

Optimization of Electric Vehicle Charging Station Location and Capacity in Xuhui District, Shanghai Based on Multi-Source POI Data and Genetic Algorithm

Ruijie Sun

*Qibaodwight High School, Shanghai, China
shanghaisunruijie@163.com*

Abstract. As the number of urban users of electric vehicles continues to grow, the demand for public charging stations also increases in urban areas. This study develops an evidence-based approach to the optimal charging station location and capacity in the Xuhui District of the city of Shanghai. Several data types are incorporated, including residential, office, commercial, transportation, and parking data. Demand was evaluated in small grids and potential charging station sites were public parking lots. A location and capacity optimization model was developed to juggle the search cost for users and the annual cost of the charging station infrastructure. This was subject to limiting the number and location of stations and the capacity for each station. Because this problem was a discrete optimization problem, a genetic algorithm was used. The optimization identified 10 stations from 80 potential locations and a total of 196 chargers. The design of the charging stations with respect to the location and capacity constraints resulted in a service-radius coverage rate of 68.36% and a weighted average search distance of 1.629 km. Compared with an experience-based baseline model, the social cost was estimated to be reduced by 21.63% and the average search distance was reduced by 38.08%. The results show that demand estimation and a genetic algorithm can be used to create a public charging station network for large urban centers.

Keywords: Electric vehicles, Charging station location optimization, Genetic algorithm, POI data, Xuhui District

1. Introduction

The proliferation of electric vehicles necessitates charging stations to support sustainable urban transportation systems. Electric vehicles continue to rise globally, but charging infrastructure is vital for their broader acceptance [1]. In compact urban environments, the sheer number of charging stations may not be the primary concern. A charging network may be wholly ineffective even if stations are located within residential, workplace, and commercial areas if the local charging capacity is fragmented and placed in suboptimal locations.

Planning charging stations involves the balancing of cost and service. Research indicates that the planning of electric vehicle charging stations becomes highly complex when the dual aspects of

coverage and capacity are considered [2]. This trade-off is also evident in the cost of infrastructure and the user travel behavior in the context of public deployment models [3]. Studies conducted at the neighborhood level indicate that travel behavior and the availability of parking also play a role in the planning of service facilities [4]. Additional research indicates that charging infrastructure can be enhanced through a combination of mobility data, parking data, and heuristic optimization [5].

Xuhui District was selected as the study area as it has a diverse set of residential and commercial urban activities, for example, offices, hospitals, schools, transit nodes, and parking facilities. Using Xuhui District as the urban study area, the proposed charging station network optimization integrates POI data with a genetic algorithm to meet the demand defined by urban activities.

2. Data and demand estimation

Data on residential, office, commercial, transport, and parking-lot POIs were collected. POI data are critical because they capture existing physical locations and feature observable instances of urban settlements. These data help represent urban physical structures and the spatial relations and interactions among residents, workers, and visitors. A recent literature review highlights that, for urban management and smart city applications, POI data are critical digital place data depicting real-world urban phenomena [6]. In this study, raw data are collected from a web service interface, and cleansing is done to remove duplicate records, abnormal coordinates, and data points outside the study area. The final demand dataset is comprised of 2,602 valid demand POIs. The parking-lot data provide the initial framework for the selection of candidate stations. The mapping interface was utilized for data access and geocoding [7]. However, optimization and other related tasks were performed using Python.

The study area is composed of 500 by 500-meter grids. For each active grid, four Point of Interest (POI) counts (residential, office, commercial, and transportation) are computed. It is assumed that the charging demand for the central district is generated from home and office parking, shopping and service parking, and transport-related activities. To mitigate the effect of demand surfaces with overly concentrated POI clusters, square-root transformation is applied. The demand score for a grid is expressed as:

$$D_i = 0.42\sqrt{R_i} + 0.28\sqrt{O_i} + 0.22\sqrt{C_i} + 0.08\sqrt{T_i} \quad (1)$$

In relation to Equation (1), R_i , O_i , C_i , and T_i are the residential, office, commercial, and transportation POI counts of grid i , respectively. The weighting is based on the reasoning that residential and work-related parking is more likely to result in stable charging demand as compared to parking related to commercial activities and demand from transit. It is assumed that the total daily charging demand is 3,200 sessions after grid scores are normalized. This is then apportioned to the 209 grids. The aim of this method is not to quantify the charging demand for each grid, but to provide a reasonable demand estimate that reflects the spatial distribution, which is subject to modification with enhanced charging transaction data.

