A New Hybrid Model for Optimal Bike-Sharing Locations in Montreal

1 INTRODUCTION AND BACKGROUND

In North America, urban mobility has been dominated by private vehicles. However, the current approach is to shift from vehicles to more sustainable transportation modes, such as cycling. Bike-Sharing Systems (BSSs) can help encourage a shift toward cycling. BSSs offer convenience as the primary motivator for cycling through simple payment options, straightforward membership processes, and easy station access (*1*). Station-based BSSs provide ready-to-use bicycles for users, and they can unlock a bicycle at a station and return it to another (or the same) station. A key factor in implementing station-based BSSs is the location of the stations (*2*). Therefore, different optimization models have been developed to optimize the stations' locations, and some of the recent studies are summarized in Table 1.

Table 1 – *A summary of the studies on BSS station location optimization*

Reference	Optimization method	Objective function	Scale (network-time)	Equality	Dynamic demand
Sun et al. (*3*)	MILP	Maximize the user demand	10 locations-2 hours	-	-
Jin et al. (*4*)	Ant colony	Minimize circuit among sites	5 districts-NA	-	-
Frade and Ribeiro (*2*)	XPRESS	Maximize demand coverage	29 zones - a day	-	-
Mix et al. (*5*)	Gurobi Optimizer	Maximize demand covering	14 municipalities - a month	-	MLR
Nikiforiadis et al. (*6*)	Weighted QCP	Maximize the demand; Maximize the need for redistribution	193 candidate stations-4months	-	-
Caggiani et al. (*7*)	Genetic algorithm	Minimise inequalities	3.0km×3.6km-NA	Equal access to 3 bus lines	-

As can be perceived, most of the previous studies considered bike-sharing demand static in location optimization. However, the location of stations and bike-sharing demand are intertwined, and the location of stations can influence the demand. Only Mix et al. (*5*) applied a dynamic demand model in the optimization process, while the dynamic demand prediction model was generated by multiple linear regression, which is generally defined as an inaccurate prediction model due to its predetermined structure (*8*). Further, all the presented studies (except (*6*)) considered a single objective function or applied converted a single objective optimization algorithm (i.e., Weighted QCP) to solve the problem. Hence, an application of multi-objective powerful multi-objective optimization problems has not been considered. Further, the dimension of previous studies was limited to 193 candidate stations and maximizing the demand for 4 months. As another research gap, equality was only considered in (*7*), which was defined as equal access to three bus lines, and equality has not been considered in station location optimization at a city scale. To address these limitations, the novelties and contributions of this study are: (a) Develop an accurate demand prediction model, which dynamically predicts the demand in the station location optimization process; (b) Develop a multi-objective framework to simultaneously maximize bike-sharing use, maximize equality, and minimize implementation cost; (c) Adjusting Reference Vector Guided Evolutionary Algorithm (RVEA) to solve multi-objective BSS location optimization; and (d) Introduce a hybrid machine learning-optimization model that optimizes station locations considering dynamic demand and many objective functions.

2 METHODOLOGY

2.1 Demand prediction

Various factors can influence the demand for BSSs, including the availability of stations, built environment and land use, availability of bikeways, access to public transit, weather conditions, and socio-demographics of dwellers (*1*). Hence, we apply a data fusion approach to merge nine different datasets to include these vital variables in the demand prediction model. The applied data sources include bike-sharing trips (Bixi) in the recent three years, Canada proximity measure data, Walk Score, Open Street Map, deprivation index dataset, Canada Census data, dissemination area information, the Canadian Bikeway Comfort and Safety (Can-BICS), and Canada Weather Stats. Independent variables include the week of the year, day type (categorized as weekday peak hours, weekday off-peak hours, and weekend hours), Walk Score, cycling distance to the city center, and the number of Bixi dock stations within the dissemination area. Additional factors encompass proximity to parks, access to transit stations, the CanBICs index (a measure of nearby bikeways quality and quantity), material and social deprivation indices, household average income, the percentage of French-speaking residents, educational attainment levels, employment rate, population density, total population, and year (noted as 2022, 2023, or 2024). Lastly, environmental conditions include temperature variation, average temperature, average precipitation, and age demographics, specifically the percentages of residents aged below 15 and over 64.

