UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Final Project Report

Developing a Residential Waste Generation and Service Analysis Model

Llewellyn James Steyn

supervised by
Elias Willemse

September 28, 2016

Abstract

Municipal solid waste collection is a costly service provided by governments all over the world. As a result of the costs and operational complexity associated with this service it has become the focus of many studies on optimisation. Key to successfully optimising the collection of waste is having access to data relating to waste generation rates and service levels. These differ significantly from city to city and therefore developing waste generation models for large metropolitan cities is crucial. Not only do effective waste generation and service analysis models aid in developing optimisation models but they also lead to better strategic and operational decision making on the part of city planners. This paper used GPS traces collected from the City of Cape Town's waste collection fleet to estimate the amount of waste produced per suburb, the mean service time per service area, the mean travel times between landfills and service areas and finally the velocity distributions inside and outside service areas. Results show that a waste generation and service analysis model is an easy, accurate and cost effective method for developing vital operational and strategic waste management data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and background

1.1 Introduction

Most people are familiar with the weekly chore of taking out the trash, it is one of the realities of every day life. However people seldom think about their trash for very long. In a well functioning municipality solid waste is collected from residential and commercial properties periodically. It is therefore a service that we often take for granted. The Constitution of the Republic of South Africa, considered one of the most progressive globally, makes provision for human dignity and a clean and safe environment for all [12]. Municipal solid waste collection is therefore implicitly a constitutional obligation to which municipalities must adhere. Furthermore, the Public Finance Management Act 1 of 1999 [13] obliges provincial and local governments to manage all resources efficiently and effectively. The City of Cape Town spends an estimated two billion rand per annum on municipal solid waste collection [11]. Given the fiscal pressures experienced by the South African Government as a result of economic stagnation, a greater emphasis is being placed on efficient, cost effective service delivery. This, coupled with rapid urbanisation, means that municipal solid waste collection is increasingly becoming a focus area for optimisation. As a result, numerous models have been developed by the University of Pretoria's Center for Transport Development, however the effectiveness of these models are limited by the current waste generation and service data available.

Ghiani [5] considers the literature on operations research as applied to municipal solid waste collection in a review article. In the paper he describes models for determining collection districts such that the total waste in a collection district equates to the collection capacity allocated to the district. Furthermore he reviews studies on collection patterns, in particular the frequency of waste collection. The common denominator in almost all waste related models is that of generation rates. Ghiani admits that despite the importance of good waste generation data "very few papers deal with it". Waste generation rates are the foundation of good waste collection optimisation yet it is still a novel research area.

This is in part due to the fact that estimating and predicting solid waste generation rates is very difficult. In contrast to other municipal services such as water and electricity, solid waste can not be measured down to a household level. The supply of water and electricity is usually done in conjunction with metering which gives municipal managers and policy makers accurate consumption data. In contrast it is impractical and expensive to measure each household or commercial enterprise's waste as it is collected. As a result there is very little data on solid waste generation

rates. While there might be waste data on a municipal level (using weigh bridges at landfills for example) very little data exists on a household or suburb level that can aide in decision support. This has the obvious effect of making solid waste collection systems incredibly hard to plan and manage both on a strategic and operational level. The need for accurate, reliable solid waste data is therefore imperative to the effective and efficient functioning of any solid waste collection system.

This project aimed to develop a descriptive model in order to understand and forecast waste generation rates in the city. In addition GPS data used for waste generation modelling was used to do service analysis to help waste managers with decision support on both a strategic and operational level. Since waste management is a service provided all over South Africa and the world this study has the potential to be applicable to many cities and municipalities.

1.2 Problem statement

In considering literature it becomes apparent that while waste generation rates can be estimated using existing techniques, none of these techniques satisfy the problem completely. Currently there is a gap in literature in terms of a waste generation model that will improve the optimisation of collection routes, scheduling, and strategic decision support. Currently there are no models that satisfy the following criteria:

- describe waste generation and vehicle service patterns with a reasonable amount of accuracy down to suburban level;

- make use of recent data to allow ongoing analysis of waste generation levels;

- are simple and easy to construct and maintain on an ongoing basis.

1.3 Research design

The outcome of this project was a descriptive statistical model that accurately reflects the waste generation rates in the city down to a suburban level. Due to the fact that weighbridge data was unavailable the data used from landfills is fictitious. The model therefore simply illustrates the feasibility of using GPS data for waste generation modelling. As output the model gives the mean waste generation rates per suburb over a period of time. Coupled with this it gives the maximum and minimum generation rates. The model also performs service analysis and produces key indicators of service levels. These include average service times per service area, travel times and information regarding vehicle velocity. These are based on actual data and have enormous potential for city planners. In terms of both waste generation and service analysis the model will have a number of distinct features that makes it useful. The first of these is the ability to duplicate the model with relative ease; any city with GPS trackers fitted on its waste collection vehicles and weigh bridges at its landfills will be able to build a similar model. The model will produce appropriate results in a short amount of time. As opposed to source sampling for example the model can be continuously updated with the latest data to give the user current waste generation and service data. Lastly the model acts as a pilot study for more detailed research into the area of waste generation modelling. Using a similar approach it is likely that subsequent models with improved accuracy could predict waste generation patterns down to a street level.

1.4 Research methodology

The development of an accurate and robust waste generation model followed the steps described below.

1.4.1 Identify facility visits using waste vehicle tracking data

The first step involved understanding when waste collection vehicles visit landfills and waste processing facilities. The location of the city's twenty five waste drop off facilities are known. Using GPS traces any waste collection vehicle that enters these areas and spends time in the area is deemed to have visited the landfill for a drop-off. To identify when a vehicle has made a facility visit the work of Cich et al [2] will be expanded on. The author develops a method for separating GPS traces into trips and stops. In the study the finding is made that predicting stops is a lot easier and more accurate than predicting trips. Using this it is possible to identify, with calibration, when waste vehicles have made stops. However it is conceivable that a stop does not definitely mean a facility visit and for this reason only stops within a reasonable radius of the facility were considered facility visits. The possibility also exists that simply using a co-ordinate's proximity to a facility will be sufficient in identifying facility visits. This will be further discussed in the results chapter.

1.4.2 Use waste vehicle tracking data to link vehicles to service areas

The city has waste collection schedules published on its website, it is therefore easy to determine which areas should have been serviced on a particular day. Given this knowledge when a truck enters an area scheduled to be serviced and reduces speed while in this area, the area will be deemed serviced by that particular truck. In determining the average speed of the collection vehicle through an area the framework for determining instantaneous and average speeds from GPS traces developed by Jimenez et al [7] was used. A is defined as the vertical distance between lat_2 and lat_1. B relates to the horizontal distance between $long_2$ and $long_1$. The absolute distance between the two sets of co-ordinates is then defined as d:

$$A = 69.1 \times (lat_2 - lat_1)$$

$$B = 69.1 \times (lon_2 - lon_1) \times \cos(lat1/57.3)$$

$$d = \sqrt{A^2 + B^2} \times 1609.344$$

The next step is determining the time, t, between two GPS points. This is described by Jimenez as follows:

$$t = (ct_2 - ct_1) \times 86400$$

Finally the instantaneous velocity, v, is calculated as follows:

$$v = \frac{d}{t} \times 3.6$$

It is important that both criteria are applied as a vehicle might enter a service area but only pass through it and not collect any waste. Simply looking at average speed through a segment will also not stand up to scrutiny. There are many factors such as traffic or a breakdown that can potentially

lead to a low average speed over a segment without any collection taking place. By applying both criteria the probability of predicting serviced segments is high. It must be emphasised at this point that calibration will be critical to this step of the project in order to identify a velocity interval that reasonably represents a vehicle picking up waste. It is also likely that when the comparison between velocities inside and outside the service areas is made, that there will be enough of a difference to use velocity as a proxy for service taking place. The model will then simply use the vehicle's speeds to predict whether or not the vehicle is within a service area.

1.4.3 Link collection vehicles to weighbridge data

The next stage of the project involved linking collection vehicles to weighbridge data. At this stage it was known which vehicles made a facility visit at which time, as determined in step 1. The weighbridge data for a particular facility could then subsequently be linked to a collection vehicle.

1.4.4 Link waste collected to service areas

It follows from the previous step that the waste collected and measured at the facility weighbridge can be linked to a truck and through step 2 to service areas. At this point the problem may arise that a truck services more than one area in a single trip, meaning that the waste must be apportioned to the service areas. This was done by apportioning waste relative to the amount of distance covered per service area or, alternatively, the amount of time spent within the service area. For future work it is envisaged that a synthetic population, based on census data, can be used to apportion waste.

1.4.5 Model validation

Model validation was done by taking sample data and manually going through the routes that vehicles follow. If the model is able to accurately identify known service areas and landfills then the model will be considered validated. Throughout the model building process data was generated at regular intervals. This data was continuously scrutinised for validity. This process as it happened in practise is explained in the conceptual model chapter.

1.4.6 Experimentation

Upon completion of the calibrated and validated model the model was used to run a number of experiments using a more significant data set than during the model development phase. Key decisions during this phase included the selection of service areas and the selection of sample days.

1.4.7 Extract waste generation rates

The resulting data from experimentation was then used to extract mean waste generation rates for selected suburbs over the study period.

1.4.8 Extract service data

Finally the vital data relating to service patterns was generated. This includes average service time per service area, mean travel duration between landfills and service areas, mean landfill duration and mean vehicle velocity.

1.5   Document outline

Following this the document will take the following format:

- A Literature review is included in chapter 2 that explores the problem environment and potential solutions to the problem.

- Thereafter the conceptual model is discussed in chapter 3. Each model component is explained and the results it produced during the developmental phase is discussed. The model is also validated before any conclusive results are discussed.

- In Chapter 4 the experimentation phase is explained. Included in this chapter is detail on the sampling decisions taken to produce the results reported on.

- Chapter 5 deals with model results, detailed results is given and the significance of these results are explored.

- Finally in chapter 6 a conclusion is reached on the success of the project and in particular on the feasibility of using GPS data for waste generation modelling and service analysis. The appendix contains all relevant supporting documentation and source code for the model subcomponents.

Chapter 2

Literature review

In order to define the problem at hand it is important to give a clear, concise idea of the existing solutions to the problem of waste generation modelling. The existing body of literature will be discussed and subsequently considered for relevance to this project. At first literature will be considered that have delivered tangible results on waste generation modelling, this aims to answer the question of whether it is necessary to duplicate a waste generation study for the city of Cape Town. From there existing methodologies will be considered in the hope that these can be duplicated or expanded for the City of Cape Town. Finally consideration will be given to literature on handling GPS data. While the application of these studies might differ, their methodologies are relevant.

2.1 Existing studies on solid waste generation rates

Beigl et al [1] in reviewing municipal solid waste generation finds that at settlement level (or suburb level) data sources include census data, market research based on geo-demographic classification packages and questionnaires. Some of the more significant studies that use these data sources are discussed below. The focus in citing these sources are on the results they have produced. Greater detail on the methods used is given in the next section.

