Impact of Data Resampling on the Accuracy of Bicycle Travel Time Estimations

Maroje Luka Borsic
Supervisor(s): Elvin Isufi, Ting Gao
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology, In Partial Fulfilment of the Requirements For the Bachelor of Computer Science and Engineering June 8, 2025

Name of the student: Maroje Luka Borsic
Final project course: CSE3000 Research Project
Thesis committee: Elvin Isufi, Ting Gao, Jing Sun

Abstract

Estimating bike trip times is becoming more and more important in many different areas such as urban mobility and route planning. However, especially in real-world, the GPS data used to generate these estimations is frequently noisy, irregularly sampled, or incomplete. With an emphasis on how these strategies interact with trip length and speed variance, this study intends to examine the effects of various data resampling techniques on the precision of bicycle travel time estimations. To analyze the impact of different preprocessing methods, we apply and assess a graph neural network model using various resampling techniques. Instinctively, the assumption that we expect to be concluded from this research is that there is no single resampling technique works well for every kind of trip. Rather, trip parameters like duration and speed fluctuation have a significant impact on accuracy.

1 Introduction

There are twice as many bicycles in the world than cars and it is currently estimated that about 100 million bicycles are in use all around the world [Hosmer, 2025]. As amazing and interesting as this sounds, this also emphasizes the importance of smart city planning and sustainable mobility becoming more and more vital because more bicycles means cities manage more traffic and they have to ensure safety for cyclists in a safer manner. This also highlights how crucial the ability to accurately estimate the time it takes to ride a bicycle from one point to another really is. In contrast to motor vehicles, bicycles are usually utilized in a wider variety of unpredictable ways. For example, cyclists may take shortcuts, go on tight paths or paths not intended for bicycle use, stop frequently and even violate traffic regulations more frequently in comparison to motor vehicles. Apart from GPS devices such as trackers installed in the bicycles, this is the reason that the GPS data that records these trips is frequently noisy, erratic, or lacking in some areas. However, machine learning models for predicting trip times are frequently trained using this same noisy and irregular GPS data.

Resampling, or changing the spacing of time or distance points to smooth out irregularities, is one of the first stages in producing such data. Although resampling is frequently used in general time series modeling, it is yet unknown how exactly it affects bicycle trip time estimation. Neural network-based methods for predicting bicycle trips using raw trajectory data have been studied in the past [Reggiani et al., 2020], frequently with encouraging outcomes. But few research have examined the impact of resampling options on model accuracy, and those that have tended to concentrate on data from cars or public transportation instead of cycling. Nonetheless, a study that focuses more intently on resampling techniques in the medical area exists [Khushi et al., 2021]. Even though this study and the previously mentioned one are done in vastly different fields, the methods and techniques used could have a big overlap.

This paper seeks to fill that gap by asking the following research question:

What is the impact of data resampling on the accuracy of travel time estimations, particularly for trip length and speed variation?

We explore this through a set of sub-questions:

How do different resampling methods (e.g., fixed interval, rolling average, interpolation) affect prediction accuracy? Are shorter or longer trips more sensitive to the resampling method used? How does variation in cycling speed affect the performance of models trained on resampled data? Is there a single resampling method that performs reliably across most trip types?

In order to address these issues, we employ a graph neural network model architecture that takes into account both route structure and trip dynamics, using real-world GPS records gathered from bicycle excursions. This real-world dataset includes data collected via the SimRa project conducted in Germany, mainly concentrated around the city of Berlin. The plan is to investigate several resampling techniques, group the findings by trip duration, and speed variation before evaluating the results using RMSE, MAE, and MAPE.

The main contributions intend to be:

A systematic analysis of how data resampling affects model accuracy in bicycle trip time estimation. An evaluation of different resampling methods using a graph neural model that handles spatial road data. Insights on which strategies work best for different trip types, and practical guidelines for improving future mobility prediction models.

