

# Spatiotemporal variability and prediction of e-bike battery levels in bike-sharing systems

Aleix Bassolas<sup>1,\*,+</sup>, Jordi Grau-Escalano<sup>1,2,+</sup>, and Julian Vicens<sup>1</sup>

<sup>1</sup>Eurecat, Centre Tecnològic de Catalunya, Barcelona, Spain

<sup>2</sup>Universitat Politècnica de Catalunya, Barcelona, Spain

\* aleix.bassolas@eurecat.org

<sup>+</sup>these authors contributed equally to this work

September 23, 2024

## Abstract

Bike Sharing Systems (BSSs) play a crucial role in promoting sustainable urban mobility by facilitating short-range trips and connecting with other transport modes. Traditionally, most BSS fleets have consisted of mechanical bikes (m-bikes), but electric bikes (e-bikes) are being progressively introduced due to their ability to cover longer distances and appeal to a wider range of users. However, the charging requirements of e-bikes often hinder their deployment and optimal functioning. This study examines the spatiotemporal variations in battery levels of Barcelona's BSS, revealing that bikes stationed near the city centre tend to have shorter rest periods and lower average battery levels. Additionally, to improve the management of e-bike fleets, a Markov-chain approach is developed to predict both bike availability and battery levels. This research offers a unique perspective on the dynamics of e-bike battery levels and provides a practical tool to overcome the main operational challenges in their implementation.

## Introduction

By 2022, more than 2,000 Bike Sharing Systems (BSSs) were available worldwide, with approximately 9 million bikes in circulation [33]. These systems enhance the efficiency of public transport by improving access to stations [35, 25], which in turn reduces car usage and alleviates traffic congestion [32, 18]. Moreover, the increased use of bikes and the resulting shift in transport modes have been shown to positively impact public health by promoting physical activity [37], benefit the environment by reducing pollutant emissions [13], and lead to economic savings [39].

The recent increase in data availability related to BSSs has led to greater interest in understanding the mobility patterns within these systems [28, 29]. Studies have shown that stations with strong flows tend to be located nearby [11], and they often display a certain level of mesoscopic organisation within communities [53]. Additionally, the development of models to understand the usage of micro-mobility [46] has improved understanding of factors such as topography [26], day of the week [46], weather conditions [16, 27, 5], and the built environment [46]. Furthermore, efforts have been made to develop predictive models that enhance the management of BSSs infrastructures

efficiently and guide system expansion through machine learning and probabilistic approaches [42, 30, 10, 3, 51, 4]. In this line, one of the major challenges in optimising BSS performance is ensuring bike availability, as stations often experience significant flow imbalances [52, 43, 17]. To address this, models focused on optimising rebalancing operations have been developed [38, 14, 15, 12].

In recent years, BSSs have undergone progressive electrification [33, 20], with a sharp increase in the number of e-bikes compared to their mechanical counterparts [20, 7], as e-bikes allow for longer trips and attract a broader user base [41, 49]. E-bikes have been shown to be less sensitive to factors such as longer distances, adverse weather conditions, and casualties [48, 9, 40]. Despite these advantages, e-bikes present additional challenges to maintaining bike availability, primarily due to low battery levels, battery malfunctions, and complications arising from high temperatures [24, 19, 54, 47]. Moreover, the scarcity of battery data at the system level complicates the development of predictive models. While several studies have implemented frameworks to predict the battery consumption of individual bikes during trips [8, 45], most approaches at the BSS level focus on the station-level granularity [55].

This publication addresses the challenge that the state of charge poses to e-bike availability by analysing mobility and battery level data from Barcelona’s BSS (Bicing) and proposing a model that predicts future bike locations with their corresponding battery levels. The structure of the paper is as follows: first, the spatial and temporal variability in average battery levels is analysed to identify regions more vulnerable to battery depletion. Next, a Markov chain approach is implemented to predict the future locations of bikes. Subsequently, the parameters in the equations describing battery power consumption are optimised. Finally, by combining the models presented in the two previous sections, a framework is provided to support the management of e-bike fleets by projecting future scenarios.

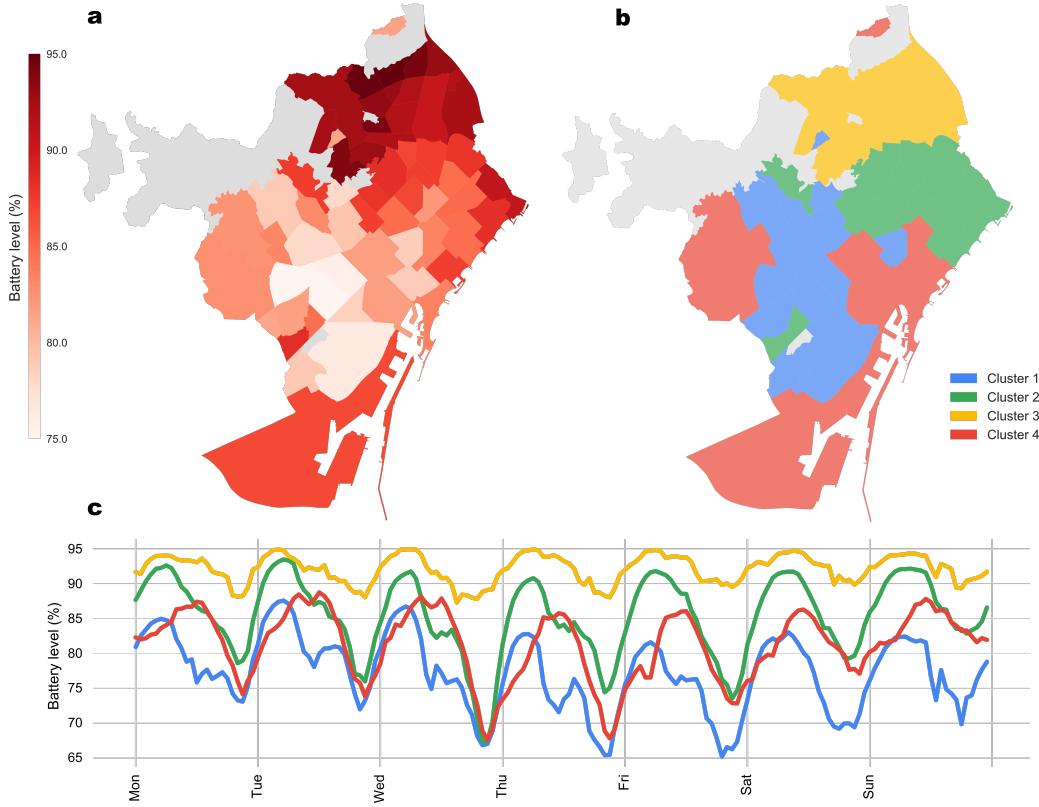
## Results

### Assessing e-bikes’ battery levels and usage patterns

This research analyses the bike usage and battery levels of Barcelona’s BSS [44, 50, 22], using data supplied by its operator, Bicing, which includes trip and battery information:

- The mobility data spans between April 2019 and December 2022, comprising 53 million trips, of which 62% were performed with m-bikes and 38% with e-bikes. Each trip record includes the station of origin and destination, the start and end date of the trip, a bike identifier, the bike model (m-bike or e-bike), and an anonymised user identifier [22]. In this work, we focus specifically on the mobility of e-bikes.
- To analyse e-bike battery consumption, two datasets were used: one with snapshots taken every 8 hours over the entire year of 2022, and another with snapshots taken every 30 minutes over three weeks from 29 March to 18 April 2023 (Figures S1 and 3). The relevant fields in both datasets include percentage, voltage, current, and temperature of the battery, bike station, timestamp, bike status (operative, low battery, or inoperative for several reasons such as maintenance), and location (at station, in use, or at the workshop). A detailed analysis of these datasets is provided in Sections S1 and S2.

Additionally, trip data from March and April 2023 was available to assess battery consumption in the 30-minute dataset.



**Figure 1: Analysis of average battery levels and temporal clustering.** **a.** Average battery levels across neighbourhoods. The average for each neighbourhood is computed using the battery values of all the bikes in that neighbourhood. **b.** Clustering of the neighbourhoods according to the temporal pattern of battery levels using K-means into four distinct clusters. **c.** Temporal series of average battery levels for each cluster. The neighbourhoods in grey do not have bike-sharing stations.