Points of interest (POIs) in public parking lots provide the basis for generating candidate stations. While locations denoted by random coordinates may be valid candidate station locations, parking lots are inherently designed for vehicle access and parking. Hence, the candidate stations are developed by computing a local demand score for all parking lots, which is done by summing the demand for the surrounding cells within an influence radius of 1.5 km. In the context of the genetic algorithm (GA), demand and computational concerns are addressed by selecting the top 80 sites.

This further assists in the elimination of parking lots with low surrounding demand, thus minimizing the unnecessary assessment of locations with low strategic value.

3. Location-capacity optimization model

The first stage of the optimization model establishes which parking lots will be selected as charging station sites, while the second stage identifies how many chargers will be assigned to each of the chosen stations. The need for a balance between the geographical coverage of a charging network and the potential charging capacity of each station means that, without this balance, a model that locates charging stations will fill a coverage area with a number of small charging stations that cannot meet the capacity needs of the surrounding area. Conversely, a model that addresses demand needs by extending the capacity of a limited number of stations will likely increase the average search distance for a charging station.

The distance between the demand grids and the candidate stations is calculated based on the latitude and longitude of the grids and the stations. At the district scale of this preliminary model, the great-circle distance will give a sufficient approximation for calculating distance based on the geographic coordinates. The distance matrix is based on the following expression:

$$d_{ij} = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\text{lat}_i - \text{lat}_j}{2} \right) + \cos(\text{lat}_i) \cos(\text{lat}_j) \sin^2 \left(\frac{\text{lon}_i - \text{lon}_j}{2} \right)} \right) \quad (2)$$

While the estimation method used in this model does not require external routing APIs and is thus both convenient and reproducible, it is likely that the use of geodesic distances will underestimate actual road-network distance or travel time compared to an estimation based on road travel time. The use of geodesic distance computation has been the subject of considerable research [8]. The matrix generated in this model will identify the closest candidate site to each demand grid and will allow the service radius to be evaluated.

The primary goal of the model is to minimize the annual social cost. The annual social cost is comprised of the annual cost of the infrastructure, the travel cost incurred by the users, and penalty terms.

$$\min Z = C_{\text{infra}} + C_{\text{user}} + P_{\text{coverage}} + P_{\text{capacity}} + P_{\text{site}} \quad (3)$$

The fixed annual cost related to each chosen station, charger installation expenses, and periodic maintenance expenses make up the infrastructure expense. The cost of user travel is related to the anticipated daily travel, average travel distance, and the distance-cost coefficient. Demand-related penalties include exceeding station demand, exceeding the maximum number of selected stations, and demand-related penalties for service requirement shortfalls. The model presented in this paper indicates the cost of infrastructure and the cost of user access, which need to be balanced for the planning of charging stations to be effective [9].

The primary limitations include the maximum number of stations set to 12, each chosen station having a maximum of 26 and a minimum of 4 chargers, and a maximum service radius of 1.8 km. Each of the chargers is assumed to serve 20 charging sessions per day. Users of each service grid are assumed to be assigned to the closest chosen station. These assumptions are not designed to be fixed universal values. These serve the purpose of creating a stable environment for testing in order to facilitate the comparison of the optimized and baseline plans.

4. Genetic algorithm design

A genetic algorithm is appropriate for this problem due to the presence of a complicated discrete and nonlinear solution space. The problem also involves 80 candidate locations, along with a variable number of chargers for each location, which would make a complete enumeration of all possible solutions exceedingly inefficient. Genetic algorithms simulate the process of natural selection and were initially proposed as adaptive search methods [10]. Now, they have a well-established role in dealing with a variety of complex optimization problems, both in engineering and management [11].

In this particular problem, a chromosome is an integer vector of a length equal to the number of candidate locations. In this case, a gene value of zero means the location was not selected, and in contrast, a positive gene value means the location was selected, along with the number of chargers. For example, the chromosome [0, 12, 0, 8, 20] means location 2 is selected with 12 chargers, location 4 is selected with 8 chargers, and location 5 is selected with 20 chargers. This type of representation is a form of flexible or mixed encoding and allows the model to evolve both the location and the capacity within the same chromosome.