2 Methodology

In order to answer the main research question of this project, the CSE3000-eta-bicycle codebase will be used and worked upon. The current codebase uses a graph neural network (GNN) to estimate bicycle travel time based on GPS data. The goal is to enhance this codebase to use of different data resampling methods on the GPS data and measure whether or not these methods will have a noticeable impact on the final travel time estimations. The different resampling methods include distance-based sampling, rolling averages, downsampling and Kalman filter smoothing. These methods will be applied to the data attained by the SimRa project in Germany, which includes records of unpredictable, noisy bicycle rides. The final goal of this project is the provide a clear and unique insight into how the early stages of data selection can affect the model outputs and to help future works in mobility research make better choices when making design decisions.

The trips from the dataset will be group based on overall trip length. The categories include short trips (less than 8 minutes), medium trips (8-16 minutes) and long trips (16+ minutes). The model will then be trained and evaluated for each trip group type using the standard error metric, which includes the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). Each of these metrics will provide an insightful idea into how each resampling method is interacting with each trip group type. The results will help show if performance changes based on these trip types or if there is one method that is best fit for every trip. Parameters such as the seed or batch size are to be kept the same for each method in order to isolate the effects of resampling to carry out a controlled study that focuses solely on the preprocessing step.

3 Literature Review

Estimating bicycles travel time is regarded as a challenging issue due to the variability and noise of the data. Bicyclists frequently use their bikes in unexpected ways, stopping more often than vehicles, taking shortcuts, and riding through parks or alleys. All of these events add up to a great deal of unpredictability to GPS data. Bikes frequently travel more flexible routes in comparison to cars or public transportation, which hold traffic rules and regulations to a higher degree of importance. Due to so many bikers using paths that they are not predicted to use, makes using GPS to estimate travel time even more difficult than it already is.

In the past, older models such as time-based regression or historical averages have been applied, however they are not effective enough at handling the kind of volatility or noise presented by this issue (Amini et al., 2016; Pan et al., 2012). Additionally, these models will often disregard the layout of the road network or variations in key variables such as speed along the trip. Research that is more recent improves accuracy through the use of machine learning. For predicting factors like trip length or speed on particular road types, random forests, gradient boosting and support vector machines have demonstrated superior results (Lin et al., 2018; Yan et al., 2020; Cheng et al., 2019). However, these models also do not take into consideration the spatial layout or interactions between different road segments, and they primarily treat trips as simple time series.

A more recent method that uses the real road layout is graph-based models. These models can deal with data structured as road networks using Graph Neural Networks (GNNs), where each node represents a segment and each edge illustrates the connections between segments (Yu et al., 2017; Chai et al., 2018). This increases the model's awareness of route structure and helps capture more realistic interactions, including slower speeds in turns or crossings. GNNs or graph-based variations such as TAGCN or STGCN have been used to bicycle systems in a number of articles. By learning from past GPS traces and connecting them to road segments, these models, which were frequently first employed for demand estimation, have been modified to estimate time and speed (Zi et al., 2021; He & Shin, 2020). Before entering such models, noisy travel data is additionally cleaned up using Kalman filters and other smoothing methods (Huang et al., 2022).

What is missing in most of the current research is a detailed look at how preprocessing affects these predictions. Many works use raw or smoothed data without comparing the impact of different resampling or filtering methods. Since trip time predictions depend a lot on clean and structured input, this paper focuses on that early stage, resampling, and how it affects GNN-based predictions of bicycle trip time.

4 Analysis of Segment-Level Resampling Effects

This section provides a more detailed look into each resampling technique, including how it is used in the project, how does the data changes, how this affects the performance and how this impacts the graph neural network's input. The techniques used in the project focus on transforming unprocessed trip data before it is being used in training. The resampling techniques being used are: rolling average, downsampling, fixed distance sampling, and Kalman filter smoothing. Although the resampling methods are applied to raw GPS traces (a series of location points over time), they affect the processed data that the model uses — like the speed, distance, or timing linked to each road segment the cyclist traveled on.

