Chapter 5

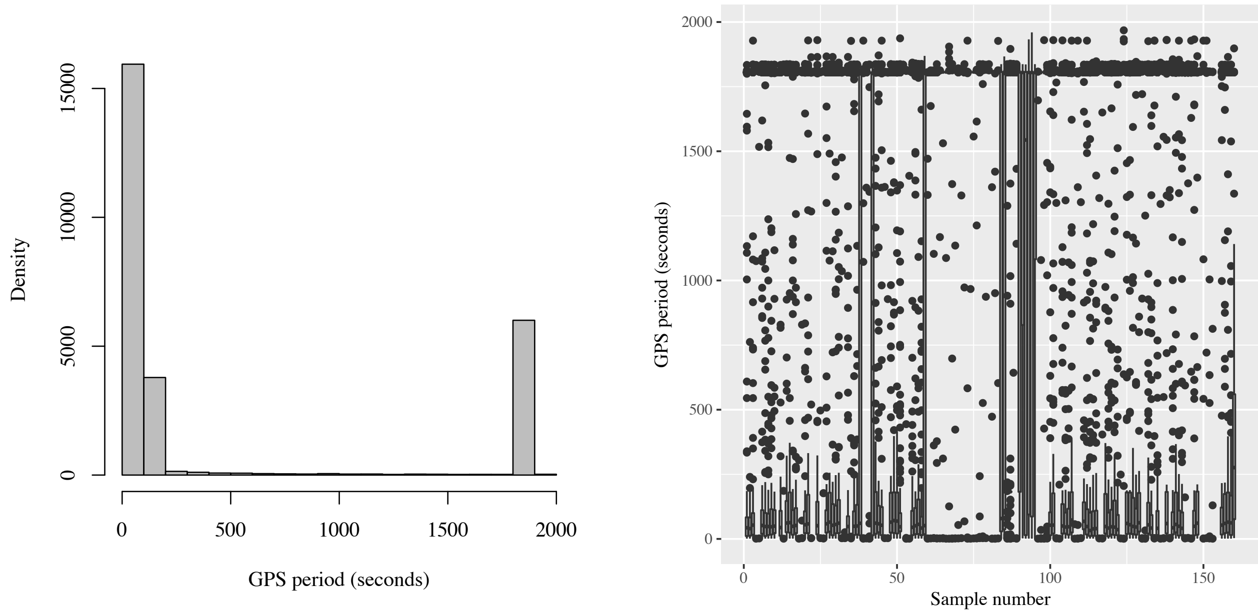
Real-world data analysis

This chapter reflects how the map-matching algorithm developed was tested not only on an a real-world network, but also on actual data gathered from a previous study on the waste collection vehicles and activities in the City of Cape Town. The dataset available for this study was the complete set of global positioning system (GPS) trajectories for a specific waste collection vehicle during the period December 2013 to July 2014. The sampling rate for GPS trajectories were very dynamic and had no fixed time interval between records. The waste collection vehicle tracking system records GPS points more frequently if the vehicle is moving or experiencing acceleration or deceleration. If the vehicle is motionless, but the engine is running, the sampling rate decreases significantly, if the engine is switched off the recordings are 30 min apart.

5.1 Exploratory data analysis

An initial analysis of the raw data revealed that the data required a significant amount of cleaning before it could be used. As a first pass, all GPS recordings which showed a GPS period of greater than 33 min were removed because it is known that when the vehicle is stationary recordings are suppose to be made every 30 min. A total of 46 points where removed from the 26824 points in the dataset. Figure 5.1a illustrates the distribution of the GPS period for all the samples in a histogram and it is clear that the data contains a significant amount of time where the vehicle is switched off. These parts of the data had to be removed as the algorithm would not provide useful information for these periods as all the points will be at the same location. In these circumstances the map-matching algorithm will infer that the vehicle was driving in circles to create such a trajectory of GPS points and provide anomalous outputs.

For Figure 5.1b, the data set was broken down into separate days and stored as separate samples. As expected, all samples had a wide variety of GPS sampling rates as well as numerous entries with rates close to 30 min, representing the times during the day that the vehicle was not in service. Some of the samples even only had data entries close to 30 min intervals, suggesting the vehicle was out of service that day, and only one sample had all entries less than 800 s or 13 min apart. Since an algorithm for clustering and cleaning raw GPS trajectory data was outside the scope of the current study, this sample was used for a detailed analysis of the map-matching algorithm's performance.