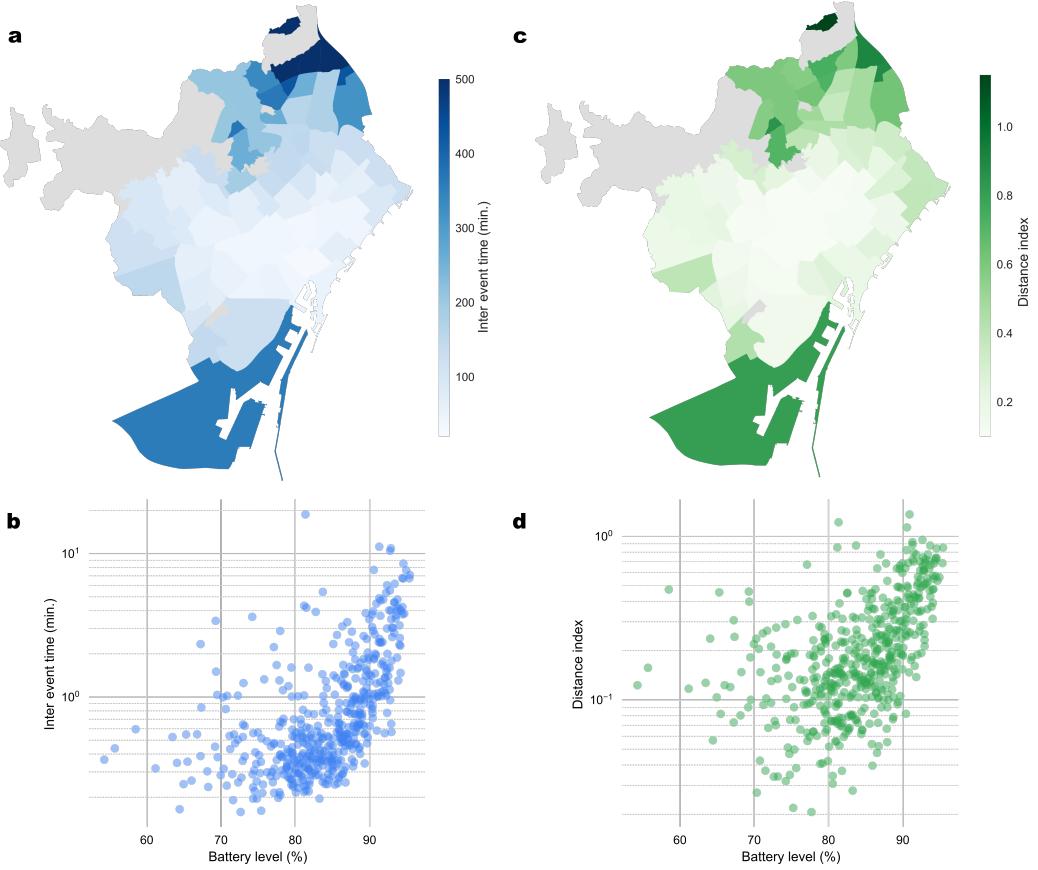
The average battery levels per neighbourhood show significant variability across neighbourhoods (Figure 1a, Figure S3). To facilitate the analysis of all the temporal series of average battery levels derived from each neighbourhood, K-means clustering was performed, resulting in four distinct clusters (Figure 1b, Figure S4). The central and southern neighbourhoods, which fall into Clusters 1 and 4, have the lowest battery levels, exhibiting steep declines in battery levels during business hours with recovery during nighttime (Figure 1c). This is likely due to the high levels of industrial and commercial activity, as well as the dense populations in these areas. Additionally, it is worth noting that some neighbourhoods in Cluster 4 have slower charging speeds at night, likely due to a stronger nightlife presence. In contrast, northern neighbourhoods in Clusters 2 and 3 are mainly residential with less commercial activity, resulting in higher and more stable battery levels throughout the day.

One factor directly related to average battery levels is the average time between trips for each bike, which can also be referred to as inter-event time. With longer inter-event times, e-bikes have more opportunity to recharge, resulting in higher battery levels. In Figure 2a, central and southern neighbourhoods (Clusters 1 and 4) have not only lower average battery levels but also shorter inter-event times, indicating a higher turnover of e-bikes due to frequent trips. This high turnover aligns with the higher density of users and increased mobility demand in commercial and densely populated areas. However, Zona Franca, located in the southernmost area of the city, presents a distinct case due to its industrial nature and significantly low population density (1 inhabitant per hectare [1]), which results in low e-bike usage. Conversely, northern neighbourhoods (Clusters 2 and 3) exhibit longer inter-event times, reflecting less frequent e-bike usage and consequently higher and more stable battery levels despite some of them being in high altitudes. Additionally, a Pearson correlation analysis was conducted between average battery levels and inter-event time at the station level (Figure 2b), revealing a correlation value of 0.39. This moderate positive correlation suggests that stations with longer inter-event times tend to have higher average battery levels, further supporting the neighbourhood-level results.

Apart from the inter-event time, two additional factors can affect the average battery levels at a station: the origins of the trips that the station receives and the number of trips coming from each origin. Trips originating from farther stations tend to deplete the battery more. Therefore, stations that receive a greater number of trips from distant origins will typically have lower average battery levels compared to those receiving trips from nearby stations. To quantify the impact of trip distances on battery levels, a distance index ( $DI_j$ ) was calculated for each station  $j$ . This index accounts for both the distance of each trip and the frequency of trips from each origin station. Specifically, the distance of each trip ( $d_{i,j}$ ) from station  $i$  to station  $j$  is weighted by the inverse of the number of trips from station  $i$  ( $n_i$ ). These weighted distances are then summed for each destination station  $j$  and normalised by the total number of trips received by station  $j$  ( $\sum_i T_{ij}$ ) to account for the overall activity level at the station:

$$DI_j = \frac{\sum_i \frac{d_{i,j}}{n_i}}{\sum_i T_{ij}} \quad (1)$$

The distance index was computed for all stations, and averaged over each neighbourhood (Figure 2c). As shown, neighbourhoods in the northern areas tend to have a higher distance index, indicating that trips to these neighbourhoods often originate from farther stations and that there are fewer incoming trips overall. This combination results in higher average battery levels. Conversely, neighbourhoods in the central and southern parts of the city exhibit a lower distance index, suggesting that trips to these areas typically come from nearby stations and that there are more



**Figure 2: Analysis of inter-event times and distance index for e-bikes.** **a.** Average inter-event time between trips across neighbourhoods. The average for each neighbourhood is computed using the inter-event time values of all bikes, with outliers removed using the IQR method at the station level. **b.** Relation between average battery levels and inter-event time at the station level, with a Pearson correlation of 0.39 ( $p\text{-value} < 0.001$ ). **c.** Distance index per neighbourhood, calculated by first computing the index at the station level to accurately reflect trip distances between stations, and then averaging it at the neighbourhood level. **d.** Relation between average battery levels and the distance index at the station level, with a Pearson correlation of 0.47 ( $p\text{-value} < 0.001$ ). The neighbourhoods in grey do not have bike-sharing stations.

incoming trips, leading to lower battery levels. Zona Franca, as previously mentioned, deviates from this trend due to its unique industrial nature and low population density. The Pearson correlation between the distance index and average battery levels at the station level revealed a value of 0.47 (Figure 2d), indicating a moderate positive correlation. This suggests that stations farther from others and receiving a lower number of trips tend to have higher average battery levels. Additionally, no significant collinearities were found between the inter-event time and the distance index (Figure S5 and Table S1).

## Predicting e-bike locations with a Markov chain-based approach

To predict bike locations, a probabilistic model based on a Markov chain approach has been developed [31, 23]. In this model, each bicycle of the system is modelled independently, with transitions based on the probability of movement from its current station. Given a starting location for a bike and a forecasting horizon  $\Delta t$ , the algorithm follows the sequence below:

- An inter-event time is sampled from the distribution  $I_i^{w,h}$  to simulate the resting time of a bike given a station  $i$  on a specific day type  $w$  and hour  $h$ . The days of the week are grouped into four categories: Working days (Monday to Thursday), Friday, Saturday, and Sunday. The distribution  $I_i^{w,h}$  is calculated by determining the time elapsed between a bike's arrival at station  $i$  and its next departure.
- If the inter-event time exceeds the desired  $\Delta t$ , the process stops, and the final location of the bike will correspond to its current station.
- If the forecasting horizon is greater than the inter-event time, a destination  $j$  is sampled based on the probability  $M_{ij}^{w,h}$ , given a weekday  $w$ , hour  $h$ , and origin station  $i$ . This destination then becomes the new location of the bike, and the average travel time between the station pairs,  $\tau_{ij}$ , is elapsed. The destination probability is calculated using the number of trips  $T_{ij}^{w,h}$  between  $i$  and  $j$  from the 2022 trip data, as follows:

$$M_{ij}^{w,h} = \frac{T_{ij}^{w,h}}{\sum_{\forall j} T_{ij}^{w,h}}, \quad (2)$$

- If the total time elapsed after completing the trip exceeds the desired  $\Delta t$ , the final predicted location will be station  $j$ . Otherwise, the process starts again.

