

Optimization of solid waste collections in Blantyre, Malawi

Author: Nicolas Seemann-Ricard

Supervisor: Elizabeth Tilley

Tutor: Florian Dörfler

21 November 2022

Abstract

Your abstract should succinctly summarise the research gap, the methods you employed, your results, conclusions, and recommendations. Don’t use acronyms if possible and keep the language as general as possible. Keep the abstract to a maximum of 500 words. The abstract stays on its own page.

Table of Contents

Contents

[1 Introduction 5](#_Toc119480419)

[1.1 Background topic 1 (Solid Waste Management) 5](#_Toc119480420)

[1.2 Background topic 2 (SWM in Africa) 5](#_Toc119480421)

[1.3 Justification and Research Questions 5](#_Toc119480422)

[2 Data analysis 6](#_Toc119480423)

[2.1 Data 6](#_Toc119480424)

[2.1.1 Dump arrivals logs 7](#_Toc119480425)

[2.1.2 Skips filling data 7](#_Toc119480426)

[2.2 Methods 9](#_Toc119480427)

[2.2.1 Filling rates 9](#_Toc119480428)

[2.2.2 Dump logs 11](#_Toc119480429)

[2.3 Results 11](#_Toc119480430)

[2.3.1 Filling rates 11](#_Toc119480431)

[2.3.2 Dump logs 14](#_Toc119480432)

[2.3.3 Extrapolating filling rates 16](#_Toc119480433)

[3 Optimization model 21](#_Toc119480434)

[3.1 Methods 21](#_Toc119480435)

[3.2 Results 23](#_Toc119480436)

[4 Conclusions 23](#_Toc119480437)

[4.1 Capital expenditure 23](#_Toc119480438)

[4.2 Operational planning 23](#_Toc119480439)

[5 Recommendations for future work 23](#_Toc119480440)

[5.1 Skip-level data 23](#_Toc119480441)

[5.2 Current organizational and operational state 23](#_Toc119480442)

[6 References 24](#_Toc119480443)

# Introduction

The project seeks to minimize the costs of operation of the municipal solid waste management service in Blantyre Malawi, while limiting overflowing of the communal skips.

## Background topic 1 (Solid Waste Management)

## Background topic 2 (SWM in Africa)

Maybe now you dive into the global differences in collection

## Justification and Research Questions

1. How can we model with limited data?
2. How many trucks are needed to service all the skips? What would be the mileage and cost of servicing all skips without overflow?
3. What would be the optimal routing schedule?

# Data analysis

## Data

In order to formulate feasible and pertinent recommendations, parameters reflecting the situation need to be calculated.

The data analysis is based on three datasets. The first is a timeseries of specific skips’ levels over a certain period of time. However, the scope of it is quite narrow, with only 12 useful filling rates extracted, all in a small area of Blantyre. A second dataset gives the arrivals at Mzedi dump, the main inorganic waste landfill in Blantyre. Though it covers all the skips studied, it presents strong limitations. Namely, arrivals at the dump do not reflect the speed at which the skips fill up. Indeed, the individual filling data show that some skips go a long time without being emptied, overflowing and presenting a public health risk.

Each skip is 7m3. All of the 53 skips are assumed to be mixed waste (organic and inorganic) except for the ones explicitly said to be organic and inorganic. The skips, the municipal dump (Mzedi dump), the truck storage facility and the compost facility are mapped in Figure 1.

