5:** Comparison between the age distribution of DENV-infected individuals, determined by the time interval between birth and notification date in the official health system, and the overall age structure in Rio de Janeiro municipality as per the 2022 census. For the calculation of the average infection age, official health records from January 2015 to December 2022 were applied. The age structure of infected persons roughly aligns with the general demographic structure.

## Appendix B: Significance of vulnerability indicators and spatial eigenvectors

Among the applied vulnerability indicators, the socio-economic variable of average income emerged as the most influential predictor, demonstrating a negative association (Table B. 1). This suggests that higher average income levels in the municipality of Rio de Janeiro are associated with a reduced risk of dengue infections. Conversely, the indicators of the hypothesized vulnerability categories accessibility and centrality did not exhibit significant predictive power, contradicting our initial assumptions. The same applied to the density of older individuals. However, the density of individuals under five years emerged as a significant

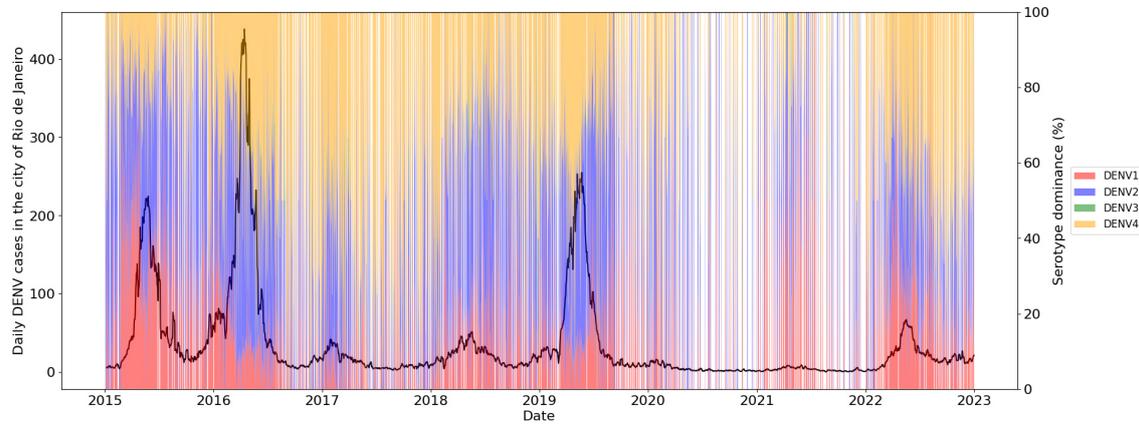
predictor in our model, exhibiting a negative estimate, which suggests that a higher concentration of children was linked to fewer dengue cases. This could potentially be explained by the fact that first dengue infection per individual have a higher probability of being clinically mild (Guzman et al., 2016). Additionally, behavioral factors could play a role, as households with young children may be more vigilant in implementing mosquito control measures, thereby reducing dengue transmission.

**Table B. 1:** Coefficients, standard errors, and p-values for the proposed vulnerability indicators in the extended QP-GLM, which considers hourly human-mosquito biting risk and spatial eigenvectors in the municipality of Rio de Janeiro. Coefficient estimates and standard errors are reported at the link scale.

Vulnerability Class	Coefficients	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	$Pr(>  z )$
Socio-economic	Average household income	-6.2719	0.7059	$< 2e^{-16}$
Accessibility	Travel time to closest healthcare facility	-0.1221	0.2665	$6.47e^{-1}$
	Total number of healthcare facilities	-1.2820	0.5748	$2.59e^{-2}$
	Travel time to closest public school	-0.3676	0.3162	$2.45e^{-1}$
	Total number of public schools	0.1972	0.2977	$5.08e^{-1}$
	Cumulative opportunity measure to access jobs in 60 minutes	-0.0901	0.2497	$7.18e^{-1}$
	Total number of formal jobs	-4.5421	2.2695	$4.55e^{-2}$
Centrality	Standardized OSM road network centrality by average travel time in car	1.4925	0.6917	$3.11e^{-2}$
	Count of public transport stations	0.0893	0.0538	$9.72e^{-2}$
Demographics	Number of residents between 0 and 5 years old	-3.8328	0.7463	$3.23e^{-7}$
	Number of residents older than 60 years	0.1224	0.6864	$8.59e^{-1}$
Immunization level	Estimated level of immunization (2015)	1.3946	0.4481	$1.90e^{-3}$
	Estimated level of immunization (2016)	1.5296	0.4095	$1.95e^{-4}$
	Estimated level of immunization (2017)	0.4041	0.2507	$1.07e^{-1}$
	Estimated level of immunization (2018)	-0.7807	0.4332	$7.18e^{-2}$
	Estimated level of immunization (2019)	1.7539	0.135	$7.00e^{-4}$
	Estimated level of immunization (2020)	0.5402	0.2855	$5.87e^{-2}$
	Estimated level of immunization (2021)	2.9644	0.5117	$8.61e^{-9}$