The first study by Qdai et al [15] used source sampling at 40 different households to determine a waste generation distribution. The study found that waste is produced at a rate of 1.76 kg per capita per day in Abu Dhabi city. Minghua et al [10] in a study in the city of Shanghai, China, produced a result of 1.11kg per capita per day. There are numerous similar studies in literature where different source sampling techniques are used. It is therefore clear that solid waste generation rates differ significantly from city to city. The generation rates in Kathmandu are ten times less than in Abu Dhabi. This is because there are many factors that contribute to solid waste generation rates. Liu et al [9] finds economic growth and urban development to be some of the major factors. In light of this it is clear that waste generation rates are unique to a city or district, further motivating the need for the City of Cape Town to build their own waste generation models. The implications of incorrectly estimating waste generation rates are severe. Due to the cost of developing the infrastructure to support waste collection (truck depots, waste transfer stations, waste sorting and compounding facilities, landfills etc.) it is important that waste generation rates are not just understood but can be extrapolated to understand future demand.

© University of Pretoria

2.2 Waste generation methodologies

Discussed below are some of the methodologies that have been successfully applied in estimating waste generation numbers. These include source sampling and systems dynamics modelling.

2.2.1 Source sampling techniques

Although numerous studies have been conducted on municipal solid waste at a regional or district level the idea of estimating waste down to a household or street level remains largely uncharted. Nevertheless some attempts have been made. An example of this is the use of stratified cluster sampling by Dangi et al [3]. The study makes use of a team of student scientists to physically visit households in Kathmandu Municipality, Nepal, to weigh and sort solid waste. The study clearly followed sound scientific method and reasoning. The waste generation rates that resulted from the study can be concluded to be a valid reflection of waste generation patterns in the particular municipality during the two weeks that the samples were taken. This allows city planners with no statistical data on waste generation to make more informed strategic and operational decisions. The study estimated waste generation in the city of Kathmandu at 161.2g per capita per day. The limitation of the study is of course that it only considers a two week period. No indication of seasonal generation rates is therefore present. The study makes use of a large number of students equipped with scales and bio-hazard gear to perform the sampling. It is therefore impractical and costly to use the sampling method described as an ongoing source of waste generation data.

2.2.2 Systems Dynamics Modelling

Another proposed solution to the problem of estimating waste generation rates is that of systems dynamics modelling. Systems dynamics is a well established methodology that was developed by Forrester [8]. Dyson et al [4] uses systems dynamics modelling to predict waste generation rates. This method considers population growth, household income, people per household and economic activity to build a number of system dynamics models that attempt to predict waste generation rates. Data is extracted from census data in order to build the models. The output is then an estimated solid waste generation rate in tons per year for the the city of San Antonio. Because the study makes use of census data it is easily replicated for other cities. However it offers more of a strategic indication of waste generation rates. This can be used in the planning of new landfills or recycling schemes but is less useful on an operational level. For a model to make a difference in route optimisation for example, as is proposed in this project, waste generation rates must be understood at a suburban level at least.

2.3 Handling GPS data

In contrast to waste generation modelling a lot of literature exists on interpreting and working with GPS traces in order to gain meaningful insight from this data source. Some of these papers have already been cited but not discussed. They are critical in solving the problem at hand and will therefore be discussed below.

Cich et al [2] looks at person traces as opposed to car traces in order to sort GPS points into trips and stops. Two methods are proposed, one that focusses on trips and one that focuses on stops. Identifying stops was found to be easier due to the large variability in trip speeds. In

addition to this sensitivity analysis is used to find the model configurations that produce the most accurate results. While the study relates to a completely different subject matter the methods are very similar to what will be used in identifying facility visits in this paper. Sensitivity analysis will also play a pivotal role in determining model parameters and therefore this paper has value from this perspective.

Wilson et al [17] attempts to estimate waste transfer station delays using GPS data. Five trucks are selected and then tracked while they go through the process of unloading waste at the transfer stations. The results of the study are useful in the sense that they give an indication of how long typical waste collection vehicle spends at a transfer station (average of 16.4 min). This is a useful starting point in calibrating the model proposed in this paper in identifying facility visits. However the true value of this study lies in the use of Geofences to separate GPS points into different areas within transfer stations. In determining whether or not areas have been serviced a similiar approach will have to be adopted to determine whether a vehicle has entered a service area. In this sense the methods and results discussed by Wilson et al has a direct bearing on this project.

2.4   Service analysis

When considering literature on service analysis it is clear that there is a lot of literature on assessing the overall effectiveness of municipal solid waste systems. Pires et al [14] looks at the systems engineering and systems assessment tools that are applied to municipal solid waste management. While these are of great relevance to South African municipalities they focus more on interpreting and sharing existing information on waste management systems and less on developing new ways of collecting information. The same can be said of Rogge et al [16] that evaluates the effectiveness of waste collection using a data envelopment analysis (DEA) model. While this too has application for the city it does not adress the shortcoming of where the data will actually come from.

Huang et al [6] however provides more relevant insight. While the focus is still on overall performance assessment and not on data gathering techniques the authors provide very useful detail on potential waste collection KPI's (Key Performance Indicators). The five KPI's that were selected for input into the aggregate indicator proposed by the authors include:

- cost per unit volume of municipal solid waste collected;

- quantity of municipal solid waste collected per unit collection time;

- quantity of municipal solid waste collected per collection vehicle;

- population served per collector;

- quantity of municipal solid waste collected per vehicle mile.

This is of relevance to this project as it gives an indication of which aspects of operational information is of most use to assessing the performance of a waste collection system.

2.5   Conclusion

The literature review proves that there is a significant gap in literature with regard to collecting accurate decision support information for waste collection assessment. Existing methods of

estimating waste generation rates are effective on a macro level but lack the detail required by optimisation models. The use of GPS data is well researched and the methods of handling GPS data are thoroughly tested, the application of GPS data to waste collection modelling remains a gap in literature though. For this reason the project attempts to address these gaps.

Chapter 3

Conceptual model

3.1 Analysis of input data

The first step in developing the conceptual model is in understanding the data that the model will use. GPS trace data, as provided by the city, has the following key characteristics. Each element of the data set contains a vehicle registration, a time stamp that reflects both date and time and finally the longitude and latitude of the vehicle at that point in time. It is the unique combination of these five fields that allows the conceptual model to be constructed. The GPS devices on the waste collection vehicles sends a regular location update while the vehicle is switched on, when switched off the GPS updates its location every thirty minutes.

The conceptual model was developed using the GPS trace data of a single day, meaning it contains the movements of multiple vehicles over the day. In developing the conceptual model great care was taken to make the model flexible to the data input. For this reason the model is capable of handling large data sets, similar to the one used during experimentation. If the data is in the format explained above, regardless of the metropolitan area, the model will produce credible results. This makes the model reproducible and therefore of great use for future work.

3.2 Model explanation

The basic philosophy behind the model is to convert unstructured GPS traces into a meaningful data set that can then be interpreted. Initially the data set contains the registration number of the vehicle as well as the date, time, longitude and latitude of the particular point. This information in isolation is pretty insignificant and for that reason the model aims to supplement this data with more meaningful fields. Each point is classified into one of three possible statuses. These are "Visiting Facility", "Servicing" or "Travelling". This is based on the assumption that a single GPS point can only represent the vehicle performing one of these three functions. In addition, these three activities are viewed as mutually exclusive in the sense that the occurrence of one excludes the occurrence of the other two. This means that if a vehicle is in a service area it cannot then also be visiting a landfill. The assumption that a point cannot be both "Servicing" and "Travelling" is less accurate however. This is because it is very likely that a vehicle passes through a service area without collecting waste. More detail is given later in dealing with this issue. "Status" therefore represents the first meaningful data attribute that is added to the data.

Once a GPS point has been classified as either "Travelling", "Servicing" or "Visiting Facility"

the next step is to identify the landfill or service area that corresponds with the point (when the point is not classified as "Travelling"). For this reason the next data attribute added to the data set is "Location Name". This is the unique name by which a landfill or service area can be identified. A GPS point therefore belongs to a specific service area or landfill when not classified as a travelling point.

The next vital piece of information that the model adds to the data set is velocity. Velocity is crucial as it could lead to detailed analysis of vehicle behaviour. For this reason the instantaneous velocity of the vehicle is added as a data attribute. Velocity is calculated using the method described by Jimenez [7] in the research methodology.

To assist in interpreting vehicle velocities the vehicle velocity is classified into different categories. This represents the final piece of information added to the data set. These additional fields help in interpreting the data and making useful conclusions from the data set. Table 3.1 shows what typical data entries look like once this process has been completed.

The next step then is to identify "trips". A trip is defined as all the points that share the same status. Therefore a trip could be a vehicle travelling from service area to a landfill, a vehicle travelling between a landfill and service area, a vehicle travelling within a service area or the time a vehicle spends at a landfill. This stage of the model is crucial as it enables both the waste generation modelling and service analysis to take place. Isolating and identifying trips is discussed in more detail later on.

Finally the waste generation rates are calculated using the trip data just mentioned. Each time a vehicle makes a visit to a landfill waste is apportioned to the vehicle. This is then split amongst service areas that were visited prior to the vehicle visiting the landfill. This in discussed in depth in the sections that follow.

Table 3.1: Amended data output from the model

Reg	Date	Time	Long	Lat	Status	Location	Velocity	VelocityStatus
CA1234	14/07/02	14:32:45	18.456	-34.563	Servicing	Dunoon	2.3456	1-10km/h
CA2387	14/07/02	18:52:50	18.444	-34.512	Travelling	NA	2.3456	1-10km/h
CA1463	14/07/02	04:32:45	18.412	-33.444	VisitingFacility	Coastal	2.3456	1-10km/h

3.2.1 Identifying facility visits

From the city's website the location of waste drop off facilities can be established. The aim of this first part of the model is simply to identify instances where a waste removal vehicle visits one of these 24 facilities. The most important criteria is the proximity of the vehicle to the facility, as a result the distance from every GPS point to every single facility is calculated using the method described by Jimenez et al [7]. This is done using the "AssignStatus.R" script as found in the Appendix. If the distance falls within the estimated radius of a facility then the function returns true and the GPS co-ordinate is considered part of a facility visit. The result of this process can be seen in figure 3.1. Blue points represent points that the model considers a facility visit while green points are considered "Travel" points. This is at a radius of 200m, it is therefore clear that the radius directly affects the model sensitivity. The accuracy of this aspect of the model can be improved with relative ease by using geofences similar to those used for service areas. While this will lead to a higher degree of accuracy it comes at the expense of processing time. For the purpose

Figure 3.1: Model identification of facility visits(blue points)

of this study therefore using proximity as a test is sufficient. The point is then also updated with the name of the facility under the "LocationName" attribute.