The dependent variable is the number of trips per hour in each dissemination area. A dissemination area is a standard geographic unit in Canada, defined by having at least one neighboring dissemination block, making it the smallest standard geographic unit in Canada (*9*). To predict the hourly demand, a powerful machine learning technique, called Light Gradient Boosting Machine (LightGBM), is employed, and it is tuned by Optuna and k-fold cross-validation.

2.2 Optimization Model

The case study of this investigation is Montreal, Canada, and the current BSS in Montreal (Bixi) includes over 900 stations. There are 787 dissemination areas in the case study, and the model aims to find the optimal number of new stations added to each dissemination area. The model includes three objective functions: maximize bike-sharing use, maximize equality, and minimize implementation cost. The optimization model encompasses 787 decision variables, representing the added stations to each dissemination area. During optimization iterations, new stations are allocated to dissemination areas, and new demand for 22 weeks (early April to early September) is predicted using LightGBM.

The optimization algorithm iteratively maximizes the overall demand. In other words, LightGBM runs inside the optimization algorithm (RVEA) to predict the demand for each solution vector. At the same time, the model maximizes the equality in the network. The equality index is defined as the standard deviation of the number of stations per population in different dissemination areas, which should be minimized. The implementation cost is defined as the number of new stations added to the city, which should be minimized. RVEA is applied to solve the optimization problem. RVEA is a powerful optimization algorithm for multi-objective optimization problems, and it outperforms many multi-objective optimization algorithms such as NSGA-III, GrEA, MOEA/DD, MOEA/D PBI, and KnEA in terms of finding the optimal solution (*10*). Although RVEA is a powerful method, its application to transportation problems has been limited.

3 RESULTS AND DISCUSSIONS

3.1 Demand prediction

The demand prediction model could predict the hourly number of trips in dissemination areas with an MAE of 0.383 trips/hour and an R^2 of 0.971 for testing data. Therefore, LightGBM performed well in predicting the demand, and its application in the optimization process can improve the model's performance.

3.2 Optimal solutions

The results of the optimization model and the number of trips and equality index in the current situation are presented in Table 2. In the current situation, the network includes 977 stations, and the optimization model finds the number and location of new stations added to the model. As shown, the model offers ten non-dominated optimal solutions by adding 4 to 290 stations to the network. To evaluate the effectiveness of the model, a random allocation model is developed, assigning new stations (equal to the number of added stations in optimal solutions) to the network, and its outcomes are also presented in Table 2. In optimal solutions, by increasing the number of stations, the number of trips in the network is increased.

Further, all optimal solutions improve the equality index of the current situation (i.e., reduce the standard deviation of the number of stations per population in different dissemination areas). In the optimal solutions with more added stations, the improvement in the equality index is greater because we need more added stations to ameliorate the unequal distribution of BSS stations and it may be impossible to address this issue with a few new stations. On the other hand, the random allocation model provides us with new solutions, significantly worse than the current situation in terms of the equality index. Moreover, the number of trips in the random allocation model is less than their corresponding optimal solutions.

Table 2 – *The current situation, non-dominated solutions, and randomly allocation models*

ID	Added stations	Number of trips	Equality index $\times\ 10^7$	ID	Added stations	Number of trips	Equality index $\times\ 10^7$
Current	0	9853927	489529				
		Optimal solutions				Randomly allocation	
ID1	4	9937210	489529	Rand1	4	9864850	489529
ID2	10	10022900	489529	Rand2	10	9868774	489529
ID3	13	10045700	489529	Rand3	13	9873256	489576
ID4	19	10073400	489528	Rand4	19	9908014	489528
ID5	70	10270100	489524	Rand5	70	10042673	722155
ID6	78	10302900	489523	Rand6	78	9974865	530811
ID7	86	10326700	489523	Rand7	86	10065608	722162
ID8	89	10309200	489522	Rand8	89	10098447	750707
ID9	284	10861100	489519	Rand9	284	10578648	722173
ID10	290	10876100	489519	Rand10	290	10563738	531238

The improvement of different solutions in terms of increasing the number of trips per added station is shown in Figure 1. As can be perceived, the improvement rate for the random allocation model is between 1200 and 2600 added trips per new station. The improvement rate of the optimization model is much better than the random allocation model. Since the network includes 977 stations and is saturated in some regions, the number of trips per added station is expected to be less than the number of trips per available station. However, ID1 to 4 found optimal locations for new stations with more trips per station than the current rate in the network. From ID 5 to 10, the number of trips per new station is below the current rate, but they increase the overall number of trips. Moreover, they allocate new stations where the inequality is minimized.