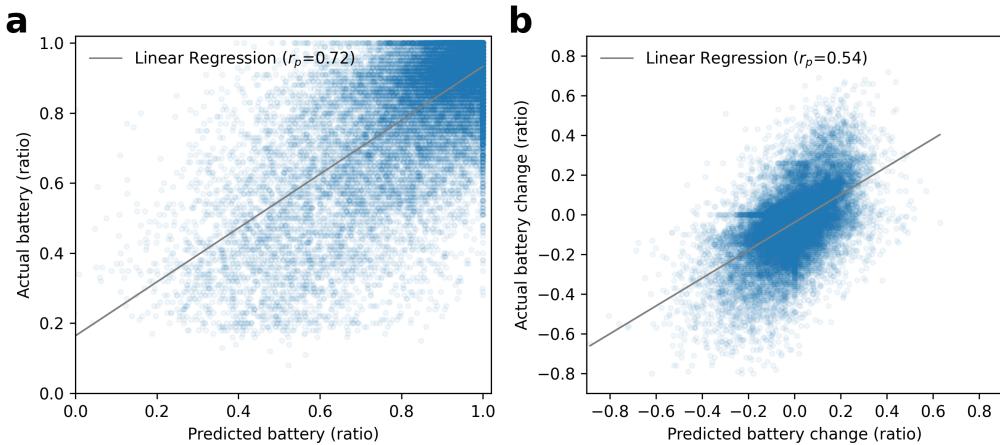
For each bike, 200 realizations are performed to capture a representation of all possible scenarios. The main result is the probability of ending up at station  $j$  after a time period  $\Delta t$ , denoted as  $f_{i,j}^{w,h}(\Delta t)$ . This probability is calculated as the number of times a bike ends up at station  $j$ , divided by the total number of iterations. To aggregate the results at the system level and determine the number of bikes available per station, the probabilities  $f_{i,j}^{w,h}(\Delta t)$  are summed for each station  $j$ , which may result in non-integer values for the number of bikes at each destination.

## Modelling e-bike battery levels

To model the battery levels of bikes, a linear increasing term was applied when docked, and a linear decreasing term during trips. The main variables available to compute battery consumption

during a trip are its duration and the distance between the origin and destination stations. Battery consumption is calculated by multiplying the trip duration by the instantaneous power consumption. To find the optimal values for the coefficients in the power consumption equation for individual e-bikes [8, 45], as well as the charging speed, the Optuna Framework was employed [2] (see Methods). The 30-minute dataset was split into 75% for training and 25% for validation.

The fit between the observed and expected battery levels, using the optimal parameters, is shown in Figure 3. In Figure 3a, the predicted battery percentage is displayed, while in Figure 3b, the predicted battery variation is presented. Despite the dispersion, a significant correlation is observed, with values closely aligned to the diagonal. Factors such as user heterogeneity, missing route information, or bikes not charging while docked could contribute to the dispersion. The parameters were also validated using the 8-hour dataset, with similar correlation values, as shown in Figure S6 (Section S3).



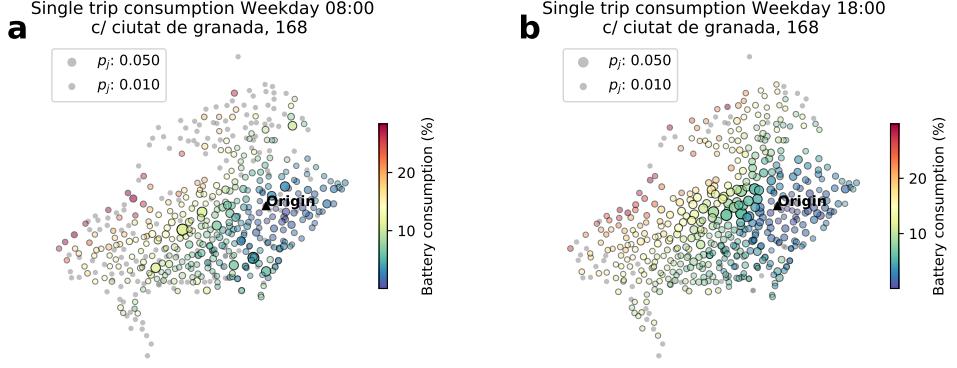
**Figure 3: Battery model predictions for the validation set of the 30-minute dataset.** **a.** Observed battery level as a function of the predicted values. **b.** Battery variation observed as a function of the predicted values. Each point corresponds to one battery-level observation. The Root Mean Squared Error (RMSE) is 0.0178 and the Pearson correlation coefficient is significant ( $p\text{-value}<0.001$ ) in both cases. The linear fit is shown in grey.

This battery model can have two key applications. The first enables the assessment of battery consumption and destination probabilities for individual trips. The second integrates the Markov chain-based approach with the battery model to predict both e-bike locations and battery levels across the entire fleet.

### Single trip application

Evaluating short-term scenarios for a single bike can be achieved by coupling the battery model with the distribution of possible destinations  $P_{ij}^{w,h}$ . In Figure 4, given *St. Ciutat de Granada, 168* as the origin station, the probability of arriving at each destination is represented by the size of the dots, while the battery consumption for each bike trip is indicated by the colour. The probability

of destinations changes significantly between 8:00 and 18:00 despite the same origin station. At 18:00, trips tend to be shorter and head towards nearby stations, while morning trips are longer and converge towards a few work-related hot spots. Additionally, battery consumption varies notably based on the elevation changes along the route, with trips going downhill consuming less battery over the same distance. The altitude of each station is provided in Figure S7 [34].



**Figure 4: Single trip location and battery prediction.** Probability of finding a bike at a station and battery consumption of the trip when departing from the *St. Ciutat de Granada, 168* station on a weekday at **a.** 08:00 and **b.** 18:00. The origin station is highlighted with a triangle. The size of the dots corresponds to the probability of having that station as the destination and color corresponds to the battery consumed by the bike. The grey points correspond to destination stations with no trips.

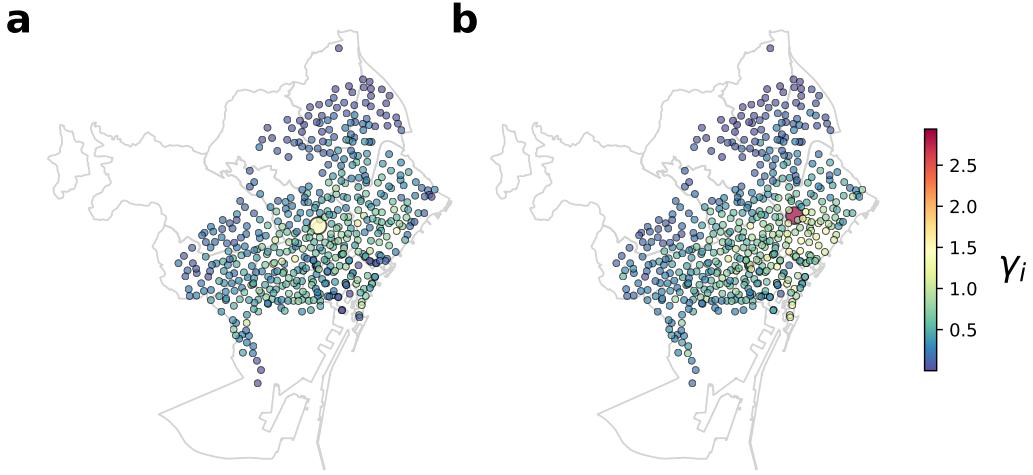
Furthermore, an impact index can be computed to assess the influence of each origin station on battery levels. This index depends on the probability of each destination given an origin station, the battery consumption of each possible trip, and the average hourly trips departing from station  $i$  to other stations. The impact index can be written as:

$$\gamma_i^{w,h} = \langle T_i^{w,h} \rangle P_{ij}^{w,h} \Delta B_{ij} \quad (3)$$

where  $\langle T_i^{w,h} \rangle$  represents the average number of trips departing from station  $i$ ,  $P_{ij}^{w,h}$  is the probability that a trip generated in station  $i$  ends in station  $j$ , and  $\Delta B_{ij}$  is the battery consumed by a trip from  $i$  to  $j$ . An example of the impact index  $\gamma$  at 8:00 and 18:00 on a regular weekday is shown in Figure 5. Due to the variation in trip patterns captured by  $P_{ij}^{w,h}$ , the stations with the highest impact change throughout the day.