3.2.2 Identifying service areas

The next step was to address the more challenging task of identifying when a vehicle services a particular area. In determining this there are two important aspects to consider. The first is whether or not a vehicle is in an area to be serviced on that day. The second is the speed at which the vehicle travels within that area. Given that a truck is in a service area for a reasonable amount of time and lowers its speed while in the suburb it is safe to assume that this area is being serviced by the vehicle.

However for this portion of the model the only consideration is whether a point falls within a service area. Instantaneous velocity is discussed at a later stage, and the application of both to determine whether or not a vehicle has serviced an area is addressed later.

In determining whether a point falls within a service area the first step was to draw geofences around service areas. This is a manual process and therefore time consuming. When using the model for experimentation during the next phase of the project a larger sample of nine service areas was used.

The model considers every single point in the data set and then compares whether or not it falls within any of the service areas' geofences. If it does, the point is considered "Servicing". The location of the point is then also updated with the service area that the point falls within. Figure

Figure 3.2: Classification of service points(red) at Vrygrond service area

3.2 illustrate this as applied to the Vrygrond service area. The full "AssignStatus.R" script in the appendix should be consulted to gain an understanding of the detail of this process.

3.2.3 Vehicle velocity

The next model aspect to be addressed is that of vehicle velocity. This is a crucial piece of information with direct application to analysis of vehicle behaviour later on. An instantaneous velocity is calculated for each point using the change in distance over time between the point and the point preceding it. To achieve this the model removes all the GPS points for a particular vehicle and calculates the velocities for all points. In order to aide in the interpretation of vehicle velocity the model then also categorises velocities into one of four categories. These are namely 1-10km/h, 10-40km/h, 40-80km/h and 80-120km/h. All velocities above 150km/h are considered outliers and are labelled as errors. At this stage it is unclear what might cause these outliers other than faulty GPS readings from the GPS devices. "AssignVelocity.R" is the script that assigns the instantaneous velocity to the data.

3.2.4 Trip identification

The next model component is the identification of "trips". A trip can be one of three things, travelling between service areas and landfills, travelling within service areas and finally travelling within a landfill. This is crucial not only to determine waste generation rates but also for the service analysis component of the project. Each trip has the following attributes: a vehicle registration, the date the trip took place, the type of trip (Travel, service or facility visit), when and where the trip starts, when and where the trip ends and finally the trip duration. The "Trip.R" script does this by identifying every time the vehicle status makes a change. This could be travel to service, service to travel, travel to facility visit, facility visit to travel, service to facility visit and

finally facility visit to service. Any of these changes in status signals the end of a trip and the start of a new trip. By identifying trips the model is then able to convert GPS data into something meaningful and simple to understand. Table 3.2 is a summarised version of what the "Trip.R" sub model does to all vehicles over the sample period. See figure 3.3 for a visual representation of table 3.2. To understand the detailed mechanics of this process it is suggested that the "Trip.R" file in the appendix is consulted. It is from the data produced during the trip identification phase that the majority of the service analysis comes from.
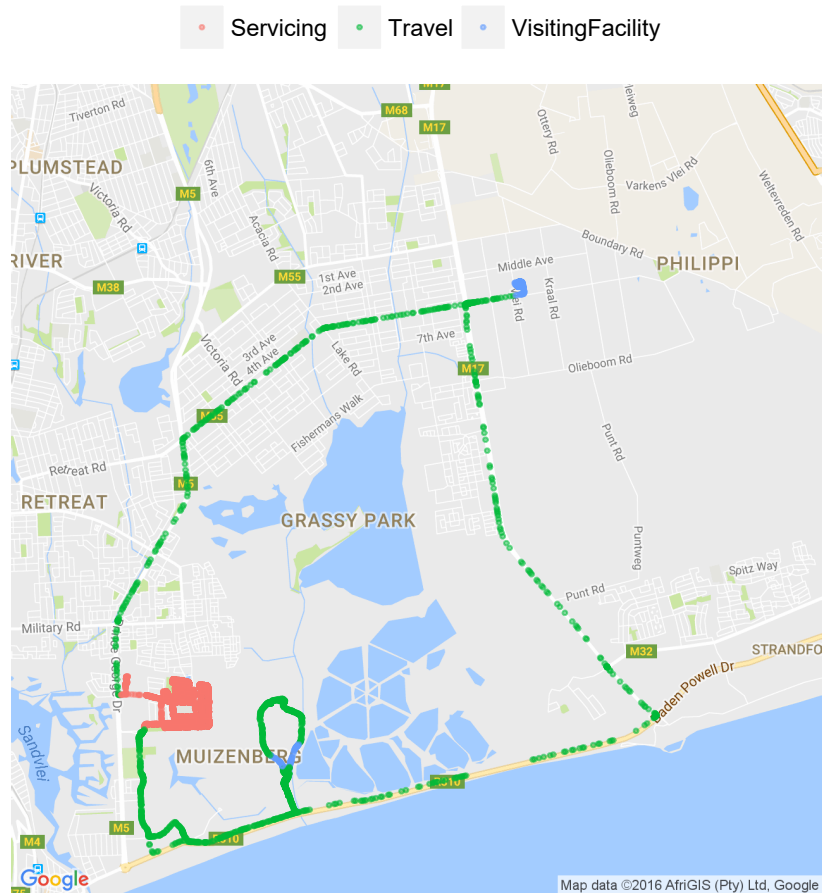
Figure 3.3: Vehicle A Status

3.2.5 Waste generation modelling

The attributes contained in a trip then make it rather simple to finally perform the waste generation modelling. "WasteGeneration.R" says that if a vehicle visits a landfill and spends a reasonable amount of time there (meaning it drops solid waste off), and the vehicle passed through one or more service areas (and also spent time within the service areas, meaning waste was actually collected) prior to visiting the landfill it is safe to assume that the waste weighed at the landfill can be allocated to the service areas. If there is more than one service areas that has likely been

Table 3.2: Vehicle A route breakdown

Location	StartTime	End Time	Duration
Schaapkraal	00:33:07	06:30:29	05:57:22
Travel	06:30:29	06:41:45	00:11:16
Vrygrond	06:41:45	07:52:05	01:10:20
Travel	07:52:05	08:16:50	00:24:45
Coastal Park	08:16:50	08:29:24	00:12:34
Travel	08:29:24	08:35:08	00:05:44
Vrygrond	08:35:08	09:32:35	00:57:27
Travel	09:32:35	09:42:34	00:09:59
Coastal Park	09:42:34	09:53:36	00:11:02
Travel	09:53:36	10:07:27	00:13:51
Schaapkraal	10:07:27	23:37:36	13:30:09

serviced before a vehicle makes a visit to a landfill the model apportions the waste. The ratio that is allocated to each service area is a function of the amount of time spent within a service area. This is not the only way that waste can be apportioned, for example distance travelled within service areas or service area population could also be used. For the purpose of this study however duration proved a sufficient attribute by which waste can be split amongst service areas. This is because the assumption is made that the amount of waste collected is a function of time.

3.3 Model validation

It is critical at this stage to scrutinise the new data set produced by the conceptual model for validity. In doing this a number of questions must be answered. These are namely whether the data makes logical sense in the current context? Furthermore, does the model explain the reality of vehicle movements with significant accuracy? These questions are answered in the sections that follow.

3.3.1 Map accuracy

The first area of interest is the accuracy of the GPS data. Despite the fact that the study focuses on the service area as the smallest area of interest and study, it is still crucial that the GPS traces used are accurate down to street level. This not only aids in the accuracy of the results produced by this project but also allows future waste generation studies for the city to delve down to the more accurate street level. Figure 3.4 shows the degree to which points correspond to roads on the map. This same figure was reproduced over multiple days and locations to validate that the GPS date is indeed accurate enough. For the purpose of this study it was found that the GPS traces proved significantly accurate.

3.3.2 Velocity accuracy

All GPS devices produce errors. As with anything under operational conditions errors. The question however is whether the errors are significant enough to affect model outcomes. As has already been
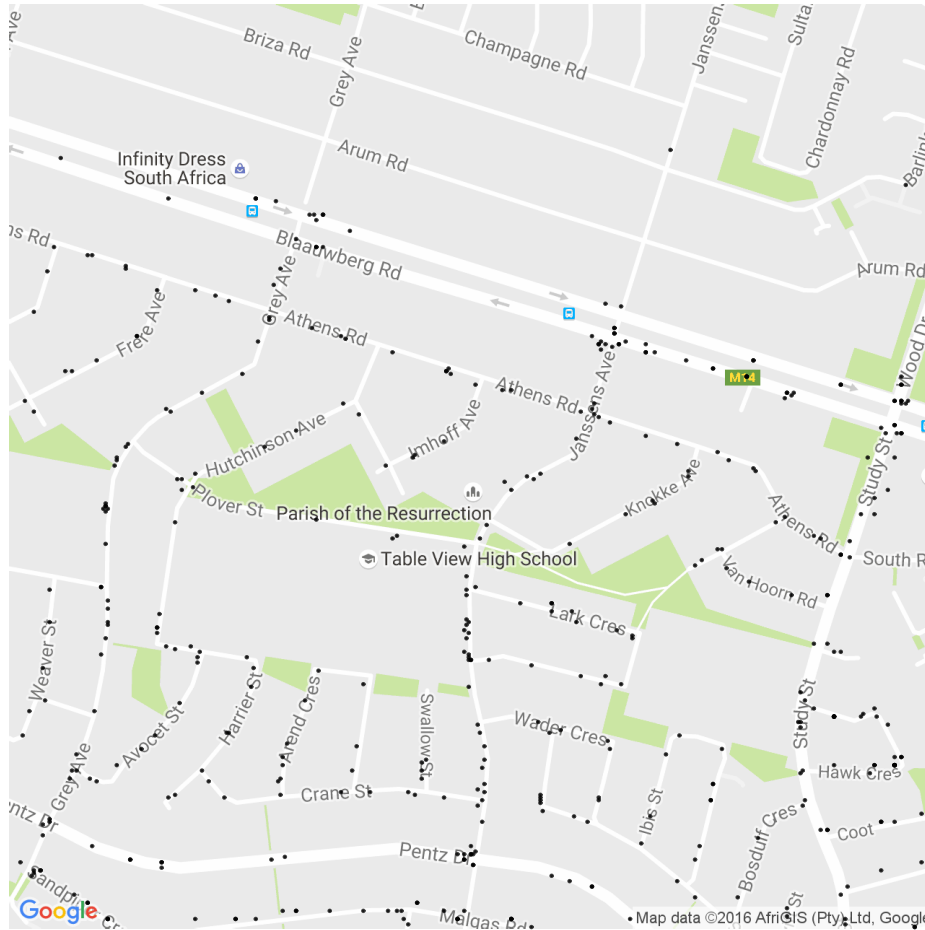
Figure 3.4: Map matching within Flamingo Vlei service area

proven the GPS devices used by city are accurate enough when plotted over a map. However when calculating distance there are always going to be errors. Currently around 5 percent of model velocities result in errors. While this is higher than expected it remains acceptable. All errors have been separated in the results that follow.