The value and the order of features such as speed, distance traveled or the percentage of the trip that was finished are changed with each different resampling technique. Below is a more detailed description of each resampling method:

- Default (None): No changes are made to the original GPS data. This specific model uses the raw data as-is and it acts as a baseline and provides the original, unprocessed values.
- Downsampling: First, map matching is performed to determine which road segments the chosen GPS points correspond to. Then, every fifth point in the GPS trace is kept and the rest of the GPS points are discarded. Therefore, if a trip had 100 points, only 20 would remain. This significantly reduces the data volume and shortens the training time, but it comes at the cost of losing the fine-grained motion patterns.
- Rolling Average: Each speed value is smoothened by averaging the speed value with nearby values using a rolling window (e.g., 5 points wide). For example, if a speed-reading suddenly jumps from 10 to 30 then back down to 12 km/h, this method will replace those values with something more stable like 16, 22, 23, 20, and 18. This helps reduce spikes and drops due to noise, but it might come at a cost of hiding real variability.
- Kalman Smoothing: Kalman smoothing uses a mathematical model to guess what the "true" values of the data (like speed or distance) should be. It does this by treating GPS readings as noisy guesses rather than perfect truths. The method assumes that each GPS point might have some error. It also assumes that the actual path of the vehicle changes gradually and that future and past points can help improve each estimate. This method works in two steps, predict and update. The predict part makes a guess about the next value based on the current trend (e.g., continuing at the same speed) and then the update part adjusts that guess based on the actual GPS reading, using a balance between trusting the data and trusting the trend. This process is repeated for all points, resulting in a version of the trip that is smoother, less noisy, and often closer to the true motion of the vehicle. Kalman smoothing is more powerful than a simple average, especially when the GPS data is inconsistent. However, it is also more complex to run and can fail if the trip is too short or the input is too messy.
- Fixed Distance Sampling: Points are kept at a fixed spatial interval (e.g., every 50 meters). A new point is kept every fixed number of meters (e.g., every 10 or 50 meters), regardless of the time it was recorded. This works well for regular-speed trips and keeps the spacing between data points physically meaningful. This aligns well with physical movement, especially for consistent-speed trips, but can underrepresent areas with complex changes.

The pipeline was set up so that the preprocessing step was the only thing that varied between tests and it was done before the training stage. In reality, this indicates that the resampling technique was used immediately following the loading of the unprocessed GPS trajectory data, but prior to building the representation of the road segments. Following resampling, segment-level variables like average speed, travel distance, and time-to-go were recovered after the data was mapped to fixed road segments using a spatial matching procedure. The graph inputs were then constructed using these. The graph structure remained unchanged. In other words, the nodes (road segments) and the connections between them were constant throughout the studies. Depending on how the resampling technique had affected the underlying trip data, the only thing that changed were the feature values linked to each node. The anticipated result would be that different trip types—such as short versus lengthy travels or trips with greater speed fluctuation—would react differently to these changes. Longer trips may benefit from smoothing by preventing overfitting to local abnormalities, whereas short trips may be more susceptible to downsampling because there are fewer data points available initially. This hypothesis is tested on a variety of real-world travels in the evaluation that is detailed in the following section.

5 Evaluation Setup and Results

The evaluation is performed based on the real-world cycling data collected through the SimRa project in Germany. This dataset contains GPS data of bicycle trips recorded in mostly urban environments, including information such as timestamped locations, speed, and distance. The data is known to be noisy and unevenly sampled, which makes it a good fit for studying the effects of different resampling methods. Each resampling technique was performed separately and used the same exact graph neural network design for all of its runs in order to assess how resampling affected prediction accuracy without changing the randomness of the seed. To prevent data bias, the dataset was divided into training, validation, and testing sets using fixed route identifiers. The resampling technique used on the input trip data was the only thing we altered between runs.

The three trip categories used include: short which is less than eight minutes long, medium which is eight to sixteen minutes long, and long trips which are more than sixteen minutes long. The model was then trained and evaluated independently for every trip length. This evaluation includes the use of the previously mentioned performance metrics of MAPE, MAE and RMSE. In Table 1. we can see the exact values and results of the evaluation for each trip type per each resampling method.