The fitness function is based on the objective value of Equation (3). Relating to the fitness function, a smaller value represents a better solution since the problem is defined as cost minimization. The role of the penalty term is important as the initial stages of the process may utilize infeasible solutions, and these may represent useful structural information. Rather than eliminating all infeasible chromosomes from the population, the penalty function helps to focus the population toward feasible and efficient layouts while still allowing the exploration of the solution space to a degree.

The methodology also incorporates tournament selection, with a greater chance of chromosome reproduction going to stronger chromosomes, yet still keeping variety in the population. Uniform crossover relates to the transfer of location and capacity content between the two chromosomes. Mutation also provides the opportunity to close a site or change the number of chargers at a site. The best chromosomes are retained, and the population of 120 chromosomes is kept over 150 generations, with the best chromosomes (elites) being 6. The rates for crossover and mutation are 0.8 and 0.08, respectively. In the current model, a single aggregated fitness function is utilized. However, in future work, multi-objective evolutionary algorithms (e.g., NSGA-II), which are able to distinguish between the cost, accessibility, equity, and utilization objectives [12], can be implemented.

5. Results and discussion

According to the proposed optimal plan, 10 out of the 80 candidate parking lots will be developed into charging station lots with 196 chargers. The proposed plan predicts a service-radius coverage of 68.36% and a weighted average search distance (WASD) of 1.629 km. The proposed plan indicates that the annual cost of the required infrastructure will be 559.784 ten-thousand RMB and the annual user travel cost will be 570.947 ten thousand RMB. The proposed plan will result in an annual total social cost of 1,130.731 ten thousand RMB. Locations of high-capacity charging stations propose 2 sites with 26 chargers, 4 sites with 24 chargers, and 1 site with 23 chargers. Candidate IDs are used in the text and tables in lieu of local site names, which assists English readers and avoids local place names in the text.

Main inputs and assumptions of the experiment are shown in Table 1. Based on the provided information, it seems that the model is more of a planning simulation than a real-time operating system. In order to distribute the relative demand across the grids, an assumption of 3,200 sessions

per day was made. This assumption can be changed in the upcoming scenario analyses without affecting the integrity of the model.

Table 1. Core data and model configurations

Parameter	Configuration
Demand POIs after cleaning	2,602
Demand grids	209
Parking-lot candidates kept for GA	80
Estimated charging sessions	3,200
Grid cell dimensions	0.5 km by 0.5 km
Service radius	1.8 km
Maximum station count	12
Daily sessions per charger	20
GA population size and generations	120 and 150

Figure 1 provides a spatial representation of demand grids, candidate parking lots, and selected charging stations. The optimization strategy results in the majority of charging stations being deployed in the central and eastern demand clusters, and one charging station being deployed in the western demand cluster. In addition to demand, the algorithm considers the number of charging stations, their capacity, and their service radius, which explains why charging stations are not deployed only in the densest demand clusters. Candidate parking lots in high-demand areas compete with one another; consequently, the final network is not constrained to a set of adjacent candidate parking lots.