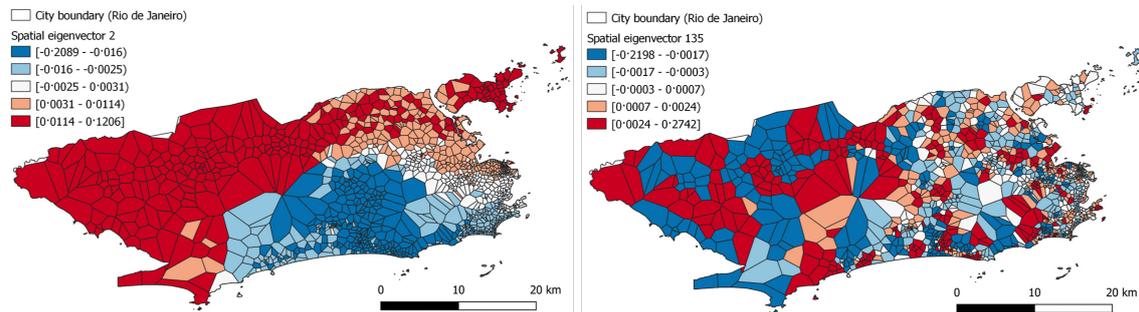
In the assessment of the vulnerability class of immunization, our analysis indicated that the calculated significance values for each year of past infections depend on the magnitude of the outbreak pattern (cf. Figure B. 1). The years with minor outbreaks and less spatial variance in DENV occurrence (2017, 2018, 2020) exhibited either no significant association or marginally statistically significant association, whereas the major outbreak years with larger spatial variance in DENV occurrence (2016, 2019) showed lower p-values. The most recent year in our analysis, 2021, yielded the highest p-value among the hypothesized immunological vulnerability indicators, despite the occurrence of lower dengue incidence. Surprisingly, most estimates of this vulnerability category were positive, contrary to our initial hypothesis about past infections conferring population immunity. We hypothesized that this is related to the fact that environmental factors facilitating transmission are overruling marginal gains in population immunity (under the assumption that cross-immunity between subsequent serotypes or genotypes is relevant). The complex immunological interactions between infections with the four dengue serotypes over time are not further discussed in this context (Simmons et al., 2012). In brief, past infections with a heterologous serotype confer short-term cross-immunity, while past infections with a homotypic serotype confer long-term immunity to the same serotype. The duration and effect size of the heterologous cross-immunity and potentially enhancement is dependent on the time interval between the infections as well as on the specific sequence of the serotypes and their genotypic similarity (Guzman et al., 2016; Katzelnick et al., 2015). These complex immunological interactions between dengue serotypes make it challenging to utilize spatial distribution patterns of dengue cases from

previous years to model immunity levels. Here we can only show the possible existence of a confounding factor not accounted for in the model but influential in driving the spatial distribution of dengue cases at the sub-neighborhood scale.



**Figure B. 1:** Daily fluctuations of DENV case counts and serotype dominance in the municipality of Rio de Janeiro from 2015 to 2022. Larger outbreaks in 2015, 2016, and 2019 co-occur with dominant serotype switches. Days indicated by white stripes indicate a lack of serotype tests in the official health database.

It is important to note that the presented results are dependent on the selection and calculation methods for vulnerability indicators and spatial eigenvectors. Not all additional variables showed significance in our model, underscoring the nuanced impact and selective relevance of certain variables within such broader predictive frameworks. The significance of the proposed human-mosquito biting risk indicator did also diminish in a more extended spatial model. The spatial eigenvectors (cf. Figure B. 2) effectively absorbed a significant portion of the spatial autocorrelation present in the residuals of the proposed QP-GLM. They can be instrumental in formulating additional hypotheses regarding potential missing covariates or confounding factors for integration into future models. Furthermore, they can serve as a tool for testing spatially varying regression coefficients.



**Figure B. 2:** Exemplary spatial eigenvectors derived from the model's spatial weight matrix illustrating distinct patterns: Spatial eigenvector 2 shows gradual spatial gradients, and spatial eigenvector 135 depicts localized clustering. These eigenvectors unveil varying spatial structures within the study area, providing valuable insights into the underlying spatial relationships influencing the observed phenomena.