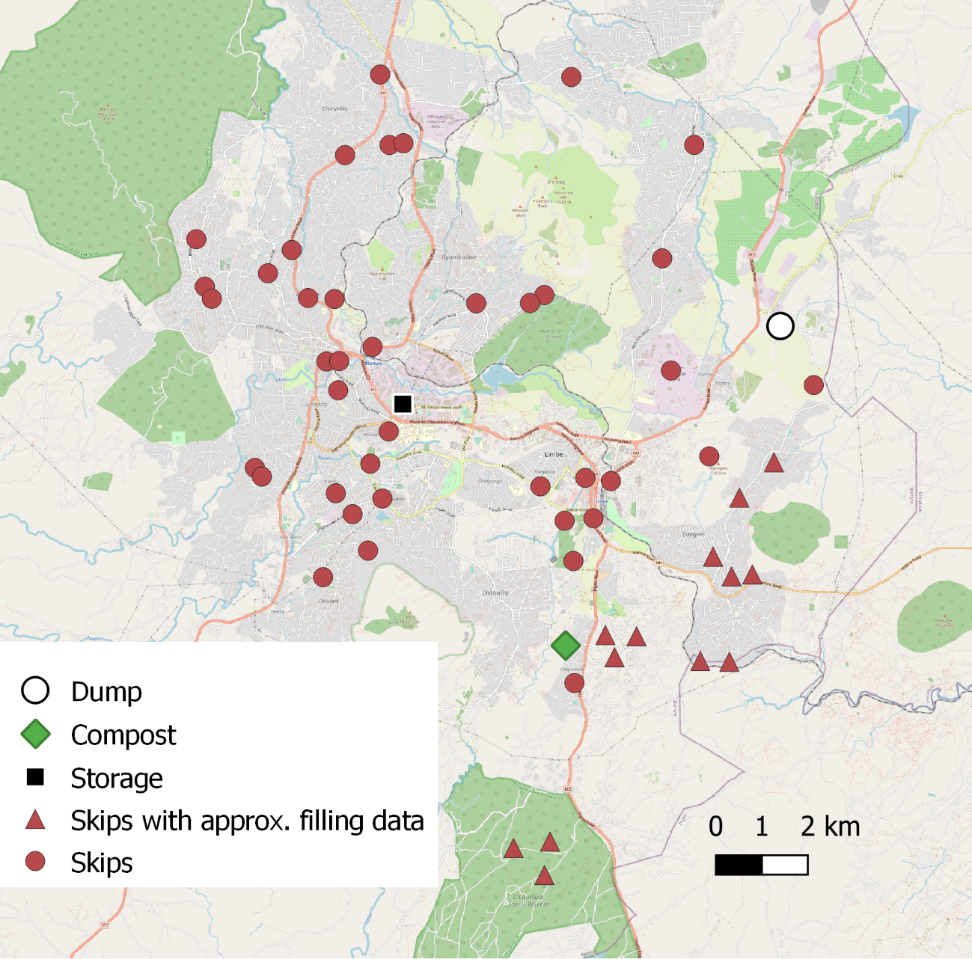


Figure 1 Locations of skips, dump, compost facility and truck storage in Blantyre

### Dump arrivals logs

This is a log lists arrivals at the Mzedi dump, along with the origin of the skip carried by each truck. The origins match exactly the geographical locations in Figure 1. The period of this series is 2020-12-05 to 2021-12-31.

### Skips filling data

Filling data for several skips is provided. Which skips from Figure 1 they exactly refer to is ambiguous, but the skips within the area which could match are annotated as “Skips with approx.. filling data”. The areas are *Bangwe*, *BCA*, *Chigumula* and *Naizi*.

Over a certain period (depending on the skip), a measurement on a scale from 1-5 was taken visually (generally) every day at those skips. A score between 0 and 4 indicate the estimated fullness of the skip, while a 5 means the skip was overflowing. Three of the 14 provided sets are shown inFigure 2**-**4.

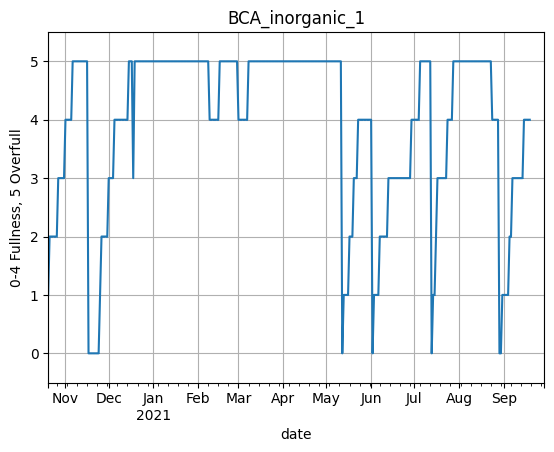


Figure 2 Skip filling data for “BCA inorganic 1”



Figure 3 Skip filling data for “Chigumula\_inorganic\_2”



Figure 4 Skip filling data for “Bangwe\_inorganic\_1”

*“BCA Inorganic 1”* gives a good example of a profile from which rising trends (ramps) are discernible and allow to easily calculate a filling rate. This is useful in estimating the frequency at which the skip needs to be serviced. *“Chigumula Inorganic 2”*, however, simply does not have enough data points and has to be dismissed. Finally, the data from *“Bangwe inorganic 1”* indicates that it fills up extremely fast, sometimes within a day.