Table 1. Table of performance metrics by resampling method for short trips

Resampling Method	RMSE	MAE	MAPE
Default	68.77	50.79	36.89
Downsampling	70.72	53.78	43.50
Rolling Average	59.65	44.46	32.23
Kalman	60.89	44.88	29.31
Fixed Distance	76.24	56.52	36.59

Table 2. Table of performance metrics by resampling method for medium trips

Resampling Method	RMSE	MAE	MAPE
Default	107.05	73.55	16.11
Downsampling	111.44	75.37	17.05
Rolling Average	99.64	67.47	15.13
Kalman	96.45	62.13	14.12
Fixed Distance	107.06	78.37	17.76

Table 3. Table of performance metrics by resampling method for long trips

Resampling Method	RMSE	MAE	MAPE
Default	694.14	215.89	14.99
Downsampling	719.34	226.00	15.36
Rolling Average	681.71	207.43	14.50
Kalman	665.56	216.21	15.06
Fixed Distance	669.21	237.92	17.11

Table 4. Table of performance metrics by resampling method for overall trips

Resampling Method	RMSE	MAE	MAPE
Default	529.02	151.88	19.16
Downsampling	541.19	155.94	20.32
Rolling Average	512.39	141.86	17.50
Kalman	500.26	145.31	17.06
Fixed Distance	503.90	163.87	20.39

Table 5. Table of time performance for each resampling method

Resampling method	Epoch	Time per Epoch (s)	Total Time (s)
Default	49	28.04	1373.96
Downsampling	25	10.53	263.25
Rolling Average	21	26.16	549.36
Kalman	10	43.16	431.60
Fixed Distance	19	9.53	181.07