# Bibliography

- Althouse, B. M., Scarpino, S. V., Meyers, L. A., Ayers, J. W., Bargsten, M., Baumbach, J., Brownstein, J. S., Castro, L., Clapham, H., Cummings, D. A., Del Valle, S., Eubank, S., Fairchild, G., Finelli, L., Generous, N., George, D., Harper, D. R., Hébert-Dufresne, L., Johansson, M. A., Konty, K., Lipsitch, M., Milinovich, G., Miller, J. D., Nsoesie, E. O., Olson, D. R., Paul, M., Polgreen, P. M., Priedhorsky, R., Read, J. M., Rodríguez-Barraguer, I., Smith, D. J., Stefansen, C., Swerdlow, D. L., Thompson, D., Vespignani, A., and Wesolowski, A. (2015). “Enhancing disease surveillance with novel data streams: challenges and opportunities”. *EPJ data science* 4.1. DOI: 10.1140/epjds/s13688-015-0054-0.
- Baik, L. S., Nave, C., Au, D. D., Guda, T., Chevez, J. A., Ray, A., and Holmes, T. C. (2020). “Circadian Regulation of Light-Evoked Attraction and Avoidance Behaviors in Daytime-versus Nighttime-Biting Mosquitoes”. *Current biology: CB* 30.16, 3252–3259.e3. DOI: 10.1016/j.cub.2020.06.010.
- Banerjee, S., Aditya, G., and Saha, G. K. (2015). “Household Wastes as Larval Habitats of Dengue Vectors: Comparison between Urban and Rural Areas of Kolkata, India”. *PloS one* 10.10, e0138082. DOI: 10.1371/journal.pone.0138082.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I. (2013). “The global distribution and burden of dengue”. *Nature* 496.7446, pp. 504–507. DOI: 10.1038/nature12060.
- Bivand, R. (2023). *Spatialreg R package documentation*.
- Bomfim, R., Pei, S., Shaman, J., Yamana, T., Makse, H. A., Andrade, J. S., Lima Neto, A. S., and Furtado, V. (2020). “Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas”. *Journal of the Royal Society, Interface* 17.171, p. 20200691. DOI: 10.1098/rsif.2020.0691.
- Chowell, G. and Nishiura, H. (2014). “Transmission dynamics and control of Ebola virus disease (EVD): a review”. *BMC medicine* 12, p. 196. DOI: 10.1186/s12916-014-0196-0.
- Cohen, J. M., Le Menach, A., Pothin, E., Eisele, T. P., Gething, P. W., Eckhoff, P. A., Moonen, B., Schapira, A., and Smith, D. L. (2017). “Mapping multiple components of malaria risk for improved targeting of elimination interventions”. *Malaria journal* 16.1, p. 459. DOI: 10.1186/s12936-017-2106-3.
- Dengue explorer (2023). *Dengue Explorer*.
- Egid, B. R., Coulibaly, M., Dadzie, S. K., Kamgang, B., McCall, P. J., Sedda, L., Toe, K. H., and Wilson, A. L. (2022). “Review of the ecology and behaviour of *Aedes aegypti* and *Aedes albopictus* in Western Africa and implications for vector control”. *Current research in parasitology & vector-borne diseases* 2, p. 100074. DOI: 10.1016/j.crpvbd.2021.100074.

- Ferreira-de-Lima, V. H. and Lima-Camara, T. N. (2018). “Natural vertical transmission of dengue virus in *Aedes aegypti* and *Aedes albopictus*: a systematic review”. *Parasites & vectors* 11.1, p. 77. DOI: 10.1186/s13071-018-2643-9.
- Finger, F., Genolet, T., Mari, L., Magny, G. C. de, Manga, N. M., Rinaldo, A., and Bertuzzo, E. (2016). “Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks”. *Proceedings of the National Academy of Sciences of the United States of America* 113.23, pp. 6421–6426. DOI: 10.1073/pnas.1522305113.
- Funk, S., Salathé, M., and Jansen, V. A. A. (2010). “Modelling the influence of human behaviour on the spread of infectious diseases: a review”. *Journal of the Royal Society, Interface* 7.50, pp. 1247–1256. DOI: 10.1098/rsif.2010.0142.
- Getachew, D., Tekie, H., Gebre-Michael, T., Balkew, M., and Mesfin, A. (2015). “Breeding Sites of *Aedes aegypti*: Potential Dengue Vectors in Dire Dawa, East Ethiopia”. *Interdisciplinary perspectives on infectious diseases* 2015, p. 706276. DOI: 10.1155/2015/706276.
- Griffith, D. A. (2000). “Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses”. *Linear Algebra and its Applications* 321.1-3, pp. 95–112. DOI: 10.1016/S0024-3795(00)00031-8.
- Griffith, D. A. (2019). *Spatial regression analysis using Eigenvector spatial filtering*. S.l.: Academic Press.
- Guzman, M. G., Gubler, D. J., Izquierdo, A., Martinez, E., and Halstead, S. B. (2016). “Dengue infection”. *Nature reviews. Disease primers* 2, p. 16055. DOI: 10.1038/nrdp.2016.55.
- Harrington, L. C., Scott, T. W., Lerdthusnee, K., Coleman, R. C., Costero, A., Clark, G. G., Jones, J. J., Kitthawee, S., Kittayapong, P., Sithiprasasna, R., and Edman, J. D. (2005). “Dispersal of the dengue vector *Aedes aegypti* within and between rural communities”. *The American Journal of Tropical Medicine and Hygiene* 72.2, pp. 209–220. DOI: 10.4269/ajtmh.2005.72.209.
- HeiGIT gGmbH (2023). *openrouteservice*.
- Hladish, T. J., Pearson, C. A. B., Toh, K. B., Rojas, D. P., Manrique-Saide, P., Vazquez-Prokopec, G. M., Halloran, M. E., and Longini, I. M. (2020). “Designing effective control of dengue with combined interventions”. *Proceedings of the National Academy of Sciences of the United States of America* 117.6, pp. 3319–3325. DOI: 10.1073/pnas.1903496117.
- Honório, N. A., Da Silva, W. C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003). “Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil”. *Memorias do Instituto Oswaldo Cruz* 98.2, pp. 191–198. DOI: 10.1590/s0074-02762003000200005.
- Humanitarian Data Exchange (2019). *Brazil: High Resolution Population Density Maps + Demographic Estimates*. 01.06.2019.
- Kache, P. A., Santos-Vega, M., Stewart-Ibarra, A. M., Cook, E. M., Seto, K. C., and Diuk-Wasser, M. A. (2022). “Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases”. *Nature ecology & evolution* 6.11, pp. 1601–1616. DOI: 10.1038/s41559-022-01876-y.
- Katzelnick, L. C., Fonville, J. M., Gromowski, G. D., Bustos Arriaga, J., Green, A., James, S. L., Lau, L., Montoya, M., Wang, C., VanBlargan, L. A., Russell, C. A., Thu, H. M., Pierson, T. C., Buchy, P., Aaskov, J. G., Muñoz-Jordán, J. L., Vasilakis, N., Gibbons, R. V., Tesh, R. B., Osterhaus, A. D. M. E., Fouchier, R. A. M., Durbin, A., Simmons, C. P., Holmes, E. C., Harris, E., Whitehead, S. S., and Smith, D. J.