## Methods

### Filling rates

Filling rates are extracted from the skip-level data. The goal is to identify a rising trend over several days, and to compute the filling rate as the slope of this ramp. It performs the following steps:

1. Removing spikes

Spikes occur in the dataset frequently. These are points where the level quickly drops or increases before going back to its original level. Those are assumed to be measurement errors. Figure 5 shows the action of the filter for “*Chigumula\_inorganic\_1*”. The limitations of this method is that quick filling might be dismissed as a spike, effectively putting a lower bound of 0.5 skip/day filling rate.

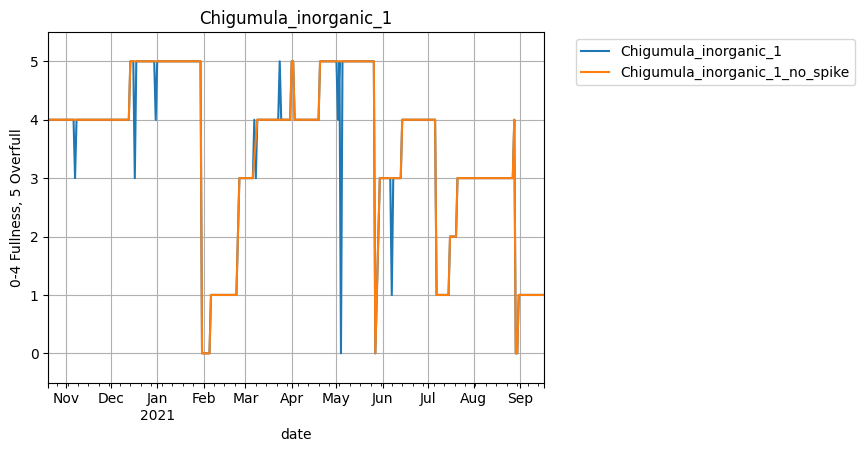


Figure 5 Action of filtering spikes

1. Detecting the top of ramps

The top of the ramp is the end point of a rising ramp fitted to the time series. These may occur at any level. They are determined by a forward and a backward pass through the series. The forward pass identifies top “ends”, when the next value of fullness is smaller than the current one (which characterises the skip being emptied). A top end is also identified when the current value is 4 and the next is a 5 (or overfull). The reasoning behind this is the preference for ramps to finish with a 4. This level is defined as full, which is more precise than the “overfull” denomination. Still, in the absence of an intermediate 4 (e.g. a direct transition between 3 and 5), the top value will be the 5.

A backward pass detects the first date at which the top value appears. These are considered to be the time where the skip achieves the fullness level at which it is emptied, or when it reaches full capacity. Finding this date is particularly important when the skip stays overfull for long periods of time. Using the top ends to calculate the filling rate would bias it to a lower rate.

1. Eliminating downgrading

Downrating is clear in Figure 6 (a zoomed in version of Figure 2), on two occasions between February and April. These events might occur for several reasons, such as:

* Trash being burned to eliminate the overflowing waste.
* Waste being cleared by another party.
* An error in data collection or the ambiguity in the measurement scale.

In the backward pass described in 2), the top values at the end of derating periods are removed by adding a condition that deletes the current top end if the level rises.

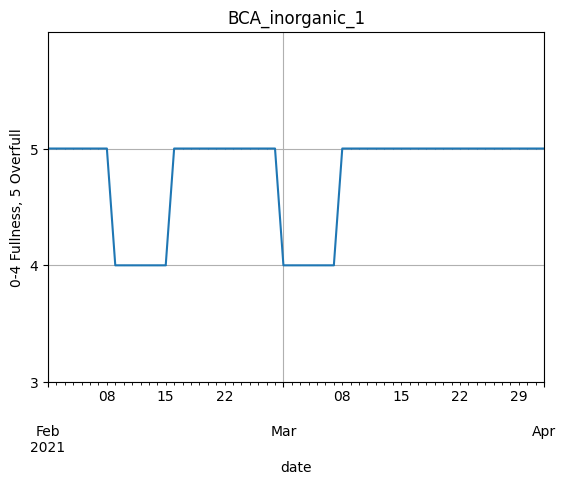


Figure 6 Zoomed in "BCA\_inorganic\_1" filling data

1. Detecting the bottom of ramps

Finally, the bottom values and dates are extracted iterating backward from each top beginning, until the previous value is larger than the current one, at which point it is assumed the skip has just been emptied.