(a) Overall distribution of sampling rate

(b) Distribution of sampling rates over consecutive samples

Figure 5.1: Exploratory data analysis of waste collection vehicle trajectory data

5.2 Map-matching on actual trajectories

Figure 5.2 displays the GPS points of the sample trajectory as well as the inferred path (IP) the algorithm matched. Since there is no true path (TP) to match the IP with, a subjective visual analysis of the match was done. The alternative options include manually creating a TP from the GPS points by making use of expert knowledge, for example asking experienced drivers to map the most likely route that a driver took in that area. This is similar to the approach discussed in section 2.2. It is typical for a waste collection vehicle to travel from one service area to the next, called *deadheading*, close to or lower than the speed restriction of the road it is travelling on. Once they reach the service area, they will travel at a much lower speed to enable workers to load the vehicle with waste. The portions of the trip that the vehicle was most likely servicing an area, versus deadheading, can be identified by the road segments where it was travelling far below the free speed.

Figure 5.3 plots the inferred vehicle speed on the road network in the top image and compares it to free speed of the network in the bottom image. This analyses clearly illustrates the areas where the vehicle was most likely collecting waste by comparing the free speed of the links to the inferred speed. Figure 5.4 plots the inferred speed versus the free speed across the entire inferred route of the vehicle. It can be noted that there are some significant outliers in the inferred speed which most likely indicates incorrectly matched links. These discrepancies occurred where the GPS points were recorded within close proximity to each other and the algorithm incorrectly identified U-turns and loops in the matched route, as can be seen Figure 5.5a. When comparing the abnormal inferred speeds to the probabilities it can be noted that the areas with abnormal inferred speeds also have low probabilities. There also appears to be sections with low probability values that do not have abnormal speeds and where the results seem credible.

The probability of incorrect matches occurring when data points are clustered close together is due to the inherent GPS error that can cause certain points to be recorded in