### Coupling the Markov chain and the battery models

The adjusted battery consumption function can be integrated into the Markov chain approach to provide joint predictions of both location and battery levels over multiple trips, either at the single-bike or BSS level. Inter-event times are used to calculate the charging percentage, while the distance and duration of trips are used to calculate energy consumption. To validate this approach,



**Figure 5: Impact index of the stations at different hours.** Impact index of the stations at **a.** 08:00 and **b.** 18:00. The station with higher impact at each hour is highlighted with a larger dot size. In the morning, the *St. Provença, 125* station has the highest impact, while in the afternoon, *St. Ciutat de Granada, 168* has the greatest impact.

systematic simulations were performed using the initial conditions from the 2023 mobility data and 30-minute battery data (Figure 6). This included the average number of available bikes and average battery levels per station. Simulations were performed across various initial scenarios and time horizons to demonstrate the capabilities of the coupled model. Additionally, to prevent data leakage, the probabilities of the Markov model were based on data from the year 2022.

First, it is possible to compute the probability of finding a bike at a given destination and its battery level as a function of its origin, following one or more trips within a time period  $\Delta t$ . The final battery levels are determined by averaging over all the Markov iterations. As an example, Figure 7 shows the probability and battery levels of bikes departing from the station at Glòries (*St. Ciutat de Granada, 168*) at 08:00 after **a.** 30 and **b.** 120 minutes. The role of distance from the origin is evident in the short-term predictions, with lower battery levels at greater distances and inclinations. However, this effect diminishes over longer time windows, as multiple trips can occur before reaching a station. For longer periods, a greater variety of final destinations is also possible.

As a second application, the Markov chain approach allows to transition from bike-level to system-level predictions. To calculate the number of available bikes per station across the entire fleet, the probabilities  $f_{i,j}^{w,h}(\Delta t)$  for each destination station  $j$  are summed, which may result in non-integer values. Battery levels, on the other hand, are calculated by multiplying each bike's probability by its corresponding battery level and then dividing by the number of available bikes. Figure 8 displays the number of available bikes and the average battery level after two hours of simulation for two different starting times, 08:00 and 18:00. Stations near the city centre exhibit lower battery levels, consistent with the data analysis results. Meanwhile, the northern part of the city, which experiences lower usage rates, shows higher battery levels. Additionally, lower battery

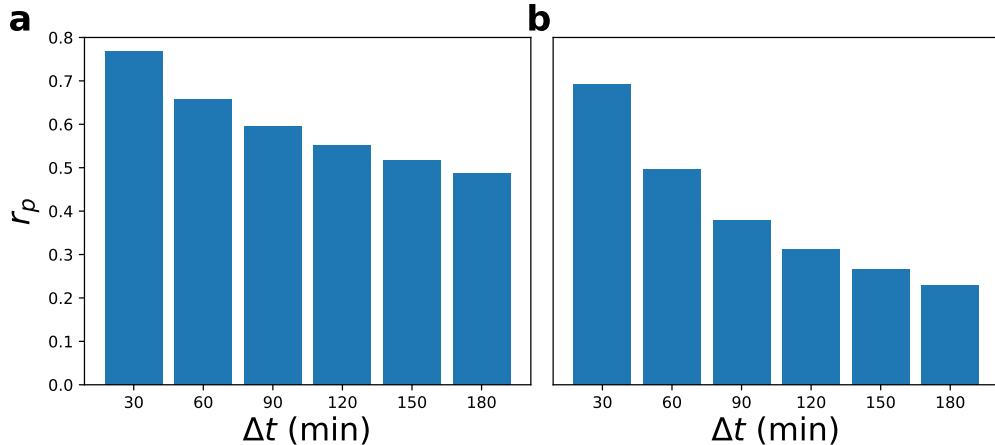


Figure 6: **Correlations for the Markov chain model over various time horizons.** **a.** Pearson correlation coefficient between actual and predicted bike availability per station as a function of  $\Delta t$ . **b.** Pearson correlation coefficient between actual and predicted average battery level per station as a function of  $\Delta t$ .

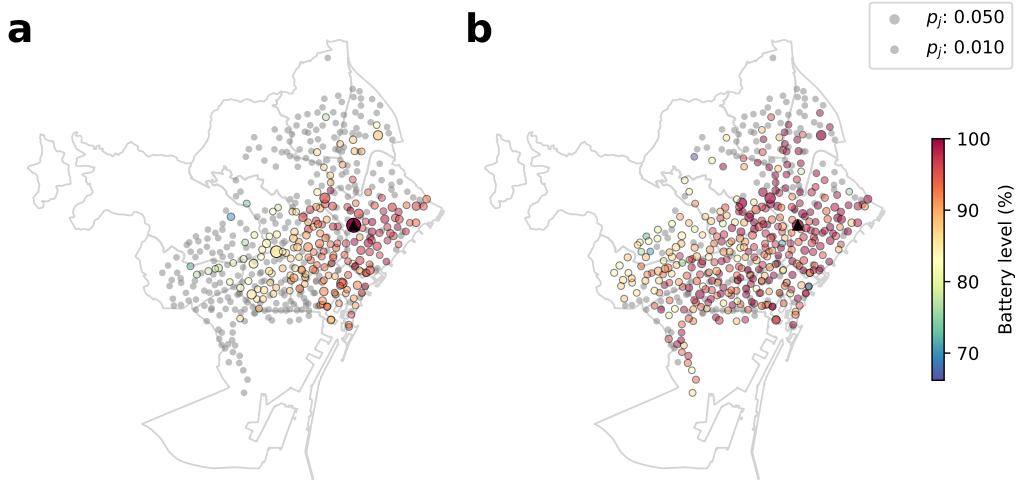
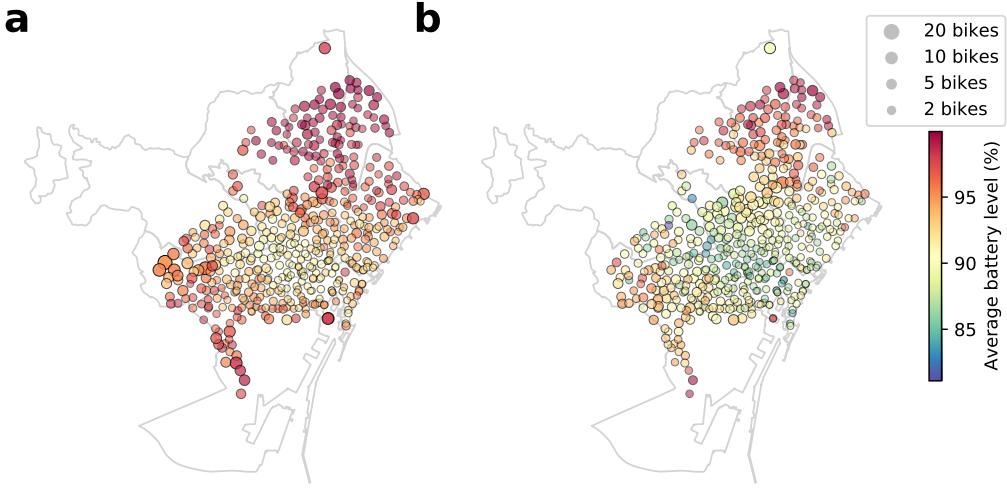


Figure 7: **Bike location and average battery level predicted with the Markov model.** Probability of finding a bike and average battery levels after **a.** 30 and **b.** 120 minutes for trips starting from the Glòries station (St. Ciutat de Granada, 168) starting at 08:00. The origin station is highlighted with a triangle.

levels are observed in the afternoon due to increased battery consumption throughout the day.



**Figure 8: Bike availability and battery levels predictions at the BSS level after 120 minutes, using the Markov chain approach combined with the battery equations.** Prediction of available bikes and their average battery levels after 120 minutes of simulation, starting at **a.** 08:00 and **b.** 18:00.

## Limitations

The main constraints of this study are the limited amount of data and the lack of detailed trip information. The first dataset has an 8-hour resolution, which complicates accurate model training and evaluation. Although the second dataset offers a 30-minute resolution, it is limited to two weeks and does not include bikes in transit. In addition to the battery data, the incomplete trip data further complicates model development. Information on the exact routes taken by users is unavailable, making it difficult to estimate battery consumption for round trips. Additionally, details such as the bike's condition (e.g., tyre inflation) or relevant rider characteristics (e.g., weight or height) are also missing.

## Discussion

The analysis revealed spatiotemporal variations in e-bike battery levels, with lower levels observed in the evenings due to daily usage, compared to higher levels in the mornings. Northern neighbourhoods tended to show higher battery levels, likely due to lower usage and longer resting times. The inter-event times and distance to other stations significantly correlate with the station's battery levels.