Figure 1. Short trip performance metrics

Figure 2. Medium trip performance metrics

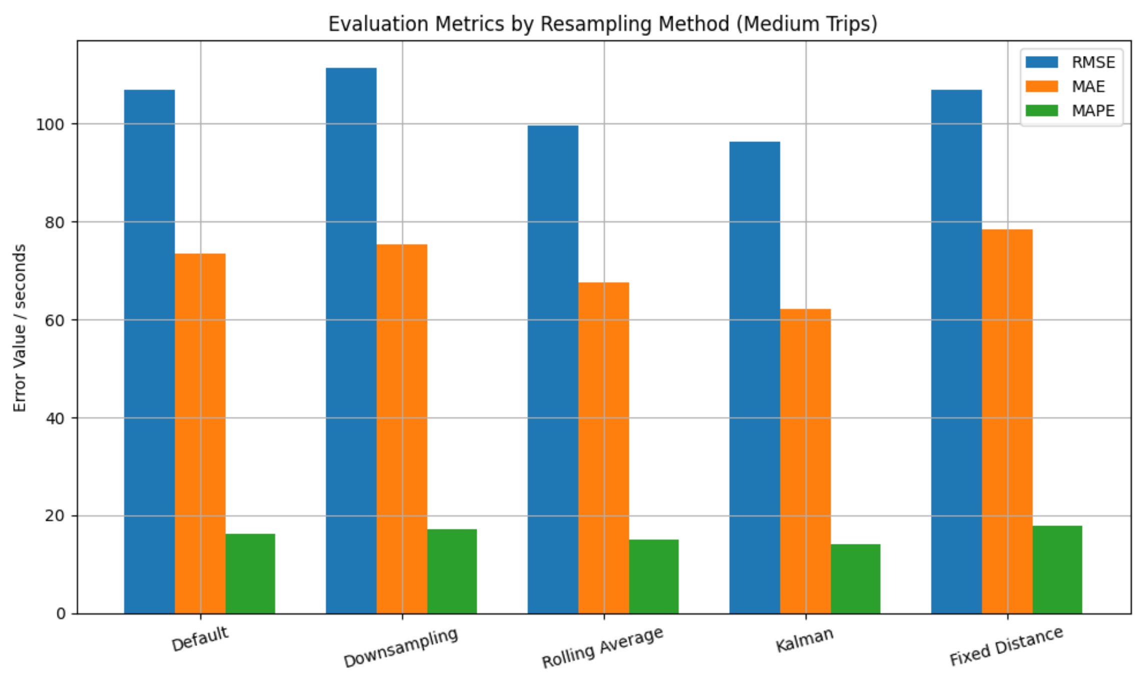


Figure 3. Long trip performance metrics

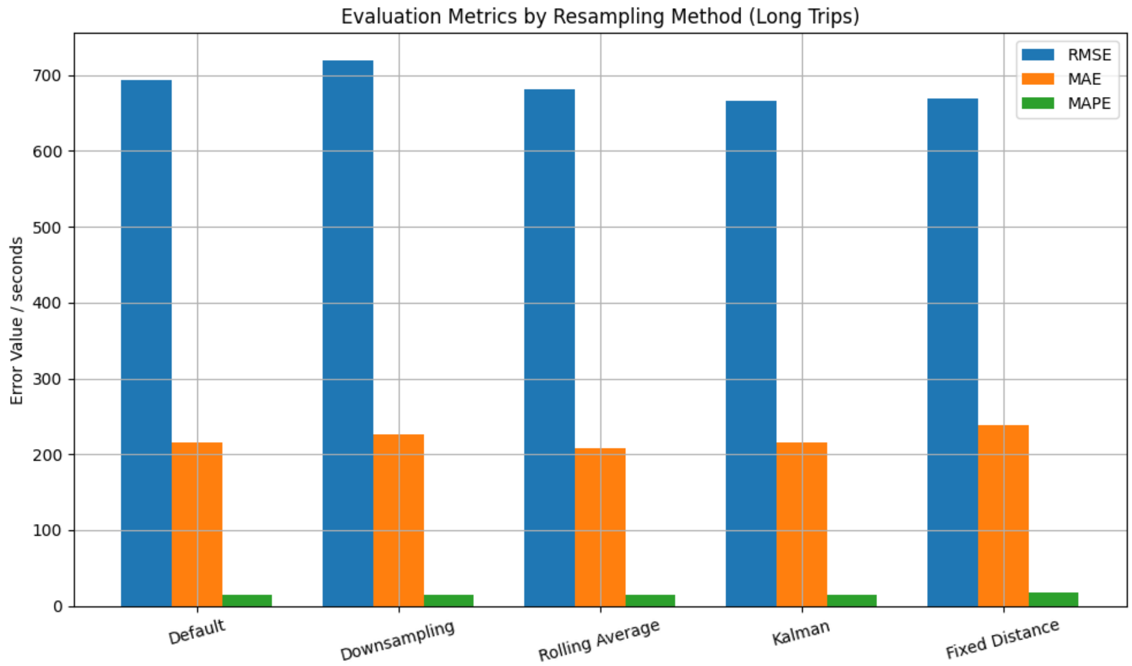


Figure 4. Overall trip performance metrics

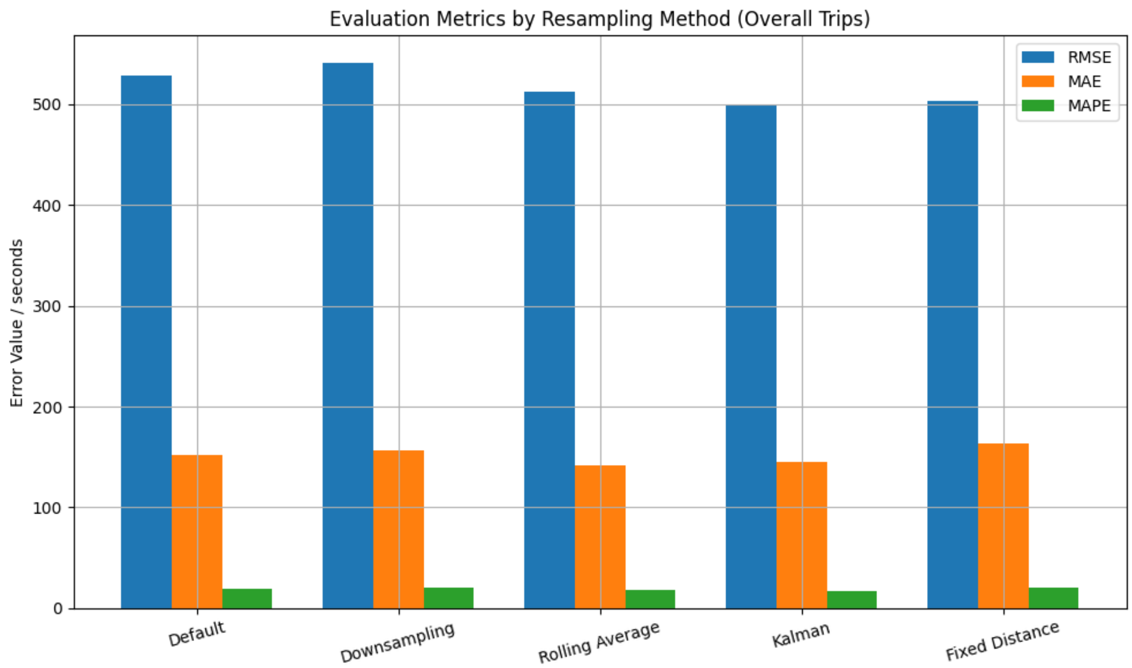
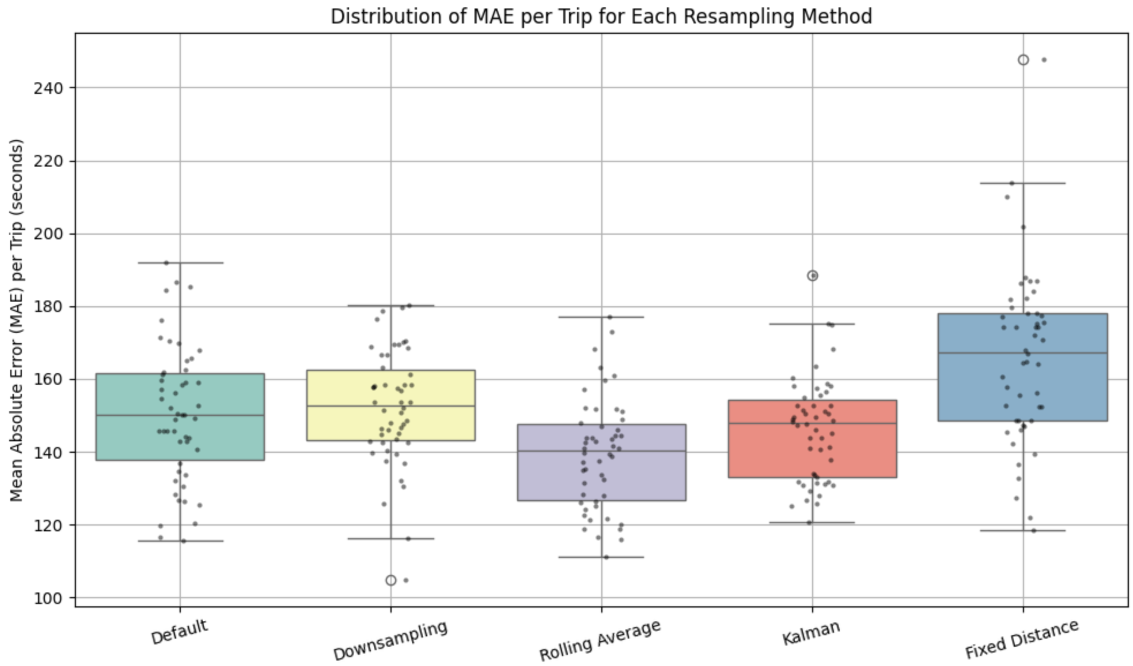


Figure 5. Variation in predication error per trip



Several significant trends can be extrapolated from the figures and tables presented. Most notably, during long trips, all approaches tend to yield more significant error value. This is rather expected since longer trips inherently involve greater fluctuations in speed, road conditions, and GPS irregularities, making precise estimation more challenging. Going in depth for each technique reveals the following:

The downsampling technique performs the worst for short and medium trips, as shown in Figure 1 and Figure 2. This is most likely due to the fact that deleting GPS locations removes important information, especially for shorter trips. Using every fifth or tenth GPS point can drastically distort the shape of already short trips, leading to higher RMSE, MAE, and especially MAPE values. In contrast, Figure 1 also shows that rolling average performs better in these shorter trips, with lower MAE and MAPE than downsampling. This is likely because it smooths out sudden changes while still preserving enough detail. Kalman smoothing, visible in all three figures, lands somewhere in the middle. It balances noise reduction with maintaining data integrity, but still introduces slight bias, especially in longer trips as seen in Figure 3. Fixed distance resampling shows major issues in short trips (Figure 1), with the highest RMSE and MAE. This might be due to poor alignment of resampled points in very short trajectories. Surprisingly, not applying any resampling method, referred to as the default method, does not perform best overall. It ranks slightly below downsampling in most cases, which may indicate that while raw GPS data can be useful, minimal resampling still helps improve model performance slightly.

Figure 5. shows a box plot comparing the distribution of Mean Absolute Error (MAE) values across individual trips for each resampling method. Unlike the bar plots which show average error values, this visualization gives us a better insight into the consistency and spread of the predictions. The middle line in each box shows the typical median error for that method, whilst the edges of the box and the lines or whiskers show how spread out the errors are. These dots represent individual trips that had unusually high or low errors.

By looking at this boxplots, it is clear that rolling average and kalman smoothening both stand out with lower median errors and tighter spreads, meaning they make more stable predictions across different trips. Downsampling performs slightly worse in terms of both error size and spread, but it still outperforms fixed distance, which shows the widest spread and some high-error outliers. Default sits between downsampling and rolling average. Although its average MAE is close to downsampling, the wider spread shows that it is less reliable on certain trips. This highlights the benefit of applying a resampling technique even if the improvement in average error seems small.

As for table 5. it reveals the trade-off between training time and model accuracy. The downsampling and fixed distance methods significantly reduced the training time due to fewer data points being used in the model training, but they often led to error values that are more significant. Reversely, the Kalman smoothing and rolling average methods had longer training time durations, but in certain cases produced more stable learning.

From these results, it can be said that across all trip types, no particular resampling technique performed better than the others for all trips and cases. For instance, by smoothing out unpredictable numbers, rolling average enhanced speed on longer flights, but it reduced accuracy on shorter trips by erasing information. However, downsampling decreased training time and noise, but it also decreased accuracy for trips with a lot of transitions. Although Kalman smoothing increased average accuracy, it occasionally caused bias in areas with large volatility.

6 Conclusion

This study focused on the impact of data resampling methods on the accuracy of bicycle travel time predictions. The main research question was:

What is the impact of data resampling on the accuracy of travel time estimations, particularly for trip length and speed variation?

In order to find the answer to this question, we trained a graph neural network (GNN) model on each version of a real-world GPS dataset of bicycle trips using a variety of different resampling techniques. These techniques include: downsampling, rolling average, Kalman smoothing, and fixed distance resampling. The original raw GPS data was used to compare these resampling techniques to a baseline. The findings demonstrate that no single resampling technique works optimally for all kinds of trips. For lengthy trips where noise reduction helps with modeling broad patterns, certain techniques, like rolling average and Kalman smoothing, performed better. However, in shorter travels, these same techniques frequently eliminated valuable variety, which decreased accuracy. On the other hand, unprocessed, raw data yielded very powerful results, particularly for short and medium-distance travel.