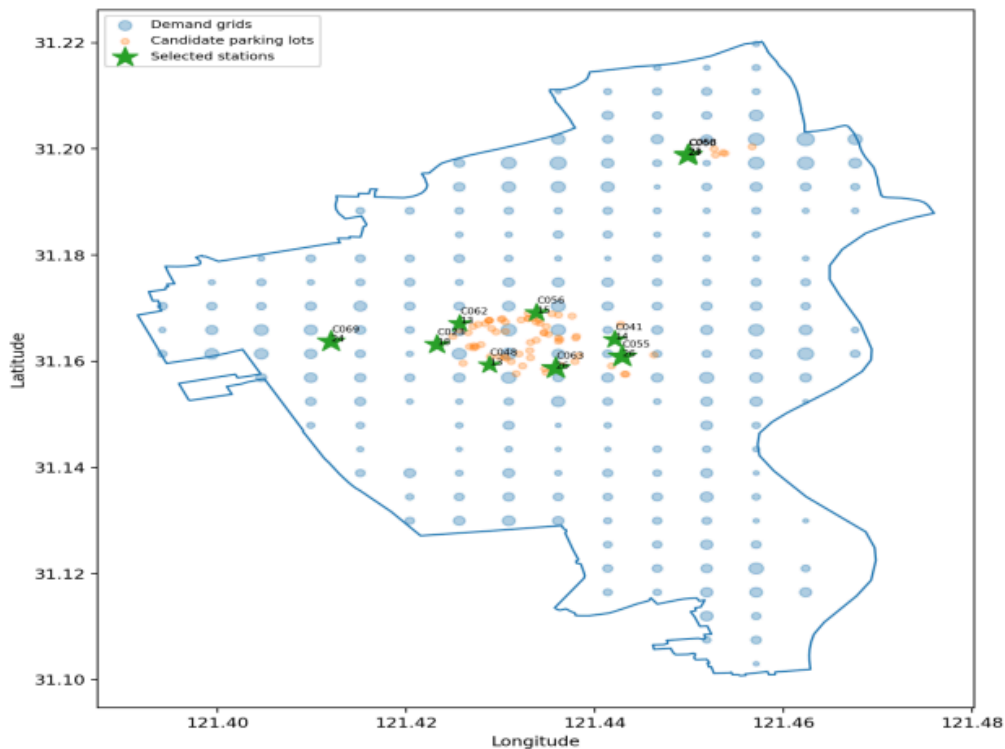


Figure 1. Optimized EV charging station layout in Xuhui district

Figure 2 illustrates convergence of the algorithm. During the earlier generations, the objective value for the best solution in the population exhibits a rapid improvement; however, over time, the improvement becomes incremental. This phenomenon is characteristic of many optimization algorithms. In the initial generations of an algorithm's execution, the solution population traverses a larger portion of the search space. As a consequence, the initial generations identify the more cost-efficient groups of solution components. In the subsequent generations, the population primarily fine-tunes the components of the identified groups. Based on the results of this study, 150 generations are sufficient when considering the specific boundaries of this study, as most solution improvements are observed in the initial generations.

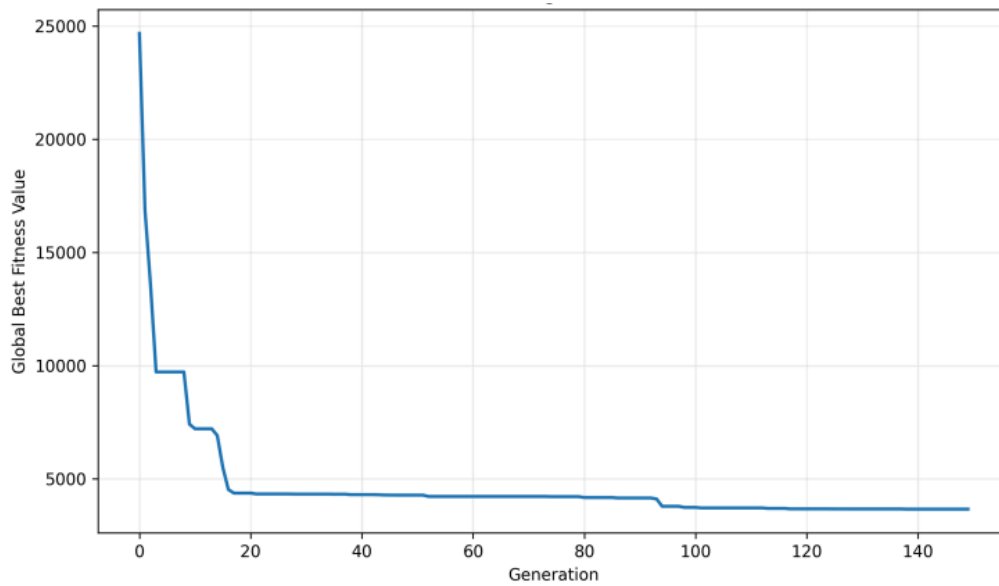


Figure 2. Genetic algorithm convergence curve

Table 2 presents a comparison between the GA-optimized plan and the plan based on experience. The baseline plan makes use of local-demand candidate sites to select stations. Consequently, this is an example of a sensible, although unoptimized plan. Compared to the baseline plan, which has 12 stations, the GA-based plan has 10 stations but contains more installed chargers (196 vs. 160). Thus, the GA plan allocates capacity more intensively. The optimized plan achieves 68.36% service-radius coverage, increasing from 32.71%, and the weighted average search distance diminishes from 2.631 km to 1.629 km. While the total annual infrastructure cost increases slightly as a result of a higher number of chargers, the total user travel cost diminishes significantly. Consequently, the total annual social cost decreases from 1,442.718 ten-thousand RMB to 1,130.731 ten-thousand RMB, a reduction of 21.63%.