3.3.3 Random vehicle selection

In essence the function that the model performs is fairly basic but tedious. When a single vehicle's movement are selected from a single day it is easy to identify when the vehicle is servicing, dropping waste off at a landfill or travelling by intuitively looking at the vehicles traces. If the model can duplicate this logic to an acceptable degree of accuracy the model is considered validated. Figure 3.3 and table 3.2 is a good example of this. One can quickly identify what looks like a service areas and landfills. The model is able to do this as well quite accurately. This same test was applied to many vehicles over multiple days and in all cases the model identified trip segments adequately. The model is therefore considered validated for the current data set.

Chapter 4

Experiment design

As explained earlier in chapter 3 the conceptual model was developed on GPS traces from the entire waste collection fleet over a single day. For the purposes of model development this was a sufficient data set, however it is too small to produce conclusive results. For this reason a thorough experimentation plan has been developed to incorporate a more diverse set of data in order to make meaningful conclusions.

4.1 Service area selection

The first important decision in terms of sampling would be to select which service areas to consider in running the model. It is important to use a significant number of service areas without compromising processing speed. For this purpose the following service areas have been selected as samples. Dunoon, Vrygrond, Durbanville Hills, Flamingo Vlei, Hout Bay, Kirstenhof, Montague Gardens, Vishoek and Vygeboom. The reasoning in selecting these service areas is to make the sample of service areas as diverse as possible. Reflected here are low, middle and high income residential areas over various parts of the city. This allows the model to identify potential characteristics associated with different service area types. The ultimate goal for future work would be to use all service areas in the city, this data would be critically important to city planners and managers.

4.2 Sample study days

The next decision in experimental design will be how to select as few samples as possible from the GPS data in order to paint an accurate picture of waste collection behaviour in cape town. In addressing this the decision was made to consider fourty different Thursdays over the period of a year. This means the sample study period is all the Thursday from the 2 January, 2014 to 2 October, 2014. All the service areas mentioned above are designated to be serviced on a Thursday according to the city's beat maps. In doing this the characteristics of individual service areas will be determinable over an extended period of time. For future study it would be interesting to see the results if an entire sample of 365 days is considered.

4.3 Experimental method

Given the two data sets described above, along with the facility data that has already been incorporated, the experiments will all be done in accordance with the flow chart in figure 4.1. The procedure is as follows:

- The entire data set is read into R, this is approximately 8 million records. From this data set the sample days are extracted. The "ExtractDays.R" script performs this function.

- The next step is in identifying on which days, which vehicles pass through the sample service areas. "ExtractVehicles.R" does this by isolating all unique date, vehicle combinations that pass through the sample service areas.

- Following this the data is passed through "AssignStatus.R". This script performs the crucial function of identifying service areas, facility visits and the specific locations where service or landfill drop-off's take place. The results from this model component is the first data that is off use and is meaningful. This will be discussed in the results chapter.

- At this point the data was run through the "AssignVelocity.R" script where the instantaneous velocities for each point are calculated.

- The "Trip.R" script then isolates individual vehicles trips, it is from this script that the majority of the service analysis is done.

- Finally the "WasteGeneration.R" script builds on the trips identified and allocates and apportions waste to landfills and service areas. This represents the final model subcomponent.
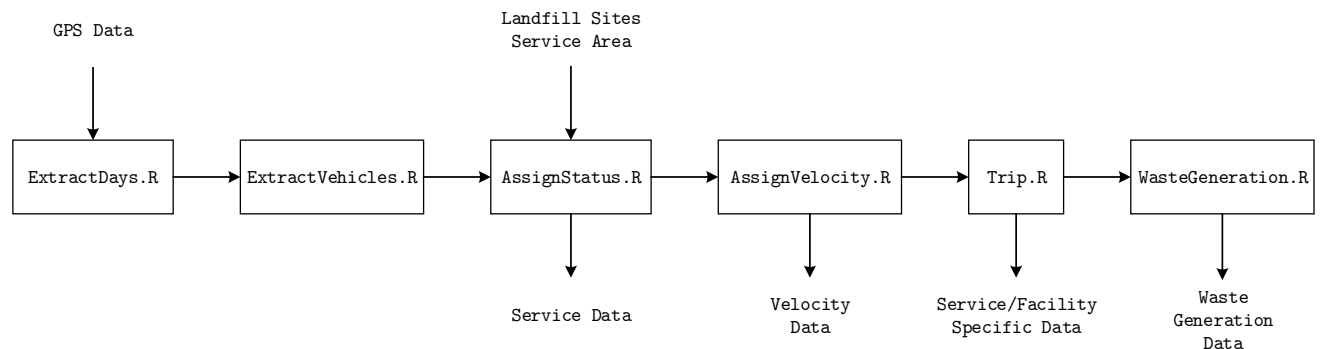

Figure 4.1: Model and data subcomponents

Chapter 5

Model results and observations

Given that the model has been explained in detail and that adequate proof has been provided of model validation and accuracy it is now appropriate to share preliminary results and observations as explained by the model.

5.1 Waste generation

Despite the fact that weighbridge data was unavailable the waste generation portion of the project produced promising results. Figure 5.1 shows the total waste collected per week over the entire sample period from three selected service areas. Table 5.1 in turn shows the mean waste generated per service area using landfill data estimates. This data immediately makes a distinction between service areas that produce more and less waste. This piece of information is novel in the sense that from this model waste generation rates can be produced for each service area over an entire city. Given real landfill data it is as simple as updating the model to produce real results. The great advantage that the model has over source sampling for example is its reproducibility. The model can be duplicated for any metropolitan area. In addition the model can be extended to forecast generation rates using the most recent GPS data. Figure 5.2 shows the total waste collected over the nine service areas per day, while the figure is not conclusive, it shows that seasonality is likely to be present in waste generation rates. Further study into this could possibly yield very interesting results.

Figure 5.1: Vehicle velocities over sample period

5.2 Service analysis

In terms of service analysis results regarding velocity, service times and travel times are shared and the significance of these results are explained below.

Table 5.1: Waste generation rates per service area

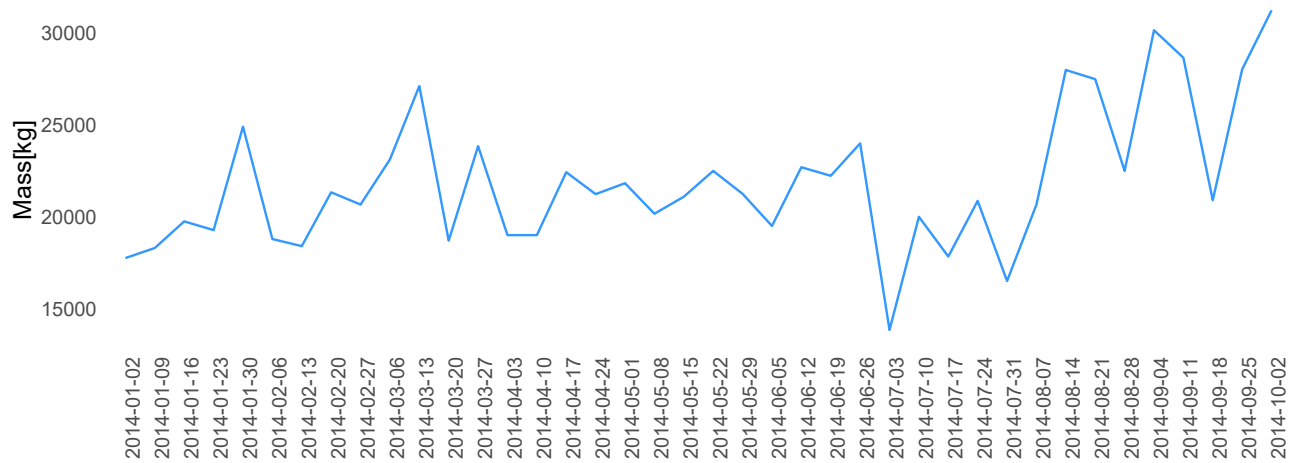| Service area	Total waste generated [kg]	Average waste generated per week[kg]
Dunoon	235390.78	5884.76
DurbanvilleHills	8111.34	202.78
FlamingoVlei	125632.20	3140.80
Houtbay	18350.00	458.75
MontagueGardens	120262.73	3006.52
Vishoek	45856.15	1146.40
Vrygrond	91090.91	2277.27
Vygeboom	1388.66	34.71

Figure 5.2: Total waste collected from sample service areas per day

5.2.1 Velocities outside service areas

The vehicle velocity forms one of the most critical pieces of information from a management perspective. In table 5.2 below a summary of vehicle velocity data is given. This is better illustrated in figure 5.4 that gives an indication of how velocities vary over the city for the entire 40 day data set. The pink dots most likely relate to service areas as the expected velocities in these areas will be very low. In addition highways clearly show up as blue as vehicles will have increased travelling speeds.

Table 5.2: Summary vehicle velocity statistics for entire data set

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0.00	0.00	4.10	77.52	32.99	48700

When considering only travel speeds, in other words travel speeds outside areas being serviced, the summary statistics are a little different. One must remember however that the model does not have access to all service areas, only nine strategic service areas have been identified for consideration by the model. It is therefore expected that the mean velocity for vehicles outside service
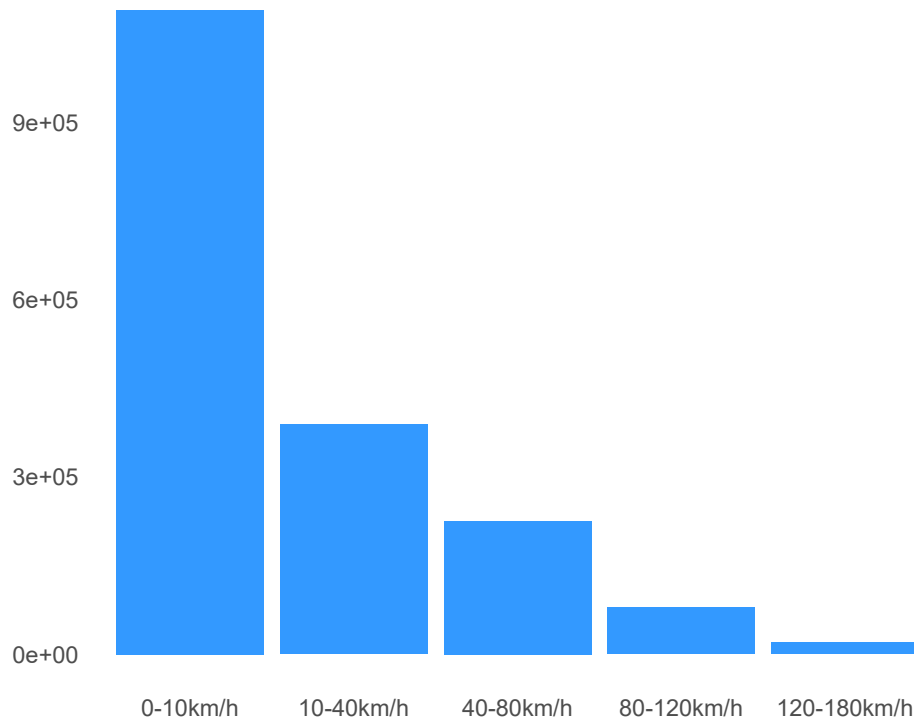
24

Figure 5.3: Vehicle velocities over sample period

areas will be considerably higher than the results show should all service areas be identified. This is nonetheless a very significant result as it proves the hypotheses that vehicle velocity can be used as an indicator of whether a vehicle is servicing an area. This is important as geofences alone do not have the required accuracy, vehicles often travel through service areas but do not collect waste(dead head). The combination of geofences and velocity is therefore an improved service indicator. In addition having an expected service velocity(in this case 0-10km/h) will help develop expected service times. Since service time is a function of velocity and the size of the population being serviced knowing the service velocity allows city planners to develop benchmark service times for collection crews. Benchmarks will play a considerable role in promoting operational efficiency for waste collection crews. This then directly fits in with the key performance indicators proposed by [6].