- (2015). “Dengue viruses cluster antigenically but not as discrete serotypes”. *Science (New York, N.Y.)* 349.6254, pp. 1338–1343. DOI: 10.1126/science.aac5017.
- Knoblauch, S., Gross, S., Lautenbach, S., Rocha, A. A. d. A., González, M. C., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2024a). “Long-term validation of inner-urban mobility metrics derived from Twitter (under review)”. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/23998083241278275.
- Knoblauch, S., Mukaratirwa, R., Pimenta, P., Wilder-Smith, A., Rocklöv, J., Lautenbach, S., Brady, O. J., Rocha, A. A. d. A., Dambach, P., Jaenisch, T., Resch, B., Biljecki, F., Haddawy, P., Myat, S.Y., Bärnighausen, T., and Zipf, A. (2023). “Suitability indicators for immature *Aedes aegypti* to guide vector control in the municipality of Rio de Janeiro (submitted)”. *The Lancet Regional Health - Americas*.
- Knoblauch, S., Su Yin, M., Chatrinan, K., Rocha, A. A. d. A., Haddawy, P., Biljecki, F., Lautenbach, S., Resch, B., Arifi, D., Jänisch, T., Morales, I., and Zipf, A. (2024b). “High-resolution mapping of urban *Aedes aegypti* immature abundance through breeding site detection based on satellite and street view imagery (under review)”. *Scientific Reports: Collection on Urbanization and communicable diseases*. DOI: 10.1038/s41598-024-67914-w.
- Kogan, N. E., Clemente, L., Liautaud, P., Kaashoek, J., Link, N. B., Nguyen, A. T., Lu, F. S., Huybers, P., Resch, B., Havas, C., Petutschnig, A., Davis, J., Chinazzi, M., Mustafa, B., Hanage, W. P., Vespignani, A., and Santillana, M. (2021). “An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time”. *Science advances* 7.10. DOI: 10.1126/sciadv.abd6989.
- Kraemer, M. U. G., Bisanzio, D., Reiner, R. C., Zakar, R., Hawkins, J. B., Freifeld, C. C., Smith, D. L., Hay, S. I., Brownstein, J. S., and Perkins, T. A. (2018). “Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan”. *EPJ data science* 7.1, p. 16. DOI: 10.1140/epjds/s13688-018-0144-x.
- Kraemer, M. U. G., Hay, S. I., Pigott, D. M., Smith, D. L., Wint, G. R. W., and Golding, N. (2016). “Progress and Challenges in Infectious Disease Cartography”. *Trends in parasitology* 32.1, pp. 19–29. DOI: 10.1016/j.pt.2015.09.006.
- Lai, Y.-J., Lai, H.-H., Chen, Y.-Y., Ko, M.-C., Chen, C.-C., Chuang, P.-H., Yen, Y.-F., and Morisky, D. E. (2020). “Low socio-economic status associated with increased risk of dengue haemorrhagic fever in Taiwanese patients with dengue fever: a population-based cohort study”. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 114.2, pp. 115–120. DOI: 10.1093/trstmh/trz103.
- Lenormand, M., Picornell, M., Cantú-Ros, O. G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frías-Martínez, E., and Ramasco, J. J. (2014). “Cross-checking different sources of mobility information”. *PloS one* 9.8, e105184. DOI: 10.1371/journal.pone.0105184.
- Lobo, N. F., Achee, N. L., Greico, J., and Collins, F. H. (2018). “Modern Vector Control”. *Cold Spring Harbor perspectives in medicine* 8.1. DOI: 10.1101/cshperspect.a025643.
- Mahabir, R. S., Severson, D. W., and Chadee, D. D. (2012). “Impact of road networks on the distribution of dengue fever cases in Trinidad, West Indies”. *Acta tropica* 123.3, pp. 178–183. DOI: 10.1016/j.actatropica.2012.05.001.
- Massad, E., Amaku, M., Coutinho, F. A. B., Struchiner, C. J., Lopez, L. F., Wilder-Smith, A., and Burattini, M. N. (2017). “Estimating the size of *Aedes aegypti* populations from dengue incidence data: Implications for the risk of yellow fever