The properties of dissemination areas selected by optimal solutions for the new stations are shown in Figure 2. As shown, the optimal solutions with a higher rate of trip increase (e.g., ID 1 to 4) are more likely to select dissemination areas near the center with higher Walk Scores. However, optimal solutions with higher equality improvement (e.g., ID 9 and 10) allocate more stations to locations with higher distance to the center and lower Walk Score because stations are currently more distributed in the city center and these optimal solutions ameliorate this unequal distribution. Moreover, optimal solutions tend to assign new stations where appropriate bikeways are available and are accessible by more people.

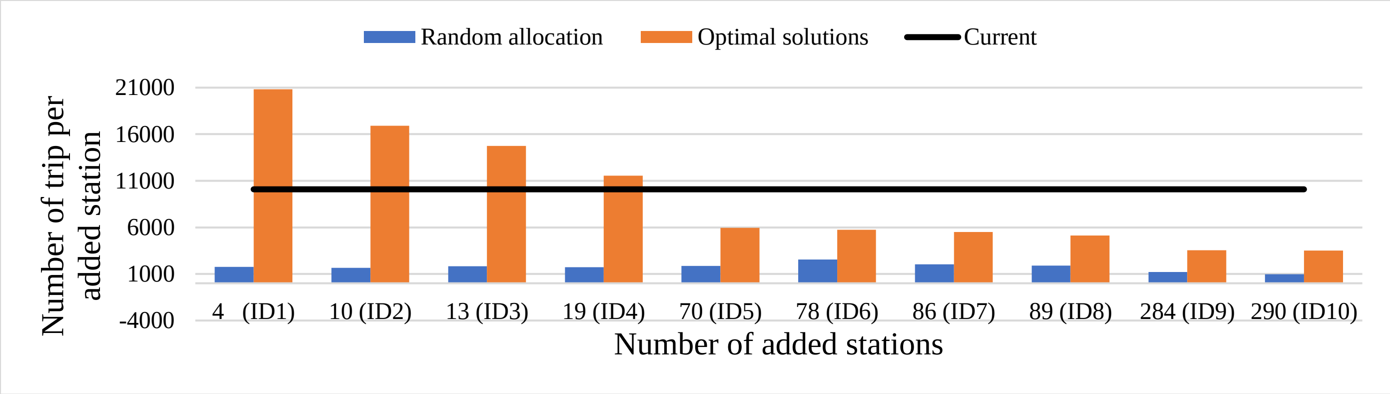

Figure 1 – *The improvement of optimal solutions regarding the number of trips per station*