Table 5.3: Summary statistics for vehicles outside service areas after error removal

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0.000	0.000	3.585	19.280	29.720	179.600

5.2.2 Velocities within service areas

The data regarding service areas is considerably more accurate despite the fact that only nine service areas have been considered. The results show that a mean velocity of 10km/h is found
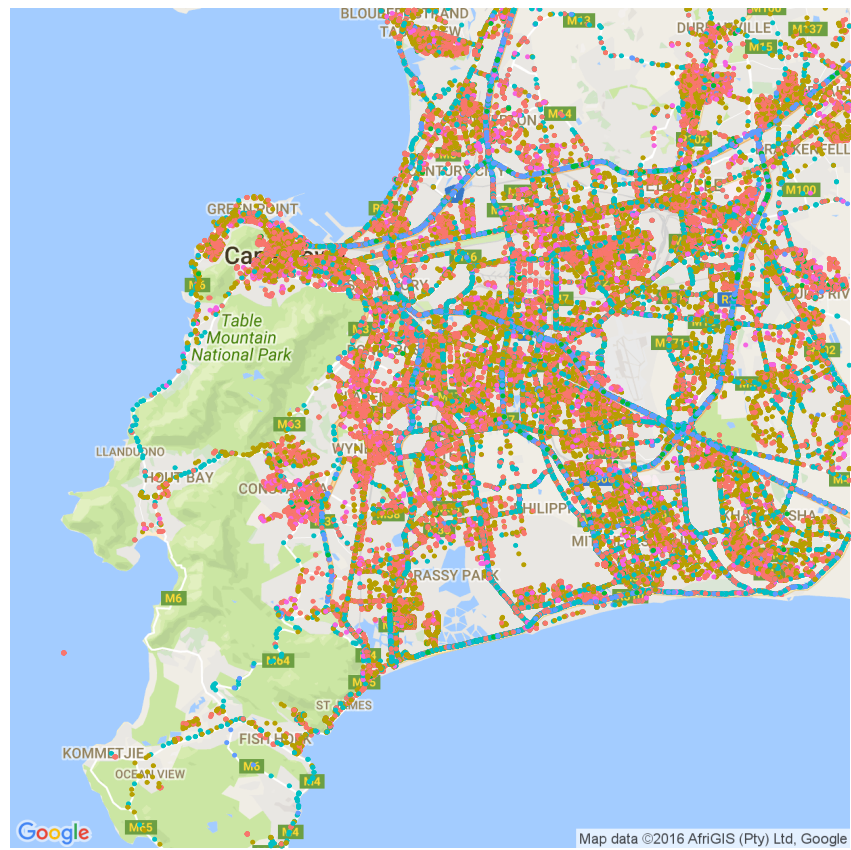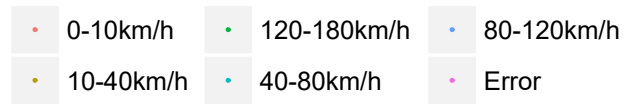
Figure 5.4: Vehicle velocities over entire sample period

26

within service areas. This is consistent with the expectation that vehicles move at reduced speeds when collecting waste. Figure 5.7 shows that all service areas share this property. The velocity category with the highest count for all service areas studied is the 0-10km/h bracket.

Another important observation can be made when service area velocity is displayed as a histogram. The expectation initially was that two distinct velocity distributions will be found within service areas, one representing vehicles collecting waste and the other representing vehicles dead-heading (driving through areas already serviced looking for waste to collect). Figure 5.5 shows that this hypothesis has been proved. By separating these two velocity populations a distinction can be made between servicing and dead heading. This could potentially improve model accuracy, in particular when street level waste generation modelling is attempted.

Table 5.4: Summary statistics for vehicles within service areas after error removal

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0.0000	0.0000	0.8822	10.1500	11.1900	179.2000
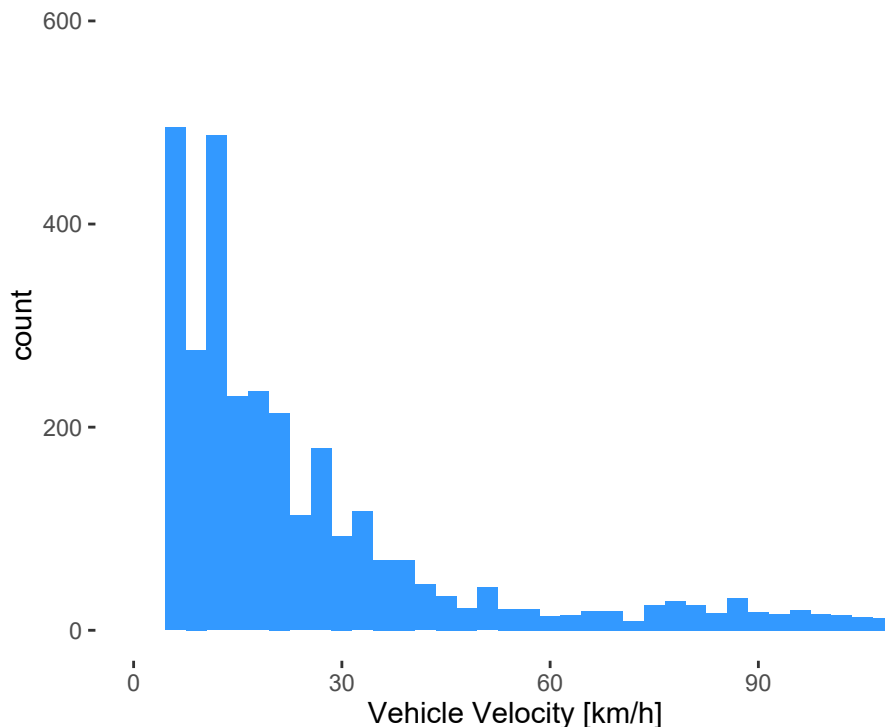
Figure 5.5: Vehicle velocities within service areas

It is also possible to investigate the velocities per service area. Factors like the population size and road conditions are likely to influence velocities within service areas. Table 5.5 shows that Vrygrond for example has a much lower service velocity than the other service areas. This is explained by the fact that Vrygrond is largely an informal settlement and therefore road quality and service time is considerably longer than would be expected for middle-class service areas. This

27
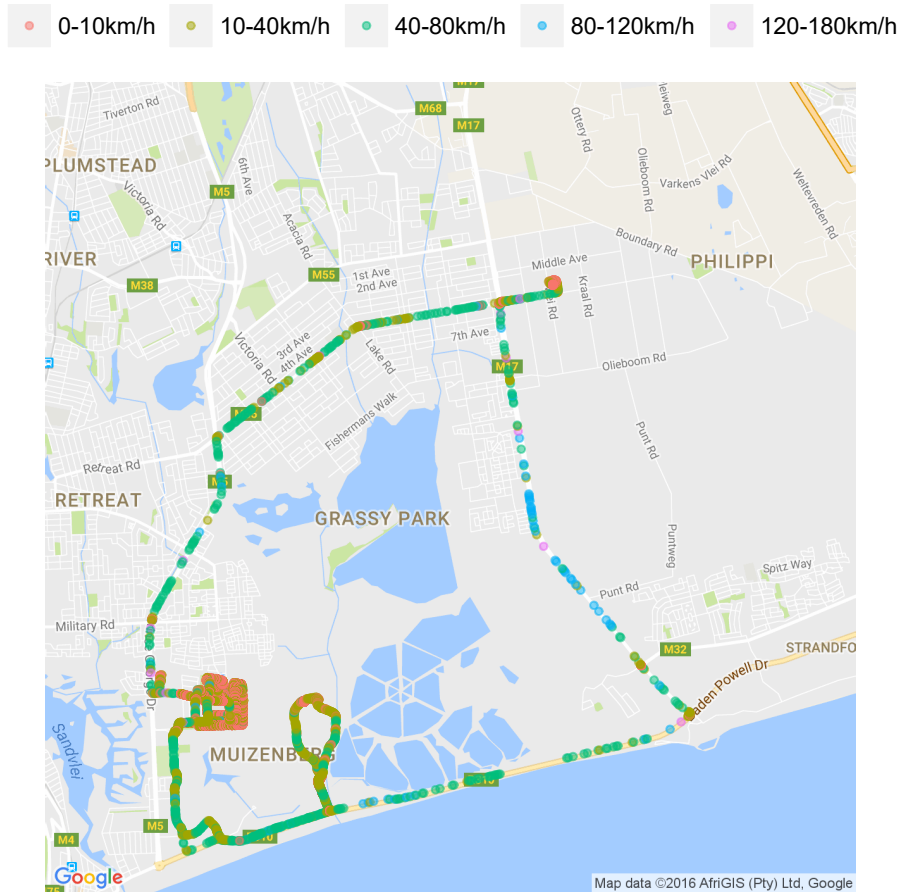
Figure 5.6: Vehicle A velocity

also explains why Vrygrond has a much higher count in figure 5.7 as the number of GPS points within the Vrygrond geofence is a lot more compared to other service areas. This information is of huge significance to a city manager in the allocation of waste collection resources to service areas. It should also be noted that the mean vehicle velocity within service areas has a small range, this serves to validate the accuracy of the model.