### Dump logs

As previously mentioned, dump logs are useful in characterizing the current operation

## Results

### Filling rates

The result is shown in Figure 7 for BCA inorganic 1 and in Figure 8 for Bangwe inorganic 1. Clearly, the fit in the second case is less ideal. In this case, it is more accurate to not to remove spikes, leading to the ramps shown in Figure 9.



Figure 7 Results of ramps for “BCA inorganic 1”



Figure 8 Results of ramps for “Bangwe\_inorganic\_1” **with** spikes removed

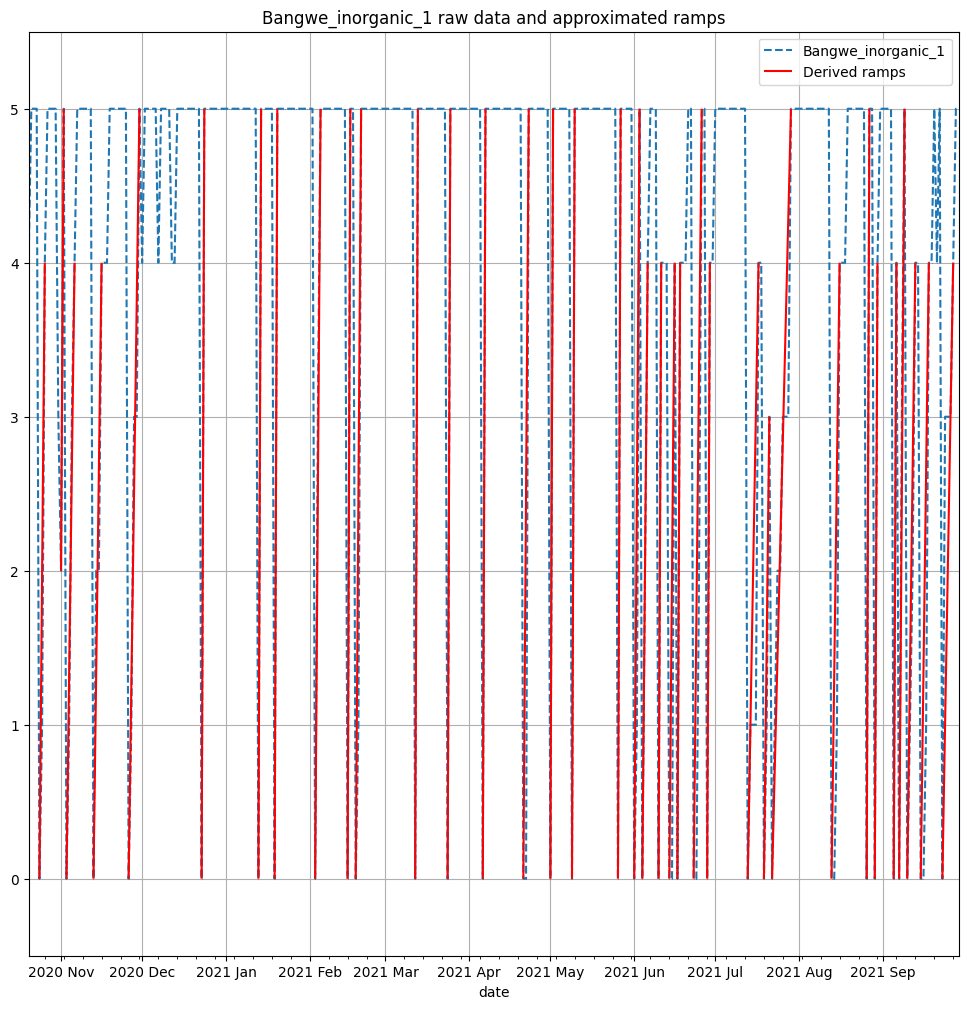


Figure 9 Results of ramps for “Bangwe\_inorganic\_1” **without** spikes removed

The average slopes of the ramps are calculated for each skip in the dataset for the case where spikes are removed and where they are not. This corresponds to the minimum and maximum average filling rates respectively in Table 1. For skips such as “*Bangwe\_inorganic\_1”*, there is a large difference, as spikes and quick filling of the skip are confounded. “*Chigumula\_inorganic\_1”* also has a large difference between the minimum and maximum filling rates. However, in this case and as shown in Figure 9, removing the spikes produces accurate ramps. The process is repeated for each skip based on the visual fit. In Table 1, the filling rates determined to be more accurate is highlighted.

Table 1 Aggregate data from ramps analysis. Highlighted filling rates are selected to be the most accurate based on visual fit of the ramps. # indicates the number of ramps fitted. Prop. overfull is the proportion of days where the level is 5. The period is the period over which levels were measured

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Skip | Avg filling rate min | Avg filling rate max | # min | # max | Prop. overfull | Period (days) |
| Bangwe\_Organic\_1 | 0.215 | 0.215 | 24 | 26 | 0.258 | 341 |
| Bangwe\_Organic\_2 | - | - | - | - | - | - |
| Bangwe\_inorganic\_1 | 0.49 | 0.766 | 23 | 36 | 0.725 | 345 |
| Bangwe\_inorganic\_2 | 0.4 | 0.541 | 16 | 20 | 0.737 | 265 |
| BCA\_Organic\_1 | 0.091 | 0.105 | 11 | 13 | 0.386 | 324 |
| BCA\_Organic\_2 | 0.077 | 0.077 | 2 | 2 | 0 | 54 |