One of the key contributions of this work is demonstrating how resampling alters the model's input features and impacts its performance over a range of trip lengths. The results indicate that trip characteristics should be taken into account when selecting a preprocessing approach and that resampling should not be used carelessly. This work also serves as a door leading to future research, which can go in a number of ways. Firstly, more sophisticated filtering techniques such as data-driven resampling algorithms or adaptive smoothing could be investigated. Secondly, it could be preferable to dynamically select the resampling method based on trip characteristics like duration, speed variance, or stop frequency rather than applying a single method to all data. Finally, only prediction accuracy is examined in this evaluation; robustness to noise, model interpretability, or training efficiency may be included in subsequent research.

In conclusion, this study emphasizes the necessity for more focused preprocessing techniques in bicycle travel time estimation, but also mobility prediction tasks as a whole. It demonstrated that choosing the proper preprocessing method has a discernible and noticeable effect on model performance.

7 References

1. Reggiani, G., Dabiri, A., Daamen, W., Hoogendoorn, S.P. (2020). Exploring the Potential of Neural Networks for Bicycle Travel Time Estimation. In: Zuriguel, I., Garcimartin, A., Cruz, R. (eds) Traffic and Granular Flow 2019. Springer Proceedings in Physics, vol 252. Springer, Cham. https://doi.org/10.1007/978-3-030-55973-1_60

2. Khushi, Matloob, et al. "A comparative performance analysis of data resampling methods on imbalance medical data." IEEE Access 9 (2021): 109960-109975.

3. Hosmer, R. (2025, April 10). Bicycles vs Cars Worldwide. Selle Anatomica. https://selleanatomica.com/blogs/homepage-blog/bicycles-vs-cars-worldwide#:~:

4. Beyond Spectral Graph Theory | TU Delft Repository. (2024.). https://repository.tudelft.nl/record/uuid:bbc5eb7b-7b5b-40da-b6c9-df0b984b215f

5. Predicting travel demand of a bike sharing system using graph convolutional neural networks. (2024). https://arxiv.org/html/2408.09317v1

6. Zhang, X., Lauber, L., Liu, H., Shi, J., Xie, M., & Pan, Y. (2022). Travel time prediction of urban public transportation based on detection of single routes. PLoS ONE, 17(1), e0262535. https://doi.org/10.1371/journal.pone.0262535

7. Amini, M.H., Kargarian, A. (Karabasoglu, O.). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. Electric Power Systems Research, 140, 378–390. https://doi.org/10.1016/j.epsr.2016.06.003

8. Pan, B., Demiryurek, U. (Shahabi, C.). Utilizing real-world transportation data for accurate traffic prediction. IEEE 12th International Conference on Data Mining, 595–604. https://doi.org/10.1109/ICDM.2012.41

9. Lin, L., He, Z. (Peeta, S.). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies, 97, 258–276. https://doi.org/10.1016/j.trc.2018.10.011

10. Yan, X., Liu, X. (Zhao, X.). Using machine learning for direct demand modeling of ridesourcing services in Chicago. Journal of Transport Geography, 83, 102661. https://doi.org/10.1016/j.jtrangeo.2020.102661

11. Cheng, L., Chen, X., De Vos, J., Lai, X. (Witlox, F.). Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society, 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002

12. Yu, B., Yin, H. (Zhu, Z.). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint, arXiv:1709.04875. https://arxiv.org/abs/1709.04875

13. Chai, D., Wang, L. (Yang, Q.). Bike flow prediction with multi-graph convolutional networks. Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 397–400. https://doi.org/10.1145/3274895.3274958

14. Zi, W., Xiong, W., Chen, H. (Chen, L.). TAGCN: Station-level demand prediction for bike-sharing system via a temporal attention graph convolution network. Information Sciences, 561, 274–285. https://doi.org/10.1016/j.ins.2021.01.065

15. He, S. (Shin, K.G.). Towards fine-grained flow forecasting: A graph attention approach for bike sharing systems. Proceedings of The Web Conference 2020, 88–98. https://doi.org/10.1145/3366423.3380151

16. Huang, F., Yi, P., Wang, J., Li, M., Peng, J. (Xiong, X.). A dynamical spatial-temporal graph neural network for traffic demand prediction. Information Sciences, 594, 286–304. https://doi.org/10.1016/j.ins.2022.02.031