5.2.3   Service duration per service area

Results regarding service duration are less conclusive. Table 5.6 shows that service durations are slightly lower than expected. A possible explanation for this is that the model regards every time a vehicle enters or exits a service area as a new "trip". For this reason one can conclude that vehicles regularly cross the service area geofences as they move from street to street collecting waste. This is supported by the fact that the number of trips that are classified as "Dunoon to Dunoon" for example is on average 40 per day. Vehicles clearly enter and exit the service area frequently as they collect waste. An improvement to the model in this regard would possibly entail merging trips that lead from a service area back to a service area as this should be regarded as part of a single service

28

Dunoon	DurbanvilleHills	FlamingoVlei	Kirstenhof	MontagueGardens	Vishoek	Vrygrond	Vygeboom
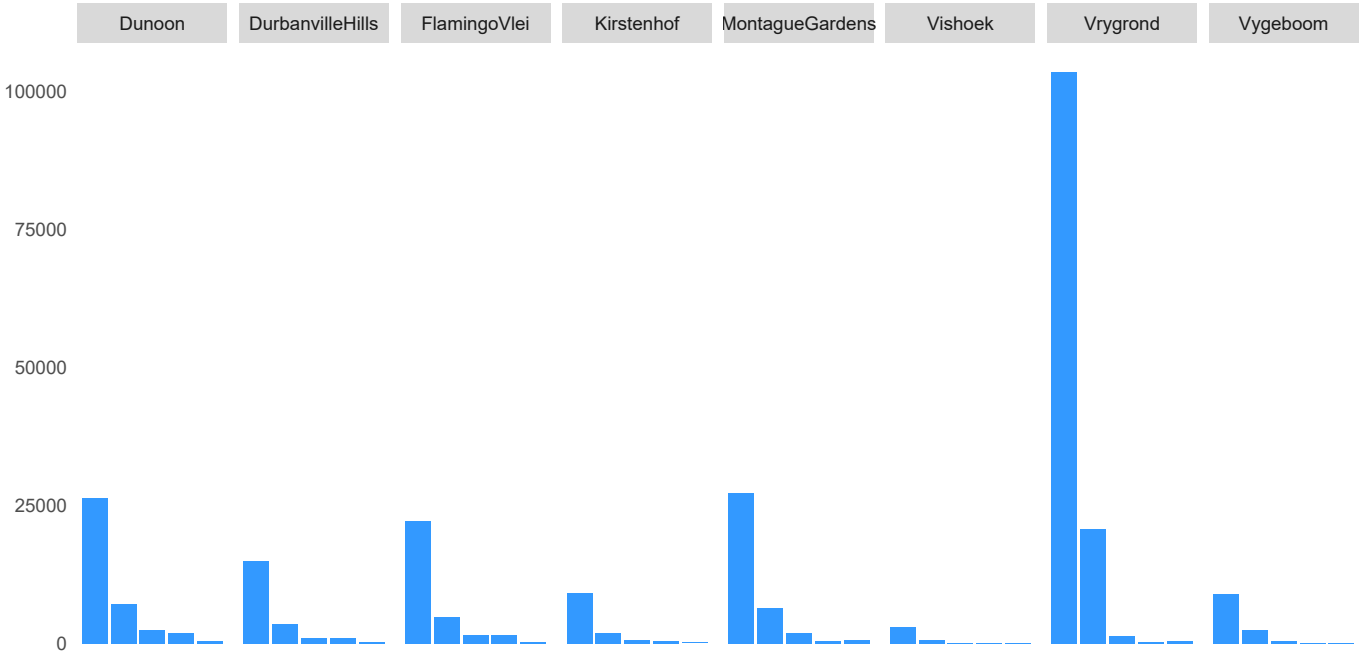
Figure 5.7: Count of points that fall within velocity categories per service area

Table 5.5: Mean vehicle velocity within service areas

Service Area	Mean Velocity [min]
Dunoon	13.86
DurbanvilleHills	10.56
FlamingoVlei	10.56
Kirstenhof	16.97
MontagueGardens	16.56
Vrygrond	5.46
Vygeboom	8.94
Zonnebloem	8.98

run.

5.2.4 Duration of facility visits

Facility visit duration produced credible results. Figure 5.8 shows how landfill waiting times are quite stratified. This is concerning as it makes vehicles routing and planning difficult, ideally landfill delays should be short and consistent across all landfills. Average duration of landfill visit is a crucial piece of information to a waste collection system. The length of time a vehicle spends at a landfill determines the number of areas it can service during a day and therefore ultimately the fleet size. In knowing what the expected delay will be at landfills city planners can improve scheduling, improve traffic flow at problematic landfills and ultimately improve operational efficiency. The data

Table 5.6: Mean vehicle velocities within service areas

Service Area	Average Service Duration[min]
Dunoon	19.87
DurbanvilleHills	31.15
FlamingoVlei	40.20
Houtbay	18.34
Kirstenhof	12.02
MontagueGardens	29.71
Vishoek	12.63
Vrygrond	32.40
Vygeboom	23.19

reported on for visit times at landfills are credible due to the large data set (23 landfills over 40 days). This is therefore a positive outcome for the project.
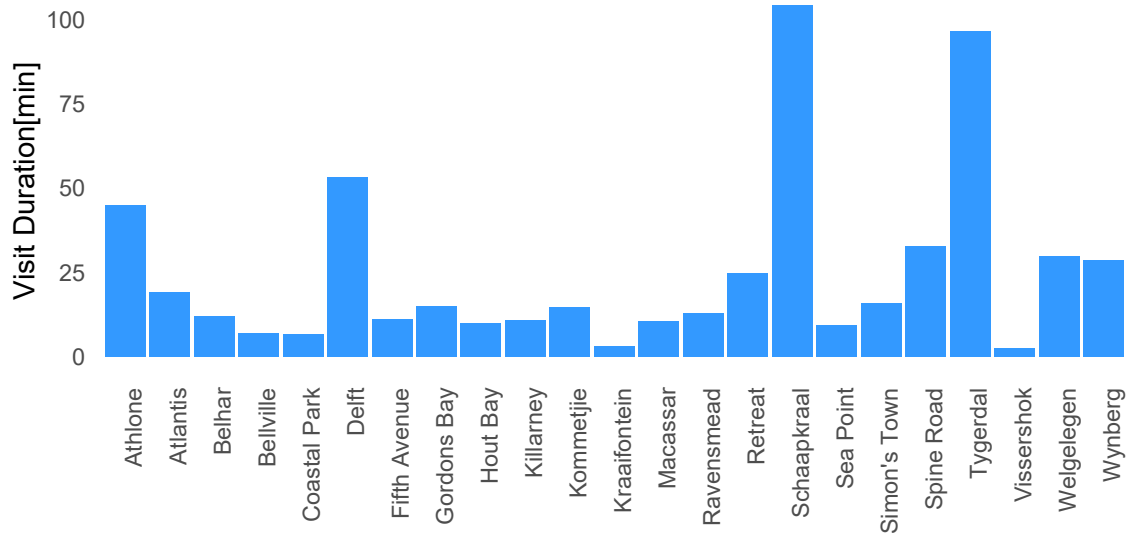
Figure 5.8: Mean visit duration per landfill site

5.2.5 Travel times

Finally travel times are addressed. This is also a very useful decision support variable as the utilisation of any vehicle is dependant on the amount of time the vehicle travels. The vehicle only truly delivers value when it is collecting and offloading waste, minimising the time in between is therefore crucial. Having an accurate idea of travel times such as those presented in the appendix is crucial in planning vehicle trips.

5.3   Results Analysis

The results above prove that using GPS data as source of operational intelligence is not only feasible for waste collection systems but should be encouraged. Incorporating credible weighbridge data into the model developed in this study is likely to produce accurate waste generation data that can be updated and reproduced across any municipality or metropolitan area. Furthermore the results achieved in analysing velocity conclusively prove that velocity is a reliable method for identifying service periods on a vehicle's route. The results also show that average service velocities can be determined per service area. This will form the foundation for developing key performance indicators that will drive efficiency. When considering service durations the results were less accurate, however the results helped in identifying a potential model shortfall. The duration of facility visits showed that there is a great disparity between delays at different landfills. The potential for optimisation is therefore significant.

Chapter 6

Conclusion

It is evident that there is a great need for accurate and reliable waste generation data as well as data relating to the operational efficiency of collection vehicles. With increased urbanisation great strain is placed on metropolitan areas to provide basic services, providing these services effectively and at minimal cost however requires access to accurate operational and strategic information.

While there are existing sources of information such as census data, these are often outdated and too broad to be of real use. Given the fact that the city already has GPS devices fitted on their collection vehicles and already use weigh bridges the potential for waste generation modelling and service analysis using GPS data is massive. This paper proves not only that there is a very real need for accurate data relating to the performance of waste collection systems but also that there is very little literature on using GPS traces to this end. The methods used in this paper are not new but are tried and tested methods of handling GPS data, the novelty being in having used these methods for waste generation modelling. This paper has proved that it is possible to:

- describe waste generation and vehicle service patterns with a reasonable degree of accuracy;

- make use of recent GPS data to develop waste generation and service analysis models that are reproducible and easy to construct and maintain.

Future work could build on the fundamentals already established to improve model accuracy, extend the scope of the model and extend the input data. Ultimately the possibility of real-time modelling on data soon after the data is generated could be realised. However the results already produced from this project also have great significance. The city now has improved methods for generating waste generation and service analysis models.

In conclusion it must be noted that a fiscally constrained country such as South Africa can always benefit from cost savings. Service delivery is a shortcoming that has still not adequately been addressed and therefore any service delivery system can benefit from optimisation. The results and methods developed and adapted in this paper will form the basis of optimisation models in future, and will go a long way in cutting waste collection costs. The application of GPS data to waste generation modelling and service analysis is therefore not only feasible but encouraged to provide efficient and effective waste collection services.

Bibliography

[1] Peter Beigl, Sandra Lebersorger, and Stefan Salhofer. Modelling municipal solid waste generation: A review. *Waste management*, 28(1):200–214, 2008.

[2] Glenn Cich, Luk Knapen, Tom Bellemans, Davy Janssens, and Geert Wets. Trip/stop detection in gps traces to feed prompted recall survey. *Procedia Computer Science*, 52:262–269, 2015.

[3] Mohan B Dangi, Michael A Urynowicz, Kenneth G Gerow, and Resham B Thapa. Use of stratified cluster sampling for efficient estimation of solid waste generation at household level. *Waste Management & Research*, 26(6):493–499, 2008.

[4] Brian Dyson and Ni-Bin Chang. Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste management*, 25(7):669–679, 2005.

[5] Gianpaolo Ghiani, Demetrio Laganà, Emanuele Manni, Roberto Musmanno, and D Vigo. Operations research in solid waste management: A survey of strategic and tactical issues. *Computers & Operations Research*, 44:22–32, 2014.

[6] You-Ti Huang, Tze-Chin Pan, and Jehng-Jung Kao. Performance assessment for municipal solid waste collection in taiwan. *Journal of environmental management*, 92(4):1277–1283, 2011.

[7] A Jiménez-Meza, J Arámburo-Lizárraga, and E de la Fuente. Framework for estimating travel time, distance, speed, and street segment level of service (los), based on gps data. *Procedia Technology*, 7:61–70, 2013.