- outbreaks”. *Infectious Disease Modelling* 2.4, pp. 441–454. DOI: 10.1016/j.idm.2017.12.001.
- McMaughan, D. J., Oloruntoba, O., and Smith, M. L. (2020). “Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging”. *Frontiers in public health* 8, p. 231. DOI: 10.3389/fpubh.2020.00231.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. (2019). “The current and future global distribution and population at risk of dengue”. *Nature microbiology* 4.9, pp. 1508–1515. DOI: 10.1038/s41564-019-0476-8.
- Ministério da Saúde Brazil (2013). *Levantamento rápido de índices para aedes aegypti*.
- Morrison, A. C., Gray, K., Getis, A., Astete, H., Sihuíncha, M., Focks, D., Watts, D., Stancil, J. D., Olson, J. G., Blair, P., and Scott, T. W. (2004). “Temporal and geographic patterns of *Aedes aegypti* (Diptera: Culicidae) production in Iquitos, Peru”. *Journal of medical entomology* 41.6, pp. 1123–1142. DOI: 10.1603/0022-2585-41.6.1123.
- Muhammad, N. A. F., Abu Kassim, N. F., Ab Majid, A. H., Abd Rahman, A., Dieng, H., and Avicor, S. W. (2020). “Biting rhythm and demographic attributes of *Aedes albopictus* (Skuse) females from different urbanized settings in Penang Island, Malaysia under uncontrolled laboratory conditions”. *PloS one* 15.11, e0241688. DOI: 10.1371/journal.pone.0241688.
- Mutebi, J.-P., Wilke, A. B. B., Ostrum, E., Vasquez, C., Cardenas, G., Carvajal, A., Moreno, M., Petrie, W. D., Rodriguez, A., Presas, H., Rodriguez, J., Barnes, F., Hamer, G. L., Juarez, J. G., Carbajal, E., Vitek, C. J., Estrada, X., Rios, T., Marshall, J., and Beier, J. C. (2022). “Diel activity patterns of two distinct populations of *Aedes aegypti* in Miami, FL and Brownsville, TX”. *Scientific reports* 12.1, p. 5315. DOI: 10.1038/s41598-022-06586-w.
- Panigutti, C., Tizzoni, M., Bajardi, P., Smoreda, Z., and Colizza, V. (2017). “Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models”. *Royal Society open science* 4.5, p. 160950. DOI: 10.1098/rsos.160950.
- Paploski, I. A. D., Rodrigues, M. S., Mugabe, V. A., Kikuti, M., Tavares, A. S., Reis, M. G., Kitron, U., and Ribeiro, G. S. (2016). “Storm drains as larval development and adult resting sites for *Aedes aegypti* and *Aedes albopictus* in Salvador, Brazil”. *Parasites & vectors* 9.1, p. 419. DOI: 10.1186/s13071-016-1705-0.
- Pereira, R. and Herszenhut, D. (2023). *Introduction to urban accessibility: a practical guide with R*.
- Perkins, T. A., Metcalf, C. J. E., Grenfell, B. T., and Tatem, A. J. (2015). “Estimating Drivers of Autochthonous Transmission of Chikungunya Virus in its Invasion of the Americas”. *PLoS Currents* 7. DOI: 10.1371/currents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc.
- Perkins, T. A., Paz-Soldan, V. A., Stoddard, S. T., Morrison, A. C., Forshey, B. M., Long, K. C., Halsey, E. S., Kochel, T. J., Elder, J. P., Kitron, U., Scott, T. W., and Vazquez-Prokopec, G. M. (2016). “Calling in sick: impacts of fever on intra-urban human mobility”. *Proceedings. Biological sciences* 283.1834. DOI: 10.1098/rspb.2016.0390.
- Perkins, T. A., Scott, T. W., Le Menach, A., and Smith, D. L. (2013). “Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission”. *PLoS computational biology* 9.12, e1003327. DOI: 10.1371/journal.pcbi.1003327.