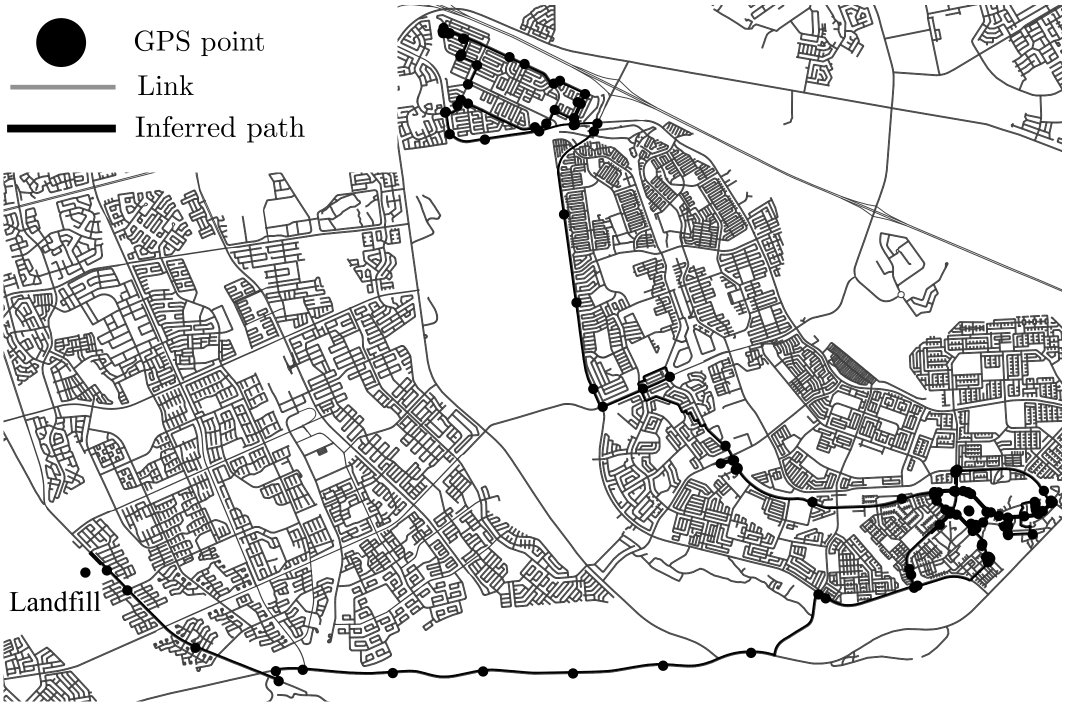


Figure 5.2: IP of real-world data sample

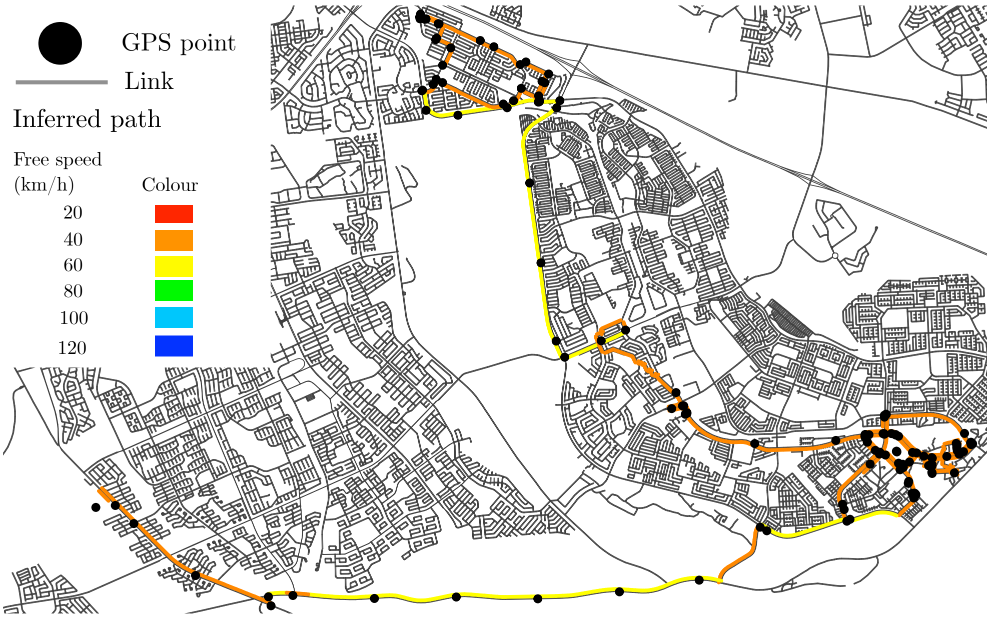
such a way that it appears as if the vehicle traveled backwards, which can sometimes be the case with waste collection vehicles. The algorithm will determine that a U-turn was made and that a circular route is the only feasible solution. This phenomena can either be addressed by cleaning up the data to remove unnecessary data points, which will also yield a decrease in computational time, or the algorithm should be improved to detect such situations and a sensitivity defined to ignore such cases.