[8] Forrester JW. Industrial dynamics. *The MIT Press*.

[9] Chen Liu and Xin-wu Wu. Factors influencing municipal solid waste generation in china: a multiple statistical analysis study. *Waste Management & Research*, 2010.

[10] Zhu Minghua, Fan Xiumin, Alberto Rovetta, He Qichang, Federico Vicentini, Liu Bingkai, Alessandro Giusti, and Liu Yi. Municipal solid waste management in pudong new area, china. *Waste management*, 29(3):1227–1233, 2009.

[11] City of Cape Town. Budget 2016. Technical report, Council of the City of Cape Town, 2016.

[12] Republic of South Africa. Constitution of the republic of south africa, act 108 of 1996, 1996.

[13] Parliament of the Republic of South Africa. Public finance management act 1 of 1999, April 2010.

[14] Ana Pires, Graça Martinho, and Ni-Bin Chang. Solid waste management in european countries: A review of systems analysis techniques. *Journal of environmental management*, 92(4):1033–1050, 2011.

[15] HA Abu Qdais, MF Hamoda, and J Newham. Analysis of residential solid waste at generation sites. *Waste Management & Research*, 15(4):395–405, 1997.

[16] Nicky Rogge and Simon De Jaeger. Evaluating the efficiency of municipalities in collecting and processing municipal solid waste: A shared input dea-model. *Waste management*, 32(10):1968–1978, 2012.

[17] Bruce Gordon Wilson and Julie K Vincent. Estimating waste transfer station delays using gps. *Waste management*, 28(10):1742–1750, 2008.

Appendix A

Tables

Table A.1: Average landfill waiting time

Landfill site	Average duration at landfill [min]
Athlone	44.98
Atlantis	19.17
Belhar	12.07
Bellville	6.84
Coastal Park	6.66
Delft	53.24
Fifth Avenue	10.99
Gordons Bay	15.09
Hout Bay	9.92
Killarney	10.74
Kommetjie	14.70
Kraaifontein	3.00
Macassar	10.45
Ravensmead	12.94
Retreat	24.72
Schaapkraal	104.24
Sea Point	9.22
Simon's Town	15.82
Spine Road	32.60
Tygerdal	96.30
Vissershok	2.66
Wynberg	28.68

Figure A.1: Travel times between landfills and service areas

| Average of Duration[min] | Column Labels | | | | | | |
Row Labels	Dunoon	DurbanvilleHills	FlamingoVlei	MontagueGardens	Vishoek	Vrygrond	Vygeboom
Athlone	158.5383333	101.1166667	141.6266667	128.0082766	257.0333333	161.2625	
Atlantis	79.47866667		37	241.2833333			
Belhar	80.73333333	335		71.13055556			
Bellville	126.29	84.01666667		86.23839286			34.96666667
Coastal Park	127.3926667			139.8966667	44.62179487	26.19036697	
Delft	61.73125	142.725	178.5	85.84047619		260.0333333	
Fifth Avenue (Parow)	221.2333333			186.4416667			
Killarney	5.066666667		6.320238095	32.43026316			
Kommetjie					54.5875		72.125
Kraaifontein		134.05		196.7833333			
Macassar							
Retreat					76.16388889	47.475	
Schaapkraal	82	62	166	91.59	71.36277778	61.82176471	
Sea Point	64.75			73.78866667		167.9333333	
Simon's Town					54.36888889		
Spine Road	111.9958333		393	135.9	189.0166667	29.6	
Tygerdal	92.98571429			49.15			
Vissershok	218.0333333						
Welgelegen	158.1071429	48.31666667		141.4833333			
Wynberg				95.29583333	221.2583333	95.03666667	

Appendix B

Source code

ExtractDays.R

Llewellyn

Tue Sep 27 14:26:55 2016

```r
Master <- read.csv("thursday_waste.csv", header = TRUE, sep = ",")
days <- read.csv("Days.txt", header = TRUE)
DayExtract <- data.frame()
NewMaster <- data.frame()
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
for(x in 1:length(days$x)){
  DayExtract <- filter(Master,Master$DATE == toString(days$x[x]))
  NewMaster <- rbind(NewMaster, DayExtract)
}

NewMaster2 <- data.frame(NewMaster$VEH_REG_NO, NewMaster$DATE,
                         NewMaster$POS_TIMESTAMP,
                         NewMaster$POS_X, NewMaster$POS_Y)
colnames(NewMaster2) <- c("Reg","Date","Time","Long","Lat")
write.table(NewMaster2, file = "2014.txt", row.names = FALSE, append = FALSE)
```

1

ExtractVehicles.R

Llewellyn

Sun Sep 25 17:20:22 2016

```r
Master <- read.csv("2014.txt", header = TRUE, sep = " ")

library(sp)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
s1 <- read.csv("Dunoon.txt", FALSE,";")
s2 <- read.csv("DurbanvilleHills.txt", FALSE,";")
s3 <- read.csv("FlamingoVlei.txt", FALSE,";")
s4 <- read.csv("Kirstenhof.txt", FALSE,";")
s5 <- read.csv("MontagueGardens.txt", FALSE,";")
s6 <- read.csv("Vrygrond.txt", FALSE,";")
s7 <- read.csv("Vygeboom.txt", FALSE,";")
s8 <- read.csv("HoutBay.txt", FALSE,";")
s9 <- read.csv("Vishoek.txt", FALSE,";")

Service <- list(s1,s2,s3,s4,s5,s6,s7,s8,s9)
ServiceNames <- c("Dunoon","DurbanvilleHills","FlamingoVlei","Kirstenhof","MontagueGardens","Vrygrond",

position <- vector()
count <- vector()
NewMaster <- data.frame()
New <- data.frame()

for(x in 1:length(Service)){
  position<- vector()
  position <- point.in.polygon(Master$Long,Master$Lat,unlist(Service[[x]][1]), unlist(Service[[x]][2]))

  for(i in 1:length(position)){
    if(position[i] ==1){
      count <- append(count, i, after = length(count))
    }
  }
}
TruckDay <- data.frame(Master$Reg[count], Master$Date[count])
TruckDay <- unique(TruckDay)
```

1

AssignStatus.R

Llewellyn

Tue Sep 27 23:03:34 2016

```r
#This Script aims to combine all model components into a single descriptive model
library(sp)

Master <- data.frame()
status <- vector()
LocationName <- vector()

gps <-read.csv("2014InServiceAreas.txt",TRUE," ")

s1 <- read.csv("Dunoon.txt", FALSE,";")
s2 <- read.csv("DurbanvilleHills.txt", FALSE,";")
s3 <- read.csv("FlamingoVlei.txt", FALSE,";")
s4 <- read.csv("Kirstenhof.txt", FALSE,";")
s5 <- read.csv("MontagueGardens.txt", FALSE,";")
s6 <- read.csv("Vrygrond.txt", FALSE,";")
s7 <- read.csv("Vygeboom.txt", FALSE,";")
s8 <- read.csv("HoutBay.txt", FALSE,";")
s9 <- read.csv("Vishoek.txt", FALSE,";")

Service <- list(s1,s2,s3,s4,s5,s6,s7,s8,s9)
ServiceNames <- c("Dunoon","DurbanvilleHills","FlamingoVlei","Kirstenhof",
                  "MontagueGardens","Vrygrond","Vygeboom",
                  "Houtbay", "Vishoek")
FacLoc <- read.csv("Facilities.csv", TRUE,";")
#
#
#
#returns whether point is near facility and thererfore a facility visit
CheckVisit <- function(Long,Lat, FacLoc, Fac_Radm){

  cur = c(Lat,Long)
  visit = 0

  ##Distance from point to each facility
  dis = 0
  for(i in 1:24){
    A = (FacLoc$Latitude[i] - cur[1])*69.1
    B = 69.1*(FacLoc$Longitude[i] - cur[2])*cos(cur[1]/57.3)
    dis = (((A^2)+(B^2))^0.5)*1609.344
    ##Whether point classifies as facility visit
    if(is.na(dis)){visit = 0}
    else if(dis<Fac_Radm){
      visit = i
      break()
    }
  }
  return(visit)
```

1

```
}
#
#
#

#Finds all points within service areas
points <- vector()
count1 <- vector()
count2 <- vector()
for(r in 1:length(Service)){
  count2 <- vector()
  points <- point.in.polygon(gps$Long, gps$Lat, unlist(Service[[r]][1]),
                             unlist(Service[[r]][2]))
  for(q in 1:length(points)){if(points[q]==1){count1 <- append(count1, q,
                                                 after= length(count1))
  count2 <- append(count2, q, after = length(count2))}}
  status[count1] <- c("Servicing")
  LocationName[count2] <- toString(ServiceNames[r])
  }
#Master is the master data frame that will contain all relevant data on a point

status[length(status):length(gps$Reg)]<- NA

for (x in 1:length(gps$Reg)){

  CheckVis <- CheckVisit(gps$Long[x],gps$Lat[x],FacLoc, 200)
  if(CheckVis>0){
    status[x] <- c("VisitingFacility")
    LocationName[x] <- as.character(FacLoc$FacilityName[CheckVis])
  }
  if(is.na(status[x]) == TRUE){status[x] <- c("Travel")
    LocationName[x] <- c("Travel")}
  else if(CheckVis == 0 && status[x]!="Servicing"){
    status[x] <- c("Travel")
    LocationName[x] <- c("Travel")
  }
}

Master <- data.frame(gps$Reg,gps$Date,gps$Time,gps$Long,gps$Lat, status,
                     LocationName)
colnames(Master) <- c("Reg","Date","Time","Long","Lat", "Status",
                     "LocationName")
write.table(Master, file = "Master.txt", row.names = FALSE, append = FALSE)
```

