

Salvador Urban Network Transportation (SUNT): A Landmark Spatiotemporal Dataset for Public Transportation (Abstract Reprint)

Marcos V. Ferreira¹, Matheus Souza², Tatiane N. Rios^{1,3}, Islame F. C. Fernandes¹, Jorge Nery^{1,2},
Joo Gama⁴, Albert Bifet³, Ricardo A. Rios^{1,3}

¹Federal University of Bahia, Institute of Computing, Salvador, Brazil

²Integra (Association of Transport Companies of Salvador), Salvador, Brazil

³University of Waikato, Hamilton, New Zealand

⁴University of Porto, Porto, Portugal

Abstract Reprint. This is an abstract reprint of the journal article by Ferreira, Souza, Rios, Fernandes, Nery, Gama, Bifet, and Rios (2025).

Abstract

Efficient public transportation management is essential for the development of large urban centers, providing several benefits such as comprehensive coverage of population mobility, reduction of transport costs, better control of traffic congestion, and significant reduction of environmental impact limiting gas emissions and pollution. Realizing these benefits requires a deeply understanding the population and transit patterns and the adoption of approaches to model multiple relations and characteristics efficiently. This work addresses these challenges by providing a novel dataset that includes various public transportation components from three different systems: regular buses, subway, and BRT (Bus Rapid Transit). Our dataset comprises daily information from about 700,000 passengers in Salvador, one of Brazil's largest cities, and local public transportation data with approximately 2,000 vehicles operating across nearly 400 lines, connecting almost 3,000 stops and stations. With data collected from March 2024 to March 2025 at a frequency lower than one minute, SUNT stands as one of the largest, most comprehensive, and openly available urban datasets in the literature.

References

Ferreira, M. V.; Souza, M.; Rios, T. N.; Fernandes, I. F. C.; Nery, J.; Gama, J.; Bifet, A.; and Rios, R. A. 2025. Salvador Urban Network Transportation (SUNT): A Landmark Spatiotemporal Dataset for Public Transportation. *Scientific Data*, 12: 1320.