5.3 Conclusion and future work - Case Study

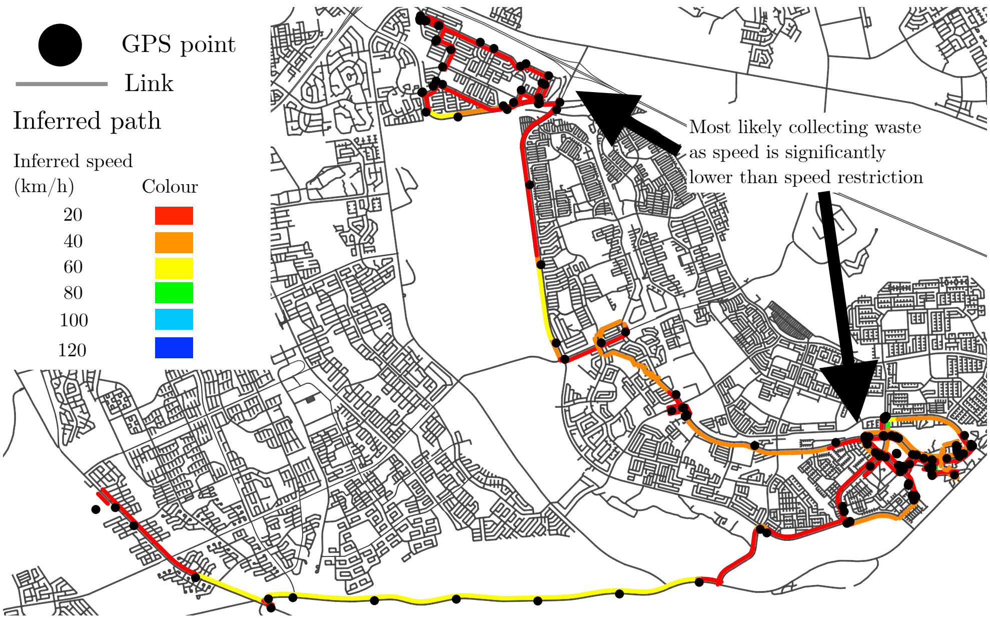
In analysing the waste collection sample, the algorithm is able to match GPS trajectories to a subjectively acceptable level of accuracy, but it seems it cannot handle situations where GPS points are recorded in close proximity. The algorithm is able to accurately match trajectories on a real-world network more accurately compared to an experimental grid network. The algorithm appears to be very robust and can handle dynamic samplings rates.

From the algorithm output, we are able to use the calculated speed as a proxy to infer where a waste collection vehicle was most likely collecting waste and where it was simply deadheading. This proves the algorithm is successful in the original requirement of the project in both the primary and secondary objectives set.

The study also provided a novel way of using the probabilities from the spatial-temporal (ST) algorithm to analyse the results in the absence of a TP. The use of the probabilities together with the inferred speed versus free speed could provide substantial insight into the confidence level of the IP. This comparison between inferred and free speed could be