- Reiner, R. C., Stoddard, S. T., and Scott, T. W. (2014). “Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal”. *Epidemics* 6, pp. 30–36. DOI: 10.1016/j.epidem.2013.12.003.
- Salje, H., Lessler, J., Maljkovic Berry, I., Melendrez, M. C., Endy, T., Kalayanarooj, S., A-Nuegoonpipat, A., Chanama, S., Sangkijporn, S., Klungthong, C., Thaisomboonsuk, B., Nisalak, A., Gibbons, R. V., Iamsirithaworn, S., Macareo, L. R., Yoon, I.-K., Sangarsang, A., Jarman, R. G., and Cummings, D. A. T. (2017). “Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size”. *Science (New York, N.Y.)* 355.6331, pp. 1302–1306. DOI: 10.1126/science.aaj9384.
- Sattenspiel, Lisa and Lloyd, Alun (2009). *The geographic spread of infectious diseases: models and applications*. 5th ed. Princeton University Press.
- Schaber, K. L., Paz-Soldan, V. A., Morrison, A. C., Elson, W. H. D., Rothman, A. L., Mores, C. N., Astete-Vega, H., Scott, T. W., Waller, L. A., Kitron, U., Elder, J. P., Barker, C. M., Perkins, T. A., and Vazquez-Prokopec, G. M. (2019). “Dengue illness impacts daily human mobility patterns in Iquitos, Peru”. *PLoS neglected tropical diseases* 13.9, e0007756. DOI: 10.1371/journal.pntd.0007756.
- Secretaria Municipal de Saude Rio de Janeiro (2022). *Weekly Dengue Occurrence*.
- Simmons, C. P., Farrar, J. J., van Nguyen, V. C., and Wills, B. (2012). “Dengue”. *The New England journal of medicine* 366.15, pp. 1423–1432. DOI: 10.1056/NEJMra1110265.
- Stoddard, S. T., Forshey, B. M., Morrison, A. C., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., Astete, H., Reiner, R. C., Vilcarromero, S., Elder, J. P., Halsey, E. S., Kochel, T. J., Kitron, U., and Scott, T. W. (2013). “House-to-house human movement drives dengue virus transmission”. *Proceedings of the National Academy of Sciences of the United States of America* 110.3, pp. 994–999. DOI: 10.1073/pnas.1213349110.
- Stoddard, S. T., Morrison, A. C., Vazquez-Prokopec, G. M., Paz Soldan, V., Kochel, T. J., Kitron, U., Elder, J. P., and Scott, T. W. (2009). “The role of human movement in the transmission of vector-borne pathogens”. *PLoS neglected tropical diseases* 3.7, e481. DOI: 10.1371/journal.pntd.0000481.
- Thai, K. T. D., Nishiura, H., Hoang, P. L., Tran, N. T. T., Phan, G. T., Le, H. Q., Tran, B. Q., van Nguyen, N., and Vries, P. J. de (2011). “Age-specificity of clinical dengue during primary and secondary infections”. *PLoS neglected tropical diseases* 5.6, e1180. DOI: 10.1371/journal.pntd.0001180.
- Trewin, B. J., Parry, H. R., Pagendam, D. E., Devine, G. J., Zalucki, M. P., Darbro, J. M., Jansen, C. C., and Schellhorn, N. A. (2021). “Simulating an invasion: unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, *Aedes aegypti*, in Brisbane, Australia”. *Biological invasions* 23.12, pp. 3891–3906. DOI: 10.1007/s10530-021-02619-z.
- Vanlerberghe, V., Gómez-Dantés, H., Vazquez-Prokopec, G., Alexander, N., Manrique-Saide, P., Coelho, G., Toledo, M. E., Ocampo, C. B., and van der Stuyft, P. (2017). “Changing paradigms in *Aedes* control: considering the spatial heterogeneity of dengue transmission”. *Revista Panamericana de Salud Pública* 41. DOI: 10.26633/RPSP.2017.16.
- Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W., and Kitron, U. (2013). “Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment”. *PloS one* 8.4, e58802. DOI: 10.1371/journal.pone.0058802.

- Vezzani, D. (2007). “Review: artificial container-breeding mosquitoes and cemeteries: a perfect match”. *Tropical medicine & international health : TM & IH* 12.2, pp. 299–313. DOI: 10.1111/j.1365-3156.2006.01781.x.
- Wei, L., FernÁndez-Santos, N. A., Hamer, G. L., Lara-Ramírez, E. E., and Rodríguez-PÉrez, M. A. (2023). “Daytime Resting Activity of *Aedes Aegypti* and *Culex Quinquefasciatus* Populations in Northern Mexico”. *Journal of the American Mosquito Control Association* 39.3, pp. 157–167. DOI: 10.2987/23-7122.
- Wesolowski, A., Buckee, C. O., Engø-Monsen, K., and Metcalf, C. J. E. (2016). “Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data”. *The Journal of infectious diseases* 214.suppl\_4, S414–S420. DOI: 10.1093/infdis/jiw273.
- Wilson, A. L., Courtenay, O., Kelly-Hope, L. A., Scott, T. W., Takken, W., Torr, S. J., and Lindsay, S. W. (2020). “The importance of vector control for the control and elimination of vector-borne diseases”. *PLoS neglected tropical diseases* 14.1, e0007831. DOI: 10.1371/journal.pntd.0007831.
- World Bank (2023). *World Development Indicators DataBank: Urban population (% of total population) 1960-2021*.
- World Health Organization (2012). *Global Strategy for Dengue Prevention and Control, 2012-2020*. Geneva, Switzerland.
- Wu, S. L., Henry, J. M., Citron, D. T., Mbabazi Ssebuliba, D., Nakakawa Nsumba, J., Sánchez C, H. M., Brady, O. J., Guerra, C. A., García, G. A., Carter, A. R., Ferguson, H. M., Afolabi, B. E., Hay, S. I., Reiner, R. C., Kiware, S., and Smith, D. L. (2023). “Spatial dynamics of malaria transmission”. *PLoS computational biology* 19.6, e1010684. DOI: 10.1371/journal.pcbi.1010684.
- Yang, J., Mosabbir, A. A., Raheem, E., Hu, W., and Hossain, M. S. (2023). “Demographic characteristics, clinical symptoms, biochemical markers and probability of occurrence of severe dengue: A multicenter hospital-based study in Bangladesh”. *PLoS neglected tropical diseases* 17.3, e0011161. DOI: 10.1371/journal.pntd.0011161.
- Yuan, Y. and Raubal, M. (2012). “Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data”. *Geographic Information Science*. Ed. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, N. Xiao, M.-P. Kwan, M. F. Goodchild, and S. Shekhar. Vol. 7478. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 354–367. DOI: 10.1007/978-3-642-33024-7{\textunderscore}26.
- Zahid, M. H., van Wyk, H., Morrison, A. C., Coloma, J., Lee, G. O., Cevallos, V., Ponce, P., and Eisenberg, J. N. S. (2023). “The biting rate of *Aedes aegypti* and its variability: A systematic review (1970-2022)”. *PLoS neglected tropical diseases* 17.8, e0010831. DOI: 10.1371/journal.pntd.0010831.