The presented methodology combines a traditional probabilistic framework with heterogeneous data sources, resulting in a multifaceted tool capable of improving the real-time management of

micro-mobility services. Although some challenges remain due to the lack of on-route data, such as speed or location, the battery function has been validated. This function, when combined with a Markov chain approach, provides a validated prediction for the availability of e-bikes, either due to an imbalance between stations, or to an overused battery. Further improvements could be made, particularly if more detailed data were available. Access to GPS information on the routes taken would likely improve power consumption estimates, while higher-resolution data could enhance battery predictions through machine learning approaches.

With the progressive electrification of bike fleets worldwide, predictions of availability alone may be insufficient, as docked bikes with low batteries remain unusable. The presented approach opens the door to developing microscopic models that provide joint predictions of bike availability and battery levels. This framework could support BSS providers in optimising bike availability and identifying stations at higher risk of disrupting urban mobility.

## Methods

### Mathematical formulation of the battery consumption

The battery levels of the Bicing e-bikes are modelled with a method based on the fundamental laws of motion. The instantaneous power consumed while riding is derived from the need to overcome three main forces opposing the motion: gravity resistance ( $F_g$ ), rolling resistance ( $F_r$ ), and aerodynamic drag ( $F_a$ ) [8, 45].

The gravity resistance  $F_g$  is given by:

$$F_g = g \cdot m \cdot \sin(\arctan(s)), \quad (4)$$

where  $g$  is the gravitational acceleration,  $m$  is the combined mass of the rider and bike, and  $s$  represents the slope. Similarly to [8],  $F_g = 0$  when  $F_g \leq 0$ .

Rolling resistance  $F_r$  is given by:

$$F_r = C_{rr} \cdot g \cdot m \cdot \cos(\arctan(s)), \quad (5)$$

where  $C_{rr}$  is the rolling resistance coefficient.

Finally, the aerodynamic drag  $F_a$  is given by:

$$F_a = 0.5 \cdot C_d \cdot A \cdot \rho \cdot (v + w)^2, \quad (6)$$

where  $C_d$  denotes the drag coefficient,  $A$  is the frontal area,  $\rho$  is the air density, and  $v$  and  $w$  are the bike and wind speeds, respectively.

The instantaneous power consumed is then given by:

$$P(t) = vF, \quad (7)$$

Assuming no energy losses during the process, the total power consumed over a trajectory of duration  $t$  is given by  $tv(F_g + F_r + F_a)$ . As the instantaneous speed information is unavailable, the speed is approximated by dividing the total distance travelled by the trip time. Consequently, the energy consumed in a trip is given by  $d_{ij}(F_g + F_r + F_a)$ , where the distance travelled is approximated using the OpenRouteService API [21], which recommends bicycle routes based on OpenStreetMap data [36]. To convert power consumed into a battery percentage, and taking into account that

the e-bikes only provide support to the rider when pedalling, a constant  $K_{\text{trip}}$  has been added by multiplying the whole equation  $K_{\text{trip}}d_{ij}(F_g + F_r + F_a)$ .

On the other hand, the rate of battery recharge during rest periods is modelled as:

$$\Delta B_{\text{rest}} = \frac{K_{\text{rest}}t_{\text{rest}}}{60}, \quad (8)$$

where  $K_{\text{rest}}$  represents the charging constant, and  $t_{\text{rest}}$  is the rest time. The inverse of  $K_{\text{rest}}$  corresponds to the time required to fully charge a bike from 0% to 100%. The division by 60 is included to express  $K_{\text{rest}}$  in units of  $\text{min}^{-1}$ . All variables are represented in the International System of Units.

## Optimisation of the battery consumption function

The battery consumption during a trip can be approximated by an expression that groups the relevant constants. This expression is written as:

$$\Delta B_{\text{trip}} = K_{\text{trip}}d_{ij}(K_a \sin(\arctan(s)) + K_b \cos(\arctan(s)) + K_c v^2), \quad (9)$$

where

$$\begin{aligned} K_a &= g \cdot m, \\ K_b &= C_{rr} \cdot g \cdot m, \\ K_c &= 0.5 \cdot C_d \cdot A \cdot \rho. \end{aligned}$$

Although the constant  $K_{\text{trip}}$  is not strictly required mathematically, it allows us to explore feasible values for  $K_a$ ,  $K_b$ ,  $K_c$ . Apart from  $g$ , the constants involved in the computation of  $K_a$ ,  $K_b$ , and  $K_c$  are unknown and may vary depending on factors such as rider characteristics, weather conditions, bike status, and terrain type. Therefore, an optimisation process is necessary to determine the optimal values of  $K_a$ ,  $K_b$ , and  $K_c$  that enable accurate battery consumption predictions, even when the constants of every trip are not known precisely.

To perform the optimisation, a different range of values is considered for each constant. The mass  $m$  is assumed to range between 79 kg and 119 kg, considering the bike's weight of 29 kg. Based on previous studies [45, 6], the rolling resistance coefficient can range between 0.001 and 0.003, the drag coefficient between 0.6 and 0.8, and the effective area between 0.4 and 0.6. The air density  $\rho$  is fixed at  $1.225 \text{ kg/m}^3$ , corresponding to a temperature of  $15^\circ\text{C}$  at sea level. Given these constants' value ranges, the following limits are defined for the hyper-parameter optimisation:  $K_a = [774.2, 1166.2]$ ,  $K_b = [0.7742, 3.4986]$ , and  $K_c = [0.147, 0.294]$ . The range for  $K_{\text{rest}}$  is set between  $[0.002083, 0.055]$ , corresponding to charging times from 480 to 180 minutes, respectively. The range for  $K_{\text{trip}}$  is set between  $[1 \cdot 10^{-8}, 1]$ .

The Optuna framework was utilised for hyper-parameter optimisation, using 75% of the three-week 30-minute dataset for training and 25% for validation. The optimal values obtained were:

$$\begin{aligned}K_a &= 911.139 \\K_b &= 3.497 \\K_c &= 0.292 \\K_d &= 1.644 \cdot 10^{-6} \\K_{\text{rest}} &= 480\end{aligned}$$

## Funding

This work has been funded by CIDAI (Center for Innovation in Data Tech and Artificial Intelligence) through the Catalan Government's CATALONIA.AI programme. J.G.E is a fellow of Eurecat's "Vicente López" PhD grant program.

## Acknowledgements

We thank Daniel Santanach, coordinator of CATALONIA.AI programme. We thank Marco Orelana from CIDAI (Centre of Innovation for Data Tech and Artificial Intelligence). We want to thank Faustino Corchero from Barcelona de Serveis Municipals and Roger Junqueras, Irene Giménez and all collaborators from Pedalem-Bicing.

## Author contributions statement

A.B., J.G.E and J.V. conceptualised the study. A.B. developed the methodology. A.B. and J.G.E. performed data analyses and models. A.B. and J.G.E. wrote the original draft; and all authors critically discussed the results, revised the paper and approved the final manuscript.

## References

- [1] Ajuntament de Barcelona. *Population density of Barcelona neighborhoods*. [https://ajuntament.barcelona.cat/estadistica/angles/Estadistiques\\_per\\_temes/Medi\\_urba/Territori/Superficie/a2021/S0403.htm](https://ajuntament.barcelona.cat/estadistica/angles/Estadistiques_per_temes/Medi_urba/Territori/Superficie/a2021/S0403.htm). Accessed: 2024-09-09. 2021.
- [2] Takuya Akiba et al. "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 2623–2631.
- [3] Vitória Albuquerque, Miguel Sales Dias, and Fernando Bacao. "Machine learning approaches to bike-sharing systems: A systematic literature review". In: *ISPRS International Journal of Geo-Information* 10.2 (2021), p. 62.
- [4] Mohammed H Almannaa, Mohammed Elhenawy, and Hesham A Rakha. "Dynamic linear models to predict bike availability in a bike sharing system". In: *International journal of sustainable transportation* 14.3 (2020), pp. 232–242.