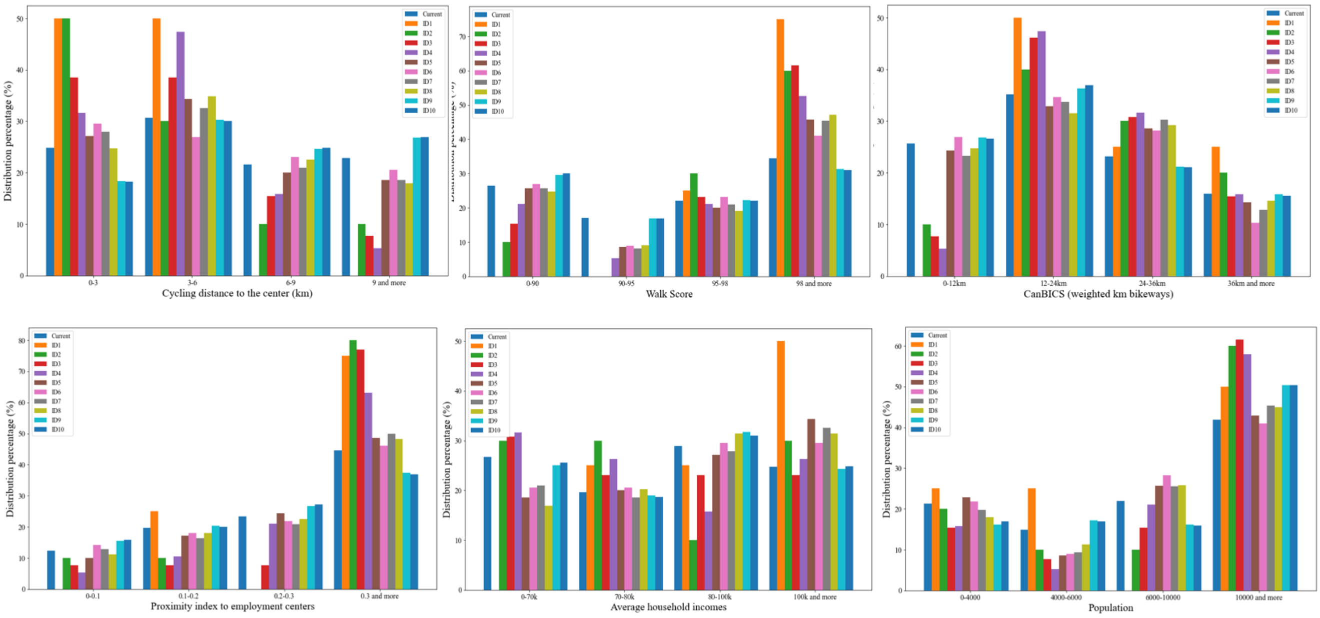

Figure 2 – *The characteristics of optimal dissemination areas for new stations*

References

1. Eren, E., and V. E. Uz. A Review on Bike-Sharing: The Factors Affecting Bike-Sharing Demand. *Sustainable Cities and Society*. Volume 54. https://www.sciencedirect.com/science/article/pii/S2210670719312387?casa_token=FnAAc1Fs_UkAAAAA:h8hlaXtSc07kXhZoog4IAaXZXIO9zwhUE91wpWUNrT_iaRWSFyy6VVM1vte3zUAD5ORjJcB62886. Accessed Nov. 20, 2021.
2. Frade, I., and A. Ribeiro. Bike-Sharing Stations: A Maximal Covering Location Approach. *Transportation Research Part A: Policy and Practice*, Vol. 82, 2015, pp. 216–227. https://doi.org/10.1016/j.tra.2015.09.014.
3. Sun, Z., Y. Li, and Y. Zuo. Optimizing the Location of Virtual Stations in Free-Floating Bike-Sharing Systems with the User Demand during Morning and Evening Rush Hours. *Journal of Advanced Transportation*, Vol. 2019, No. 1, 2019, p. 4308509. https://doi.org/10.1155/2019/4308509.
4. Jin, B., M. H. Hong, X. Lin, and J. M. Sha. Service Sites Selection for Shared Bicycles Based on the Location Data of Mobikes. *IEEE Access*, Vol. 6, 2018, pp. 54965–54975. https://doi.org/10.1109/ACCESS.2018.2872451.
5. Mix, R., R. Hurtubia, and S. Raveau. Optimal Location of Bike-Sharing Stations: A Built Environment and Accessibility Approach. *Transportation Research Part A: Policy and Practice*, Vol. 160, 2022, pp. 126–142. https://doi.org/10.1016/j.tra.2022.03.022.
6. Nikiforiadis, A., G. Aifadopoulou, J. M. S. Grau, and N. Boufidis. Determining the Optimal Locations for Bike-Sharing Stations: Methodological Approach and Application in the City of Thessaloniki, Greece. No. 52, 2021, pp. 557–564.
7. Caggiani, L., A. Colovic, and M. Ottomanelli. An Equality-Based Model for Bike-Sharing Stations Location in Bicycle-Public Transport Multimodal Mobility. *Transportation Research Part A: Policy and Practice*, Vol. 140, 2020, pp. 251–265. https://doi.org/10.1016/j.tra.2020.08.015.
8. Naseri, H., H. Jahanbakhsh, A. Foomajd, N. Galustanian, M. M. Karimi, and E. O. D. Waygood. A Newly Developed Hybrid Method on Pavement Maintenance and Rehabilitation Optimization Applying Whale Optimization Algorithm and Random Forest Regression. *https://doi.org/10.1080/10298436.2022.2147672*, 2022, pp. 1–13. https://doi.org/10.1080/10298436.2022.2147672.
9. Statistics Canada. Dissemination Block (DB), Dictionary, Census of Population, 2021. https://www12-2021.statcan.gc.ca/census-recensement/2021/ref/dict/az/definition-eng.cfm?ID=geo014.
10. Cheng, R., Y. Jin, M. Olhofer, and B. Sendhoff. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation*, Vol. 20, No. 5, 2016, pp. 773–791. https://doi.org/10.1109/TEVC.2016.2519378.