(a) Speed restriction of network



(b) Inferred speed of vehicle

Figure 5.3: Comparison of inferred speed versus free speed on the IP

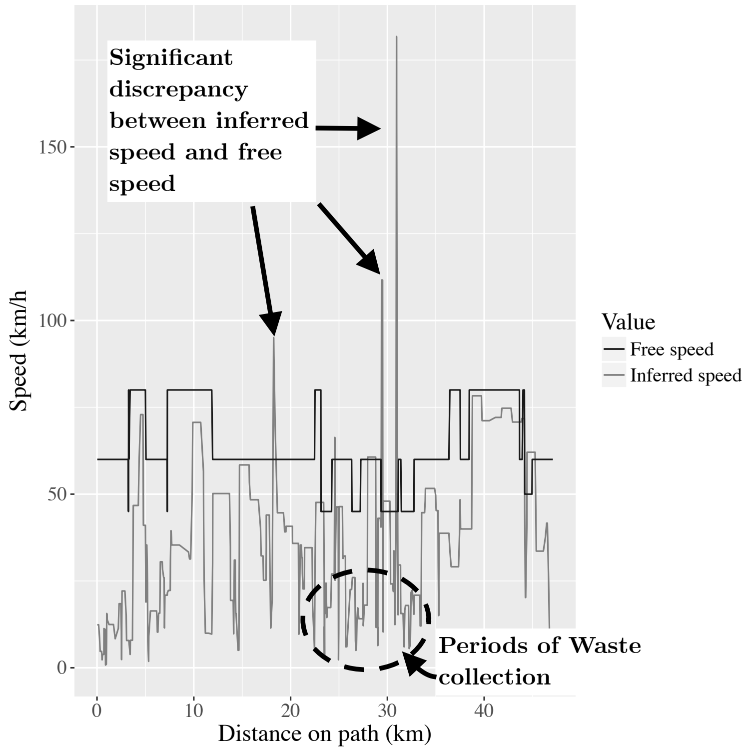
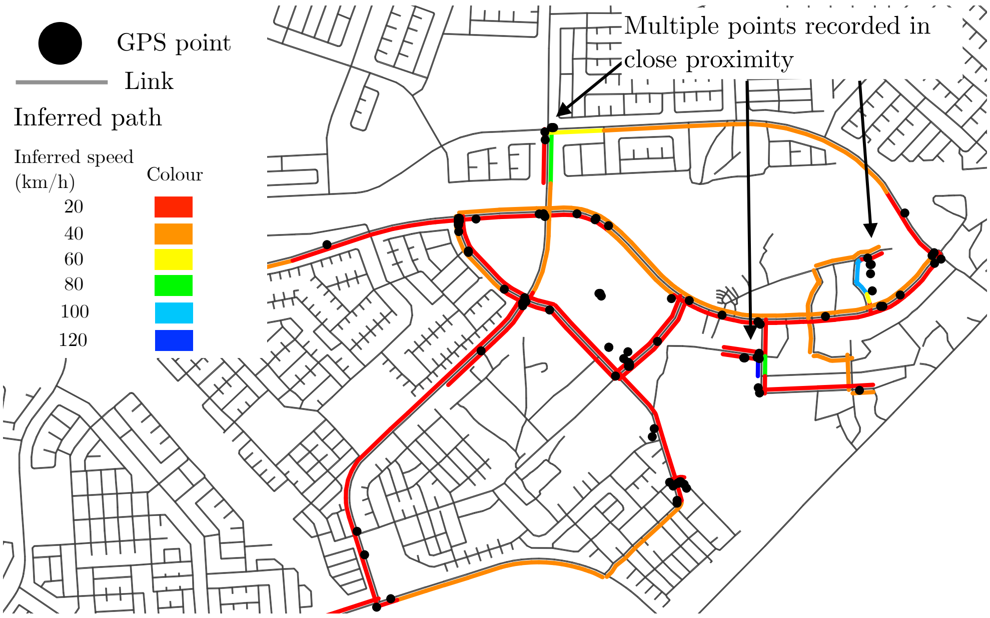
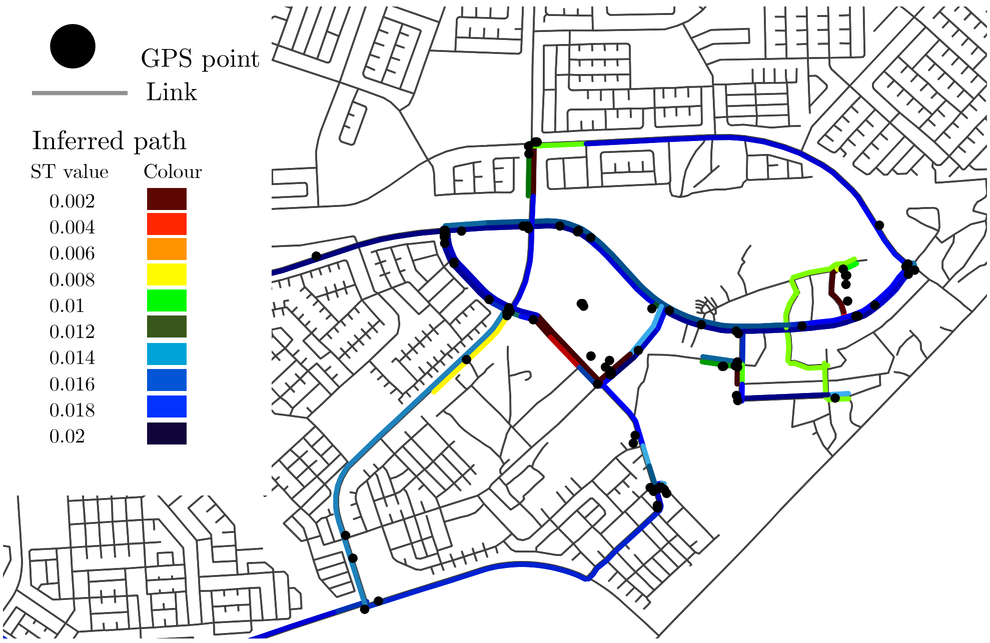


Figure 5.4: Inferred speed versus free speed on inferred path



(a) Abnormal inferred speeds



(b) Probability analysis of abnormal inferred speeds

Figure 5.5: Analyses of abnormal inferred speeds

improved by making use of historical actual speeds on the network. The historical speeds can also be used in the temporal probability calculation which will provide the algorithm with a high degree of accuracy.

Future work on waste collection data will have to include a significant level of data cleaning in order to get trajectories that only represent a single trip of an object and possibly remove or smooth out sections where points were generated in close proximity.