AssignVelocity.R

Llewellyn

Mon Sep 26 20:58:41 2016

```r
#Functions
#
#
#Returns Time between two points (text format) in seconds (numeric)
time <- function(time1, time2){
 hrs1 = as.numeric(substr(time1,1,2))
 min1 = as.numeric(substr(time1,4,5))
 sec1 = as.numeric(substr(time1,7,8))

 hrs2 = as.numeric(substr(time2,1,2))
 min2 = as.numeric(substr(time2,4,5))
 sec2 = as.numeric(substr(time2,7,8))
 dif = ((hrs2-hrs1)*3600)+((min2-min1)*60)+(sec2-sec1)
}
#Returns Distance between two points in metres
distance <- function(Lat1, Long1,Lat2, Long2){
A = (Lat2 - Lat1)*69.1
B = 69.1*( Long2- Long1)*cos(Lat1/57.3)
dis = sqrt((A^2)+(B^2))*1609.344
return(dis)
}
#
#

library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
Master <-read.csv("Master.txt",TRUE," ")
NewMaster1 <- data.frame()
NewMaster2 <- data.frame()
TruckList <- read.csv("TruckDay.txt", TRUE, sep = " ")
position <- vector()

 Truck <- data.frame()
 Velocity <- vector()
 TruckVelocity <- vector()
```

```r
VelocityStatus <- vector()
Status <- vector()
count = 0

for(i in 1:length(TruckList$Reg)){
  count = 1
  position <- vector()
  Truck <- data.frame()
  NewMaster1 <- data.frame()
  Velocity <- vector()
  VelocityStatus <- vector()

  Truck<- filter(Master, Master$Reg==toString(TruckList$Reg[i]),
              Master$Date == toString(TruckList$Date[i]))
  if(length(Truck$Reg)>1){
      for(p in 2:length(Truck$Reg)){
        Velocity[p] <- c(distance(Truck$Lat[p-1],
                              Truck$Long[p-1],Truck$Lat[p],
                              Truck$Long[p])/time(Truck$Time[p-1],Truck$Time[p])*3.6)
        if(Velocity[p]>180||is.na(Velocity[p])==TRUE){VelocityStatus[p] <- c("Error")}
        else if(Velocity[p]<10){VelocityStatus[p] <- c("0-10km/h")}
        else if(Velocity[p]>10&&Velocity[p]<40){VelocityStatus[p] <- c("10-40km/h")}
        else if(Velocity[p]>40&&Velocity[p]<80){VelocityStatus[p] <- c("40-80km/h")}
        else if(Velocity[p]>80&&Velocity[p]<120){VelocityStatus[p] <- c("80-120km/h")}
        else if(Velocity[p]>120&&Velocity[p]<180){VelocityStatus[p] <- c("120-180km/h")}
      }
  }
  NewMaster1 <- data.frame(Truck$Reg, Truck$Date, Truck$Time, Truck$Long,
                            Truck$Lat, Truck$Status,
                            Truck$LocationName, Velocity, VelocityStatus)
  NewMaster2 <- rbind(NewMaster2, NewMaster1)
}
colnames(NewMaster2) <- c("Reg","Date","Time","Long","Lat", "Status", "LocationName",
                            "Velocity", "VelocityStatus")
#write.table(NewMaster2, file = "MasterVelocity.txt", row.names = FALSE, append = FALSE)
```

Trip.R

Llewellyn

Tue Sep 27 08:28:57 2016

```r
Master <- read.csv("MasterSplit2Sorted.csv", header = TRUE, sep = ";")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```r
#
#
#Functions
#Returns Time between two points (text format) in seconds (numeric)
time <- function(time1, time2){
 hrs1 = as.numeric(substr(time1,1,2))
 min1 = as.numeric(substr(time1,4,5))
 sec1 = as.numeric(substr(time1,7,8))
 hrs2 = as.numeric(substr(time2,1,2))
 min2 = as.numeric(substr(time2,4,5))
 sec2 = as.numeric(substr(time2,7,8))
 dif = ((hrs2-hrs1)*3600)+((min2-min1)*60)+(sec2-sec1)
 return(dif/60)
}

SingleTime <- function(time1){
  hrs1 = as.numeric(substr(time1,1,2))
  min1 = as.numeric(substr(time1,4,5))
  sec1 = as.numeric(substr(time1,7,8))
  return((hrs1*3600)+(min1*60)+sec1)
}
#
```

```r
#
#
#

TruckList <- read.csv("TruckDay.txt", TRUE, sep = " ")

Trip <- data.frame()

Truck <- data.frame()
Reg <- vector()
Date <- vector()
Type <- vector()
From <- vector()
To <- vector()
StartTime <- vector()
EndTime <- vector()
Duration <- vector()

for(y in 1:length(TruckList$Reg)){
  Truck <- data.frame()
  Truck <- filter(Master, Master$Reg == toString(TruckList$Reg[y]),
                  Master$Date == toString(TruckList$Date[y]))

  if(length(Truck$Reg)>2){
    StartTime <- append(StartTime, as.character(Truck$Time[1]),
                        after = length(StartTime))
    From <- append(From, as.character((Truck$LocationName[1])),
                   after = length(From))
    for(x in 1:(length(Truck$Reg)-1)){

      if(Truck$LocationName[x+1] != Truck$LocationName[x]){
        if(x!=(length(Truck$Reg)-1)){

          EndTime <- append(EndTime, as.character(Truck$Time[x+1]),
                            after = length(EndTime))
          Reg <- append(Reg, as.character(Truck$Reg[x]),
                        after = length(Reg))
          Date <- append(Date, as.character(Truck$Date[x],
                                            after = length(Date)))
          Type <- append(Type, as.character(Truck$Status[x]),
                         after = length(Type))
          StartTime <- append(StartTime, as.character(Truck$Time[x+1]),
                              after = length(StartTime))

          #Travel to Service/Facility Changeover
          if(Truck$Status[x]== "Travel"){

            To <-append(To, as.character(Truck$LocationName[x+1]),
                        after = length(To))
            From <-append(From, as.character(Truck$LocationName[x+1]),
                          after = length(From))
          }
```

2

```r
        #Service to Service Changeover
        else if(Truck$Status[x+1]== "Servicing"&&
              Truck$Status[x]== "Servicing"){
          To <-append(To, as.character(Truck$LocationName[x]),
                    after = length(To))
          From <-append(From, as.character(Truck$LocationName[x+1]),
                      after = length(From))
        }

        #Service/Facility to Travel Changeover
        else if(Truck$Status[x]== "Servicing"||
              Truck$Status[x]== "VisitingFacility"){
          To <- append(To, as.character(Truck$LocationName[x]),
                    after = length(To))
          From <- append(From, as.character(Truck$LocationName[x]),
                      after = length(From))
        }

        }
        else{To <- append(To, as.character(Truck$LocationName[x]),
                      after = length(To))
        EndTime <- append(EndTime, as.character(Truck$Time[x]),
                      after = length(EndTime))
        Reg <- append(Reg, as.character(Truck$Reg[x]),
                    after = length(Reg))
        Date <- append(Date, as.character(Truck$Date[x],
                                        after = length(Date)))
        Type <- append(Type, as.character(Truck$Status[x]),
                    after = length(Type))}
      }
    }
    if(Truck$LocationName[length(Truck$Reg)]==Truck$LocationName[length(Truck$Reg)-1]){
      EndTime <- append(EndTime, as.character(Truck$Time[length(Truck$Reg)]),
                    after = length(EndTime))
      Reg <- append(Reg, as.character(Truck$Reg[length(Truck$Reg)]),
                  after = length(Reg))
      Date <- append(Date, as.character(Truck$Date[length(Truck$Reg)],
                                      after = length(Date)))
      Type <- append(Type, as.character(Truck$Status[length(Truck$Reg)]),
                    after = length(Type))
      To <-append(To, as.character(Truck$LocationName[length(Truck$Reg)]),
                  after = length(To))
    }
  }
}
Duration <- vector()
for(z in 1:length(Trip$Reg)){
  Duration <- append(Duration, time(toString(Trip$StartTime[z]),
                                  toString(Trip$EndTime[z])),
                  after = length(Duration))
}

Trip <- data.frame(Reg,Date, Type, From,To,StartTime, EndTime, Duration)
```

3

```
write.table(Trip, file = "Trip.txt", row.names = FALSE, append = TRUE)
```

```
## Warning in write.table(Trip, file = "Trip.txt", row.names = FALSE, append =
## TRUE): appending column names to file
```

WasteGeneration.R

Llewellyn

Tue Sep 27 09:28:43 2016

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
Trip <- read.csv("WasteGeneration.csv", header = TRUE, sep = ";")
TruckList <- read.csv("TruckDay.txt", header = TRUE, sep = " ")
NewTrip <- data.frame()
NewTruck <- data.frame()
WasteSer <- vector()
WasteFill <- vector()
EstimatedAmount = 1000
index <- vector()
Duration = 0

for(x in 1:length(TruckList$Reg)){
  Truck <- data.frame()
  Truck <- filter(Trip, Trip$Reg == toString(TruckList$Reg[x]),
                  Trip$Date == toString(TruckList$Date[x]), Trip$Type != "Travel")
  index <- vector()
  WasteFill <- vector()
  WasteSer <- vector()

  if(length(Truck$Reg)>1){

  if(Truck$Type[length(Truck$Type)] == "VisitingFacility"){
    WasteFill[length(Truck$Type)] <- c(EstimatedAmount)
    WasteSer[length(Truck$Type)] <- NA
    index <- append(index, length(Truck$Type), after = length(index))
  }
  else{WasteFill[length(Truck$Type)] <- NA
  WasteSer[length(Truck$Type)] <- NA}

  for(z in (length(Truck$Reg)-1):1){
    #Allocates estimated waste to facility visits
    if(toString(Truck$Type[z]) == "VisitingFacility"&&
       toString(Truck$Type[z+1]) != "VisitingFacility"){
      WasteFill[z] <- c(EstimatedAmount)
```

1

```r
      WasteSer[z] <- NA
      index <- append(index, z, after = length(index))
    }
    else{WasteFill[z] <- NA}
  }
    Duration = 0

    #Runs through service segments
    if(length(index)>0){
      if(length(index)>1){
      for(i in 1:(length(index)-1)){

        #Adds total service duration before waste drop off
        for(s in index[i]:index[i+1]){
          if(Truck$Type[s] == "Servicing"){Duration = Duration +
            as.numeric(Truck$Duration[s])}
        }
        for(s in index[i]:index[i+1]){
          if(Truck$Type[s] == "Servicing")
            {WasteSer[s]<- -as.numeric(Truck$Duration[s])/Duration*WasteFill[i]
          WasteFill[s] <- NA }
        }
        if(Duration == 0){WasteFill[index[i]] <- NA}
      }
    }
    Duration =0

    for(s in index[length(index)]:1){
      if(Truck$Type[s] == "Servicing"){Duration = Duration + as.integer(Truck$Duration[s])}
    }
    for(s in index[length(index)]:1){
      if(Truck$Type[s] == "Servicing")
        {WasteSer[s] <- c(-as.integer(Truck$Duration[s])/Duration*as.integer(WasteFill[i]))
      WasteFill[s] <- NA }
    }
      if(Duration == 0){WasteFill[index[length(index)]] <- NA}
    }
  }
  else if(length(Truck$Reg) == 0){WasteFill <- vector()
  WasteSer <- vector()}
  else if(length(Truck$Reg) == 1){WasteFill[1] <- NA
  WasteSer[1] <- NA
  }

  NewTruck <- data.frame(Truck$Reg,Truck$Date, Truck$Type,
                         Truck$From,Truck$To, Truck$StartTime,
                         Truck$EndTime, Truck$Duration, WasteFill,
                         WasteSer)
  NewTrip <- rbind(NewTruck, NewTrip)
}

colnames(NewTrip) <- c("Reg", "Date", "Type", "From", "To", "StartTime",
                       "EndTime", "Duration", "WasteFill", "WasteSer")
```

```r
write.table(NewTrip, file = "WasteGeneration.txt", row.names = FALSE,
            append = FALSE)
```