Table 2. GA-optimized plan compared to baseline

Metric	GA-optimized	Baseline
Selected station count	10	12
Total chargers	196	160
Coverage rate	68.36%	32.71%
Weighted average search distance	1.629 km	2.631 km
Annual infrastructure cost	559.784 ten-thousand RMB	520.713 ten-thousand RMB

Table 2. (continued)

Annual user travel cost	570.947 ten-thousand RMB	922.005 ten-thousand RMB
Annual total social cost	1,130.731 ten-thousand RMB	1,442.718 ten-thousand RMB
Social cost reduction	21.63%	—
Average distance reduction	38.08%	—

Figure 3 depicts the results of the sensitivity analysis for demand multipliers. Using a multiplier of 0.75, the social cost with 72.17% coverage is 1,040.751 ten thousand RMB. A multiplier of 1.00 means the cost is 1,234.126 ten thousand RMB with 66.79% coverage. A multiplier of 1.25 means the cost is 1,775.503 ten thousand RMB and overload sessions are recorded. This suggests that the network remains stable under moderate demand. As the charging demand rises significantly without adding candidate sites and/or charging capacity, the network will likely be negatively impacted. The results of the sensitivity analysis are to be interpreted as scenario tests and should not replace the results from the primary 150-generation GA result.

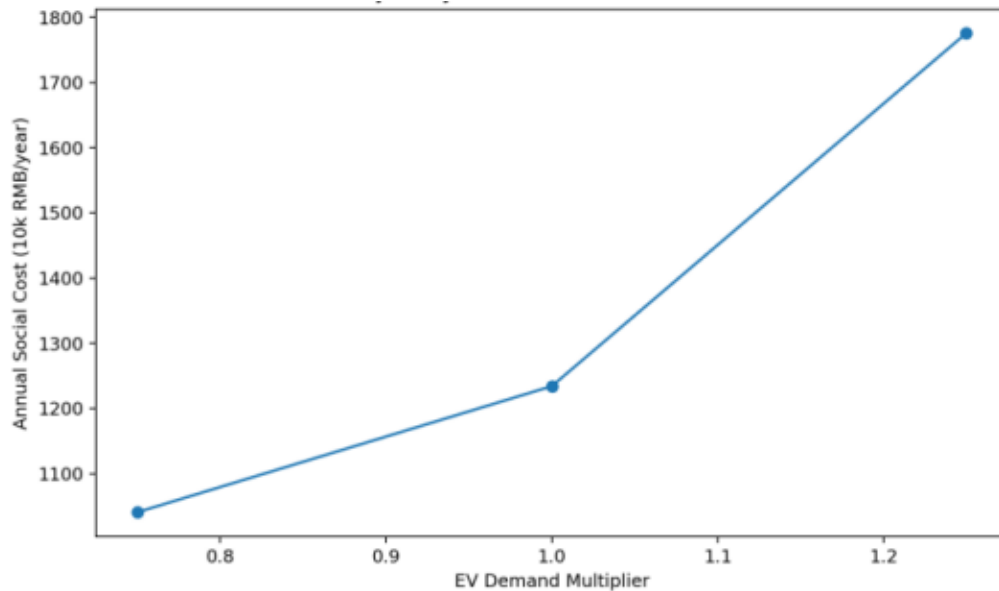


Figure 3. Sensitivity analysis under different demand levels

Three major implications arise from these findings. First, at the sub-district level, POI grid-based demand systems can identify demand clusters. From a spatial planning perspective, this system can identify potential charging demand much better than the average demand at the sub-district level. Second, location and capacity are interdependent elements of the charging station planning process. Location choices can improve coverage and accessibility, but only if the charging stations are sufficiently sized. Third, the reduction of the social cost primarily stems from the reduction of user travel cost, not the reduction of the cost of the infrastructure. This means that the additional infrastructure cost required to cover search distance and demand is smaller than the saving in user travel costs.