- [5] Huthaifa I Ashqar, Mohammed Elhenawy, and Hesham A Rakha. “Modeling bike counts in a bike-sharing system considering the effect of weather conditions”. In: *Case studies on transport policy* 7.2 (2019), pp. 261–268.
- [6] William M Bertucci, Simon Rogier, and Raoul F Reiser. “Evaluation of aerodynamic and rolling resistances in mountain-bike field conditions”. In: *Journal of sports sciences* 31.14 (2013), pp. 1606–1613.
- [7] Tomasz Bieliński et al. “Lessons from implementing a metropolitan electric bike sharing system”. In: *Energies* 13.23 (2020), p. 6240.
- [8] Erik Burani, Giacomo Cabri, and Mauro Leoncini. “An algorithm to predict e-bike power consumption based on planned routes”. In: *Electronics* 11.7 (2022), p. 1105.
- [9] Andrew A Campbell et al. “Factors influencing the choice of shared bicycles and shared electric bikes in Beijing”. In: *Transportation research part C: emerging technologies* 67 (2016), pp. 399–414.
- [10] Longbiao Chen et al. “Dynamic cluster-based over-demand prediction in bike sharing systems”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2016, pp. 841–852.
- [11] Longbiao Chen et al. “Understanding bike trip patterns leveraging bike sharing system open data”. In: *Frontiers of computer science* 11 (2017), pp. 38–48.
- [12] Federico Chiariotti et al. “A dynamic approach to rebalancing bike-sharing systems”. In: *Sensors* 18.2 (2018), p. 512.
- [13] Raeven Lynn M Clockston and David Rojas-Rueda. “Health impacts of bike-sharing systems in the US”. In: *Environmental research* 202 (2021), p. 111709.
- [14] Mauro Dell’Amico et al. “The bike sharing rebalancing problem: Mathematical formulations and benchmark instances”. In: *Omega* 45 (2014), pp. 7–19.
- [15] Mauro Dell’Amico et al. “The bike sharing rebalancing problem with stochastic demands”. In: *Transportation research part B: methodological* 118 (2018), pp. 362–380.
- [16] Wafic El-Assi, Mohamed Salah Mahmoud, and Khandker Nurul Habib. “Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto”. In: *Transportation* 44 (2017), pp. 589–613. URL: <https://api.semanticscholar.org/CorpusID:133307286>.
- [17] Ahmadreza Faghih-Imani et al. “An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville”. In: *Transportation Research Part A: Policy and Practice* 97 (2017), pp. 177–191.
- [18] Yichun Fan and Siqi Zheng. “Dockless bike sharing alleviates road congestion by complementing subway travel: Evidence from Beijing”. In: *Cities* 107 (2020), p. 102895.
- [19] David Florez et al. “Development of a bike-sharing system based on pedal-assisted electric bicycles for bogota city”. In: *Electronics* 7.11 (2018), p. 337.
- [20] Nikolaos-Fivos Galatoulas, Konstantinos N Genikomsakis, and Christos S Ioakimidis. “Spatio-temporal trends of e-bike sharing system deployment: A review in Europe, North America and Asia”. In: *Sustainability* 12.11 (2020), p. 4611.
- [21] GIScience Research Group and HeiGIT. *openrouteservice*. <https://openrouteservice.org/>. Accessed: 2024-08-20. 2023.

- [22] Jordi Grau-Escalano, Aleix Bassolas, and Julian Vicens. “Cycling into the workshop: e-bike and m-bike mobility patterns for predictive maintenance in Barcelona’s bike-sharing system”. In: *EPJ Data Science* 13.1 (2024), pp. 1–21.
- [23] Wei Huang et al. “Predicting human mobility with activity changes”. In: *International Journal of Geographical Information Science* 29.9 (2015), pp. 1569–1587.
- [24] Shuguang Ji et al. “Electric bike sharing: simulation of user demand and system availability”. In: *Journal of Cleaner Production* 85 (2014), pp. 250–257.
- [25] Christian Kapuku et al. “Assessing and predicting mobility improvement of integrating bike-sharing into multimodal public transport systems”. In: *Transportation research record* 2675.11 (2021), pp. 204–213.
- [26] Injung Kim and Konstantinos Pelechrinis. “The anatomy of the daily usage of bike sharing systems: elevation, distance and seasonality”. In: *ACM SIGKDD workshop on Urban Computing*. 2020.
- [27] Kyoungok Kim. “Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations”. In: *Journal of Transport Geography* 66 (2018), pp. 309–320. URL: <https://api.semanticscholar.org/CorpusID:158630089>.
- [28] Zhaoyu Kou and Hua Cai. “Understanding bike sharing travel patterns: An analysis of trip data from eight cities”. In: *Physica A: Statistical Mechanics and its Applications* 515 (2019), pp. 785–797.
- [29] Wenwen Li et al. “Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories”. In: *International Journal of Geographical Information Science* 34.12 (2020), pp. 2451–2474.
- [30] Yexin Li et al. “Traffic prediction in a bike-sharing system”. In: *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*. 2015, pp. 1–10.
- [31] Xin Lu et al. “Approaching the limit of predictability in human mobility”. In: *Scientific reports* 3.1 (2013), p. 2923.
- [32] Xinwei Ma et al. “Bike-sharing systems’ impact on modal shift: A case study in Delft, the Netherlands”. In: *Journal of cleaner production* 259 (2020), p. 120846.
- [33] Meddin. *The Meddin Bike-sharing World Map*. <https://bikesharingworldmap.com>. Accessed: 2024-06-30.
- [34] A. Nisbet. *OpenTopoData*. <https://github.com/ajnisbet/opentopodata>. Accessed: 2024-08-20. 2013.
- [35] Robert B Noland, Michael J Smart, and Ziye Guo. “Bikesharing trip patterns in New York City: Associations with land use, subways, and bicycle lanes”. In: *International journal of sustainable transportation* 13.9 (2019), pp. 664–674.
- [36] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. <https://www.openstreetmap.org/>. Accessed: 2024-08-20. 2013.
- [37] Isabel Otero, Mark J Nieuwenhuijsen, and David Rojas-Rueda. “Health impacts of bike sharing systems in Europe”. In: *Environment international* 115 (2018), pp. 387–394.

- [38] Tal Raviv, Michal Tzur, and Iris A Forma. "Static repositioning in a bike-sharing system: models and solution approaches". In: *EURO Journal on Transportation and Logistics* 2.3 (2013), pp. 187–229.
- [39] Miriam Ricci. "Bike sharing: A review of evidence on impacts and processes of implementation and operation". In: *Research in Transportation Business & Management* 15 (2015), pp. 28–38.
- [40] Maya Siman-Tov et al. "A look at electric bike casualties: Do they differ from the mechanical bicycle?" In: *Journal of Transport & Health* 11 (2018), pp. 176–182.
- [41] Özlem Simsekoglu and Christian A Klöckner. "The role of psychological and socio-demographical factors for electric bike use in Norway". In: *International journal of sustainable transportation* 13.5 (2019), pp. 315–323.
- [42] Divya Singhvi et al. "Predicting bike usage for new york city's bike sharing system". In: *Workshops at the twenty-ninth AAAI conference on artificial intelligence*. 2015.
- [43] Adish Singla et al. "Incentivizing users for balancing bike sharing systems". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 29.1 (2015).
- [44] Francesc Soriguera, Víctor Casado, and Enrique Jiménez. "A simulation model for public bike-sharing systems". In: *Transportation research procedia* 33 (2018), pp. 139–146.
- [45] Wynand J vdM Steyn and Janike Warnich. "Comparison of tyre rolling resistance for different mountain bike tyre diameters and surface conditions". In: *South African Journal for Research in Sport, Physical Education and Recreation* 36.2 (2014), pp. 179–193.
- [46] Tien Dung Tran, Nicolas Ovtracht, and Bruno Faivre d'Arcier. "Modeling bike sharing system using built environment factors". In: *Procedia Cirp* 30 (2015), pp. 293–298.
- [47] Muhammad Usama, Yongjun Shen, and Onaira Zahoor. "Towards an energy efficient solution for bike-sharing rebalancing problems: a battery electric vehicle scenario". In: *Energies* 12.13 (2019), p. 2503.
- [48] John Wamburu et al. "Ride substitution using electric bike sharing: feasibility, cost, and carbon analysis". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5.1 (2021), pp. 1–28.
- [49] Jonathan Weinert, Chaktan Ma, and Christopher Cherry. "The transition to electric bikes in China: history and key reasons for rapid growth". In: *Transportation* 34 (2007), pp. 301–318.
- [50] Julia Winslow and Oksana Mont. "Bicycle sharing: Sustainable value creation and institutionalisation strategies in Barcelona". In: *Sustainability* 11.3 (2019), p. 728.
- [51] Yuanxuan Yang et al. "Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems". In: *Computers, Environment and Urban Systems* 83 (2020), p. 101521.
- [52] Zidong Yang et al. "Mobility modeling and prediction in bike-sharing systems". In: *Proceedings of the 14th annual international conference on mobile systems, applications, and services*. 2016, pp. 165–178.
- [53] Martin Zaltz Austwick et al. "The structure of spatial networks and communities in bicycle sharing systems". In: *PloS one* 8.9 (2013), e74685.
- [54] Chaofeng Zhang et al. "Battery maintenance of pedelec sharing system: Big data based usage prediction and replenishment scheduling". In: *IEEE Transactions on Network Science and Engineering* 7.1 (2019), pp. 127–138.