# Part III

## Declarations





**Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die  
Naturwissenschaftlich-Mathematische Gesamtfakultät der Universität Heidelberg / Sworn  
Affidavit according to § 8 of the doctoral degree regulations of the Combined Faculty of  
Natural Sciences and Mathematics**

1. Bei der eingereichten Dissertation zu dem Thema / *The thesis I have submitted entitled*

Harnessing Geospatial Big Data to Guide Local Interventions for Aedes-borne Arboviral Infections

handelt es sich um meine eigenständig erbrachte Leistung / *is my own work.*

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. / *I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.*

3. Die Arbeit oder Teile davon habe ich wie ~~folgt~~/bislang nicht<sup>1)</sup> an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. / *I have not yet/have already<sup>1)</sup> presented this thesis or parts thereof to a university as part of an examination or degree.*

Titel der Arbeit / *Title of the thesis:*

Hochschule und Jahr / *University and year:*

Art der Prüfungs- oder Qualifikationsleistung / *Type of examination or degree:*

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. / *I confirm that the declarations made above are correct.*
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. / *I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit*

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe. / *I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.*

20.09.2024

Ort und Datum / *Place and date*

*St. Knoblauch*

Unterschrift / *Signature*

<sup>1)</sup> Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung. / *Please cross out what is not applicable. If applicable, please provide: the title of the thesis that was presented elsewhere, the name of the university, the year of presentation and the type of examination or degree.*

---

**1. Publikation/Publication:**

Vollständige bibliographische Referenz/Complete bibliographic reference:

Knoblauch S., Li H., Lautenbach S., Elshiaty Y., Rocha A., Resch B., Arifi D., Jänisch T., Morales I., Zipf A. (2023). Semi-supervised water tank detection to support vector control of emerging infectious diseases transmitted by Aedes Aegypti. International Journal of Applied Earth Observation and Geoinformation (Volume 119) DOI: 10.1016/j.isprsjprs.2021.06.011

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes  Nein/No 3. Veröffentlicht/Published  Zur Veröffentlichung akzeptiert/Accepted 

Q1/Q2\*:

Ja/Yes  Nein/No 

\*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: \_\_\_\_\_

Eingereicht/Submitted  Noch nicht eingereicht/Not yet submitted **4. Beteiligungen/Contributions\*\***

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Knoblauch, Steffen	Li, Hao	Zipf, Alexander
Methodology	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Software	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Validation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formal analysis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Investigation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resources	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Curation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Original Draft	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Review&Editing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Visualization	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Project administration	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Funding acquisition	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*\*Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature St. Knoblauch

Doktorand/in/Doctoral student

Li-Hao

Co-Autor/in 1/Co-author 1

A. Zipf

Co-Autor/in 2/Co-author 2

**Betreuungsperson/Supervisor:**

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Zipf, Alexander

Name/Name

A. Zipf

Unterschrift/Signature

09.09.2024

Datum/Date

---

**1. Publikation/Publication:**

Vollständige bibliographische Referenz/Complete bibliographic reference:

Knoblauch, S., Su Yin, M., Chatrinan, K., Rocha A., Haddawy P., Biljecki F., Lautenbach S., Resch B., Arifi D., Jänisch T., Morales I., Zipf A. (2024) High-resolution mapping of urban Aedes aegypti immature abundance through breeding site detection based on satellite and street view imagery. Scientific Reports 14, 18227. DOI: <https://doi.org/10.1038/s41598-024-67914-w>

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes  Nein/No

3. Veröffentlicht/Published  Zur Veröffentlichung akzeptiert/Accepted

Q1/Q2\*: Ja/Yes  Nein/No

\*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: \_\_\_\_\_

Eingereicht/Submitted  Noch nicht eingereicht/Not yet submitted

**4. Beteiligungen/Contributions\*\***

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Knoblauch, Steffen	Rocha, Antonio	Zipf, Alexander
Methodology	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Software	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formal analysis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Investigation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resources	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Data Curation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Original Draft	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Review&Editing	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Visualization	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Project administration	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Funding acquisition	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*\*Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature St. Knoblauch A. Rocha A. Zipf  
 Doktorand/in/Doctoral student Co-Autor/in 1/Co-author 1 Co-Autor/in 2/Co-author 2

**Betreuungsperson/Supervisor:**

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Zipf, Alexander A. Zipf 09.09.2024  
 Name/Name Unterschrift/Signature Datum/Date

---

**1. Publikation/Publication:**

Vollständige bibliographische Referenz/Complete bibliographic reference:

Knoblauch S., Tadiwa Mukaratirwa R., Paolucci Pimenta P.F., Wilder-Smith A., Rocklöv J., Lautenbach S., Randhawa S., Brady O.J., Rocha A., Dambach P., Jänisch T., Resch B., Biljecki F., Haddawy P., Su Yin M., Bärnighausen T., Zipf A.. Suitability indicators for immature Aedes aegypti to guide vector control in the municipality of Rio de Janeiro. Under review in "The Lancet Planetary Health".