The current system has some limitations. This system is designed based on the estimated charging demand, therefore, the outcomes of this design can only serve as a basis for planning, not as direct construction decisions. Among other things, the distance matrix based on the Haversine formula does not account for circuitous trips through the street network, congestion, delays in access

and parking, and roadway circulation. The ability of the roadway to accommodate a charging station is not modeled and may influence the selection of a charging station site. Future research can integrate actual charging demand, travel time on the roadway, circulation delays, queuing theory, and the power distribution grid. A multi-objective framework can be used to incorporate trade-offs between cost, accessibility, demand, equity, and grid stress.

6. Conclusion

This research includes a dataset-driven optimization framework to establish charging location and capacity for electric vehicles in Xuhui District, Shanghai. This work incorporates multiple types of POI data, a grid-based demand estimate, parking-lot candidate sites, a distance matrix, and a genetic algorithm. This workflow can be applied to urban centers where charging data is not available, but data is available for spatial activities.

Results show how the demand estimation using POI can show spatial charging demand on a smaller scale in a central district. The solution from the genetic algorithm placed ten stations with 196 chargers to reach 68.36% service-radius coverage and resulted in a weighted average search distance of 1.629 km. When compared to the experience-based baseline, the solution selected station locations to reduce the estimated social cost by 21.63% and the average search distance by 38.08%. This outcome also indicates that the cost-and-user-access-based heuristic optimization method is better than the simple location-based solution.

This work also shows that the location of the charging station and capacity are optimally planned together. Networks of stations may perform well, yet be poorly planned if capacity is overdispersed. The solution supports an investment of charging infrastructure in a planned and balanced manner for the service area.

While useful, the current model is limited in several respects. These include the use of proxy demand driven by points of interest, assumed cost and coordinate-based distance, and a number of other simplifications. Future iterations of this model should incorporate real charging transaction data, travel times on road networks, the state of parking supply, shifting traffic patterns, and other constraints on the power supply. Additionally, further refinements to the model could incorporate multi-objective algorithms that would allow for the separate evaluation of trade-offs between economic cost, user accessibility, environmental benefit, and spatial equity. Ultimately, the proposed framework for metropolitan charging infrastructure could be a useful tool for model-based decision support for planning and policy at the metropolitan level.

References

- [1] International Energy Agency. (2024). Global EV outlook 2024. *IEA*, Paris. <https://www.iea.org/reports/global-ev-outlook-2024>
- [2] Lam, A. Y. S., Leung, Y. W., & Chu, X. (2014). Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Transactions on Smart Grid*, 5(6), 2846–2856. <https://doi.org/10.1109/TSG.2014.2344684>
- [3] He, F., Wu, D., Yin, Y., & Guan, Y. (2013). Optimal deployment of public charging stations for plug-in hybrid electric vehicles. *Transportation Research Part B: Methodological*, 47, 87–101. <https://doi.org/10.1016/j.trb.2012.09.007>
- [4] Frade, I., Ribeiro, A., Goncalves, G., & Antunes, A. P. (2011). Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. *Transportation Research Record*, 2252(1), 91–98. <https://doi.org/10.3141/2252-12>
- [5] Vazifeh, M. M., Zhang, H., Santi, P., & Ratti, C. (2019). Optimizing the deployment of electric vehicle charging stations using pervasive mobility data. *Transportation Research Part A: Policy and Practice*, 121, 75–91.

- [6] Sun, K., Hu, Y., Ma, Y., Zhou, R. Z., & Zhu, Y. (2023). Conflating point of interest (POI) data: A systematic review of matching methods. *Computers, Environment and Urban Systems*, 105, 101977.
- [7] Amap Open Platform. (2026). Web service API documentation: Place search and district search. *Amap Open Platform*. <https://lbs.amap.com/api/webservice/guide/api/search>
- [8] Karney, C. F. F. (2013). Algorithms for geodesics. *Journal of Geodesy*, 87, 43–55. <https://doi.org/10.1007/s00190-012-0578-z>
- [9] Kuby, M., & Lim, S. (2005). The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, 39(2), 125–145. <https://doi.org/10.1016/j.seps.2004.03.001>
- [10] Holland, J. H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor.
- [11] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading.
- [12] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. <https://doi.org/10.1109/4235.996017>