- [55] Yaoming Zhou et al. “Dynamic battery swapping and rebalancing strategies for e-bike sharing systems”. In: *Transportation Research Part B: Methodological* 177 (2023), p. 102820.

# Supplementary Material

## Spatiotemporal variability and prediction of e-bike battery levels in bike-sharing systems

### 1 Data description

The first dataset we analysed comprises battery levels recorded every 8 hours, containing data on 2,748 bikes from July 20th to December 31st, 2022 (spanning 164 days). Battery levels are recorded exclusively when bikes are docked at stations, resulting in 1.48 million data points. The dataset fields are datetime in second granularity, bike identifier, station, status (operative or inoperative for various maintenance reasons), and battery metrics such as voltage, temperature, current, and battery level (from 0 to 1). After excluding records with missing values and filtering for those labelled as ‘operative’, the dataset was narrowed to 671,251 rows. Additionally, for the modelling phase, only data points with a subsequent data point for the same bike 8 hours later were utilised. The details are provided in Figure S1.

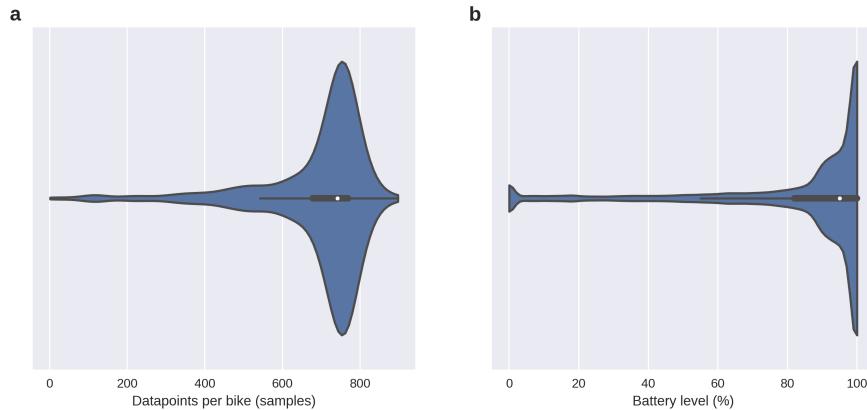


Figure S1: **30-minute dataset description.** **a.** Distribution of data points per bike. **b.** Distribution of battery levels across the dataset.

The second dataset contains battery level snapshots taken every 30 minutes for 3,004 bikes, totalling 2.71 million data points from March 29th to April 18th, 2023 (spanning 20 days). Similarly, this dataset includes the same fields, with battery level information available only when bikes are docked. After data cleaning and selecting the bikes with an ‘operative’ status, 1.87 million data points remained. For analysis, only data points with a subsequent reading for the same bike 8 hours later were used. The details are provided in Figure S2.

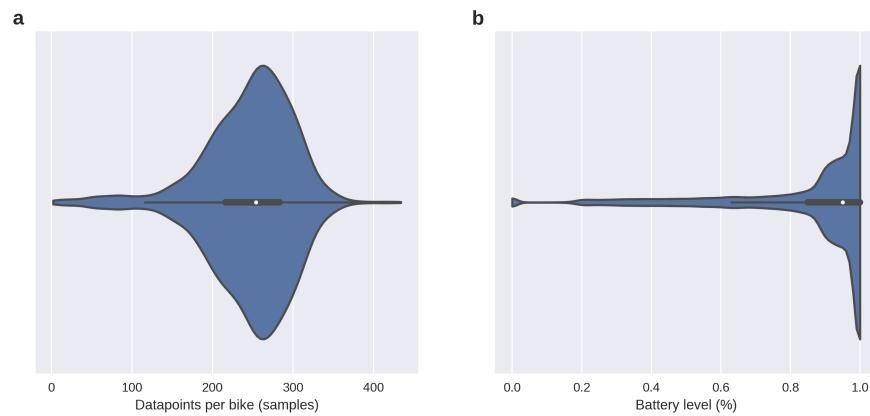


Figure S2: **8-hour dataset data description.** **a.** Distribution of data points per bike. **b.** Distribution of battery levels across the dataset.

## 2 Data analysis

### Battery levels across time for each neighbourhood

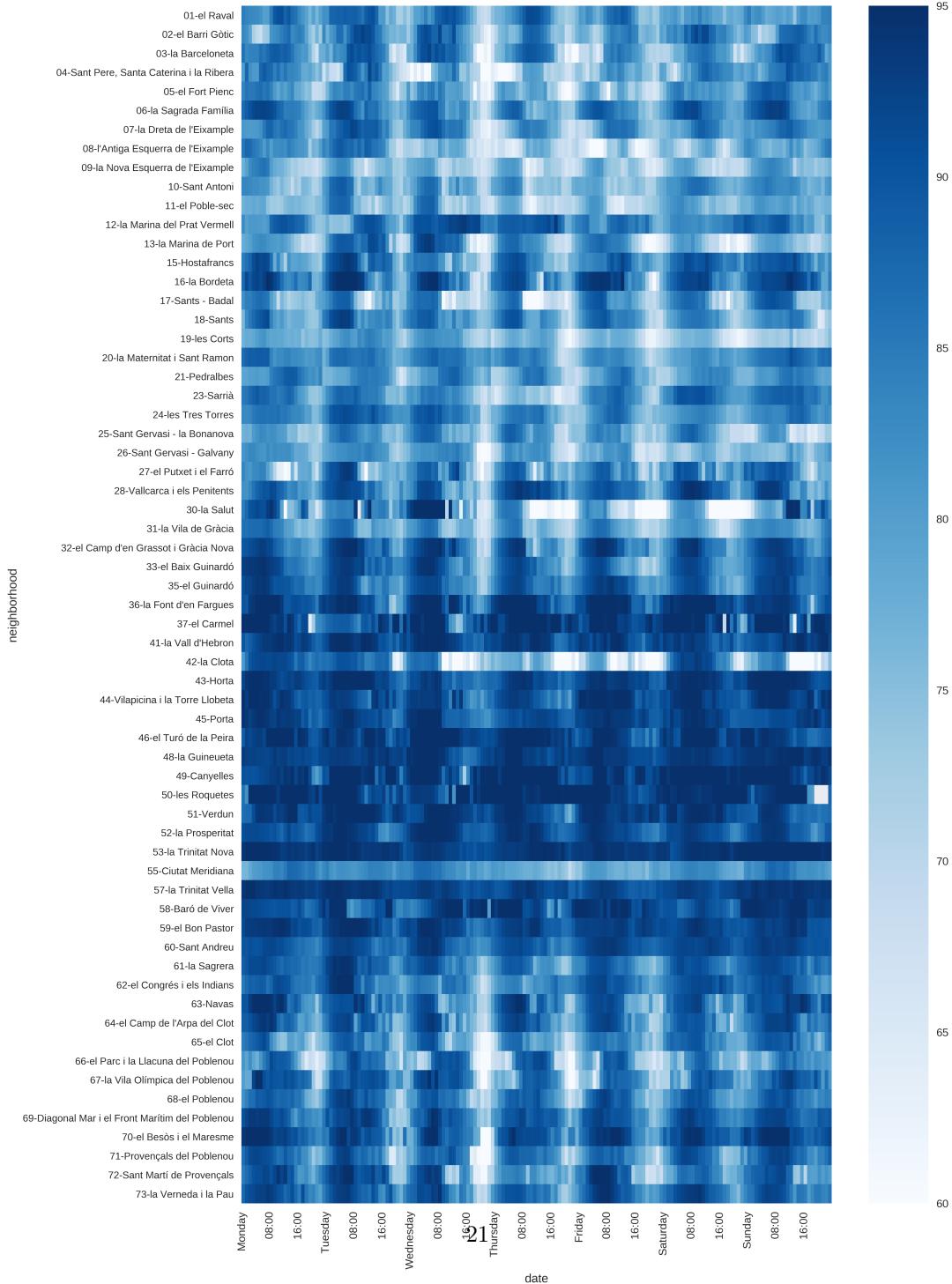


Figure S3: **Average battery levels across time for each neighbourhood.** The x-axis represents time for an average week, while the y-axis lists the neighbourhoods.

## Choice of number of clusters

To determine the optimal number of clusters for the k-means clustering analysis, we utilized two commonly used methods: the Elbow Method and the Silhouette Method (Figure S4). The Elbow Method helps identify the point where the within-cluster sum of squares (inertia) begins to decrease at a slower rate, indicating the optimal number of clusters. The Silhouette Method measures the quality of the clustering by calculating the average silhouette score for different numbers of clusters.