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes  Nein/No

3. Veröffentlicht/Published  Zur Veröffentlichung akzeptiert/Accepted

Q1/Q2\*: Ja/Yes  Nein/No

\*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: \_\_\_\_\_

Eingereicht/Submitted  Noch nicht eingereicht/Not yet submitted

**4. Beteiligungen/Contributions\*\***

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Knoblauch, Steffen	Paolucci Pimenta, Paulo Filemon	Zipf, Alexander
Methodology	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Software	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formal analysis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Investigation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resources	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Curation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Original Draft	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Review&Editing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Visualization	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Project administration	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Funding acquisition	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*\*Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature

St. Knoblauch Paulo Pimenta A. Zipf  
 Doktorand/in/Doctoral student Co-Autor/in 1/Co-author 1 Co-Autor/in 2/Co-author 2

**Betreuungsperson/Supervisor:**

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Zipf, Alexander A. Zipf 09.09.2024  
 Name/Name Unterschrift/Signature Datum/Date

---

**1. Publikation/Publication:**

Vollständige bibliographische Referenz/Complete bibliographic reference:

Knoblauch S., Gross S., Lautenbach S., Rocha A., Gonzalez M.C., Resch B., Arifi D., Jänisch T., Morales I., Zipf A. (2023). "Long-term validation of inner-urban mobility metrics derived from Twitter/X. Environment and Planning B: Urban Analytics and City Science. DOI: 10.1177/23998083241278275

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes  Nein/No

3. Veröffentlicht/Published  Zur Veröffentlichung akzeptiert/Accepted

Q1/Q2\*: Ja/Yes  Nein/No

\*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: \_\_\_\_\_

Eingereicht/Submitted  Noch nicht eingereicht/Not yet submitted

**4. Beteiligungen/Contributions\*\***

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Knoblauch, Steffen	Rocha, Antonio	Zipf, Alexander
Methodology	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Software	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formal analysis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Investigation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resources	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Data Curation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Original Draft	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Review&Editing	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Visualization	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Project administration	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Funding acquisition	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*\*Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature St. Knoblauch  
Doktorand/in/Doctoral student

Antonio Rocha  
Co-Autor/in 1/Co-author 1

A. Zipf  
Co-Autor/in 2/Co-author 2

**Betreuungsperson/Supervisor:**

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Zipf, Alexander  
Name/Name

A. Zipf  
Unterschrift/Signature

09.09.2024  
Datum/Date

---

**1. Publikation/Publication:**

Vollständige bibliographische Referenz/Complete bibliographic reference:

Knoblauch S., Heidecke J., Reinmuth M., Lautenbach S., Paolucci Pimenta P.F., Rocklöv J., Brady O.J., Rocha A., Jänisch T., Resch B., Biljecki F., Wilder-Smith A., Bärnighausen T., Zipf A.. "Assessing Dengue Risk in Urban Areas: The Role of Daytime Aedes-Human Interaction." Under review in "Scientific Reports".

2. Erst- oder gleichberechtigte Autorenschaft/First or equal authorship: Ja/Yes  Nein/No 3. Veröffentlicht/Published  Zur Veröffentlichung akzeptiert/Accepted 

Q1/Q2\*:

Ja/Yes  Nein/No 

\*SCImago Journal Rank (SJR) indicator

Im Erscheinungsjahr oder im letzten verfügbaren Vorjahr/In the year of publication or the last prior year available: \_\_\_\_\_

Eingereicht/Submitted  Noch nicht eingereicht/Not yet submitted **4. Beteiligungen/Contributions\*\***

Contributor Role	Doktorand/in/ Doctoral student	Co-Autor/in 1/ Co-author 1	Co-Autor/in 2/ Co-author 2
Name, first name	Knoblauch, Steffen	Heidecke, Julian	Zipf, Alexander
Methodology	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Software	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formal analysis	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Investigation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resources	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Curation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Original Draft	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Writing-Review&Editing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Visualization	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Project administration	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Funding acquisition	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

\*\*Kategorien des CRediT (Contributor Roles Taxonomy, <https://credit.niso.org/>)

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind/I confirm that all declarations made above are correct.

Unterschrift/Signature St. Knoblauch

Doktorand/in/Doctoral student

J. Heidecke

Co-Autor/in 1/Co-author 1

A. Zipf

Co-Autor/in 2/Co-author 2

**Betreuungsperson/Supervisor:**

Hiermit bestätige ich, dass alle obigen Angaben korrekt sind und dass die selbstständigen Arbeitsanteile des/der Doktoranden/in an der aufgeführten Publikation hinreichend und signifikant sind/I confirm that all declarations made above are correct and that the doctoral student's independent contribution to this publication is significant and sufficient to be considered for the cumulative dissertation.

Zipf, Alexander

Name/Name

A. Zipf

Unterschrift/Signature

09.09.2024

Datum/Date