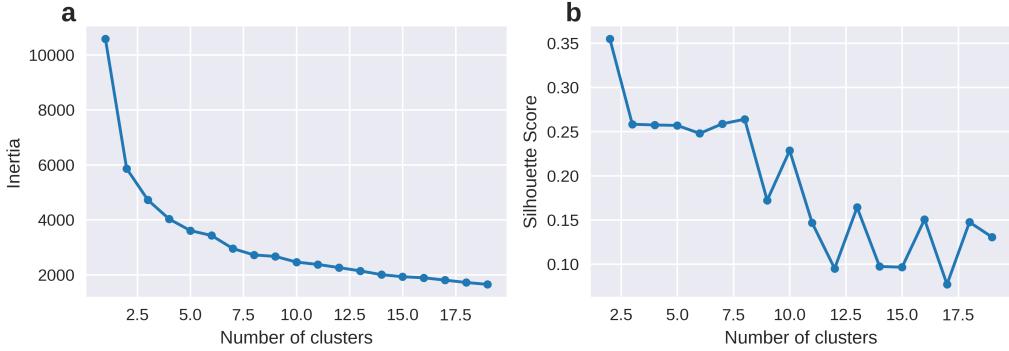


Figure S4: **Determining the optimal number of clusters.** **a.** Elbow Method for optimal k. **b.** Silhouette Method for optimal k.

Based on these analyses, we chose to use 4 clusters for our k-means clustering. This choice balances the trade-off between minimising within-cluster variance (as indicated by the Elbow Method) and maximising cluster separation (as indicated by the Silhouette Method).

## Collinearity between inter-event time and distance index

To ensure that there is no significant collinearity between the inter-event time and the distance index, we examined both the correlation matrix and the Variance Inflation Factor (VIF) for each variable in the regression model.

Figure S5 illustrates the correlation values among average battery level, inter-event time, and distance index. The correlation between inter-event time and distance index is 0.77, indicating a moderate relationship. At first glance, this might suggest potential collinearity between these two variables.

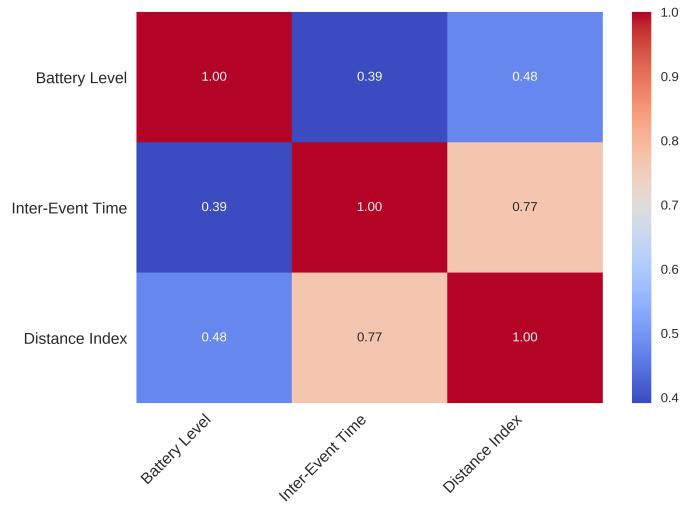


Figure S5: Correlation matrix for battery level, inter-event time, and distance index.

However, to confirm whether this correlation leads to collinearity, we computed the VIF for each variable. The VIF quantifies how much the variance of a regression coefficient is inflated due to collinearity with other predictors. A VIF value greater than 10 typically indicates high collinearity.

The table below summarises the VIF values for the variables:

Variable	VIF
constant	170.328236
battery level	1.296030
inter-event time	2.447021
distance index	2.680041

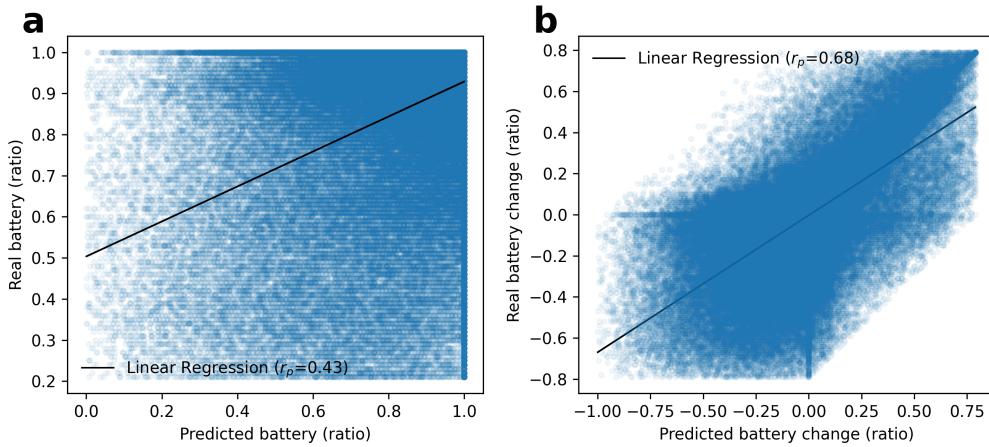
Table S1: Variance Inflation Factor (VIF) values for the variables in the regression model.

As shown in Table S1, the VIF values for the inter-event time and the distance index are 2.447021 and 2.680041, respectively. These values are well below the threshold of 10, indicating that there is no significant collinearity between these variables.

Thus, considering both the correlation matrix and the VIF values, we conclude that the inter-event time and the distance index can be used as independent predictors in the regression model without concerns of collinearity.

### 3 Battery validation and elevation data

In Figure S6, the correlation between the predicted and observed battery levels in the 8-hour dataset is shown.



**Figure S6: Battery model fit in the 8-hour dataset.** **a.** Observed battery level as a function of the predicted values. **b.** Battery change observed as a function of the predicted values. The Root Mean Squared Error (RMSE) is 0.036. The linear fit between both is shown in grey.

Battery consumption is highly impacted by the altitude difference between the origin and destination stations of a trip. The altitude of each station in the BSS of Barcelona is provided in Figure S7. A clear division between the station near the sea and those in the northern region is observed. In some cases, the disparity between stations is steep, with differences in altitude greater than 100 metres.

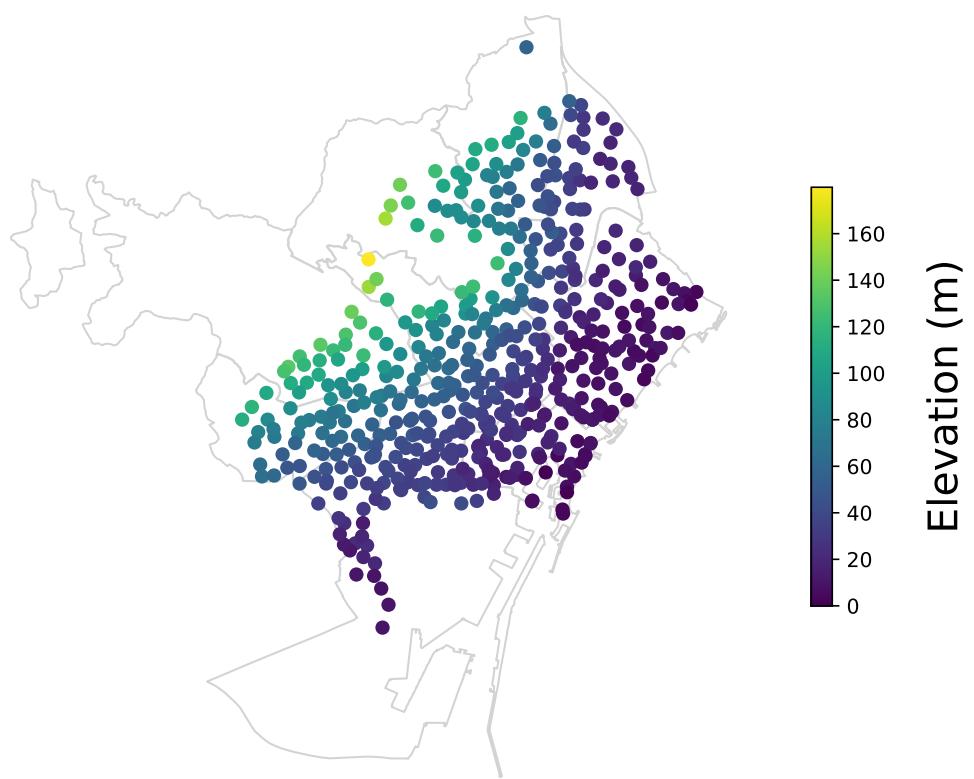


Figure S7: **Station altitude.** Altitude of the Bicing stations.