| BCA\_inorganic\_1 | 0.063 | 0.126 | 6 | 7 | 0.536 | 335 |
| BCA\_inorganic\_2 | 0.038 | 0.047 | 1 | 2 | 0.627 | 178 |
| Naizi\_Organic\_1 | 0.046 | 0.046 | 8 | 8 | 0.21 | 324 |
| Naizi\_Organic\_2 | - | - | - | - | - | - |
| Naizi\_inorganic\_1 | 0.144 | 0.161 | 15 | 17 | 0.293 | 342 |
| Naizi\_inorganic\_2 | - | - | - | - | - | - |
| Chigumula\_Organic\_1 | 0.158 | 0.212 | 21 | 24 | 0.183 | 327 |
| Chigumula\_Organic\_2 | 0.127 | 0.142 | 13 | 19 | 0.106 | 223 |
| Chigumula\_inorganic\_1 | 0.056 | 0.308 | 4 | 9 | 0.256 | 333 |
| Chigumula\_inorganic\_2 | - | - | - | - | - | - |

### Dump logs

The sum of arrivals in each week is shown in Figure 11**.** A sizeable gap is noticeable between January 20th and February 26th . This is reflected in the skips filling data, where many skips were overflowing and not emptied during this period. The reason for this gap is unknown but assumed here to be the service simply not operating. Intra-weekly, the arrivals are relatively homogeneous, except for Sunday, as illustrated in Figure 12. This pattern does not change significantly before and after the service gap.

Of particular interest is the number of days between arrivals at the dump for each skip. The analysis of these gaps is shown in Figure 13. A noticeable characteristic is the variability of the time gaps, as illustrated by the size of the boxes and whiskers. For areas that are serviced often, such as Blantyre Flea Market, Limbe, Ndirande and Chirimba (where the median gap is at most 3), the outliers are frequent, indicating gaps where the skips are likely overflowing.

Importantly, the arrivals at the dump provide useful insights into the current operation of the municipal solid waste management system in Blantyre. It does not, however, allow by itself to infer the filling rate of bins. As seen in Table 1, at least some skips spend a considerable amount of time overfull. As such, the time between collections is dependent on other factors as well.



Figure 11 Sum of deliveries (arrivals) at Mzedi dump over the entire period of measurements



Figure 12 Sum of deliveries to Mzedi dump per weekday over the entire period of measurements

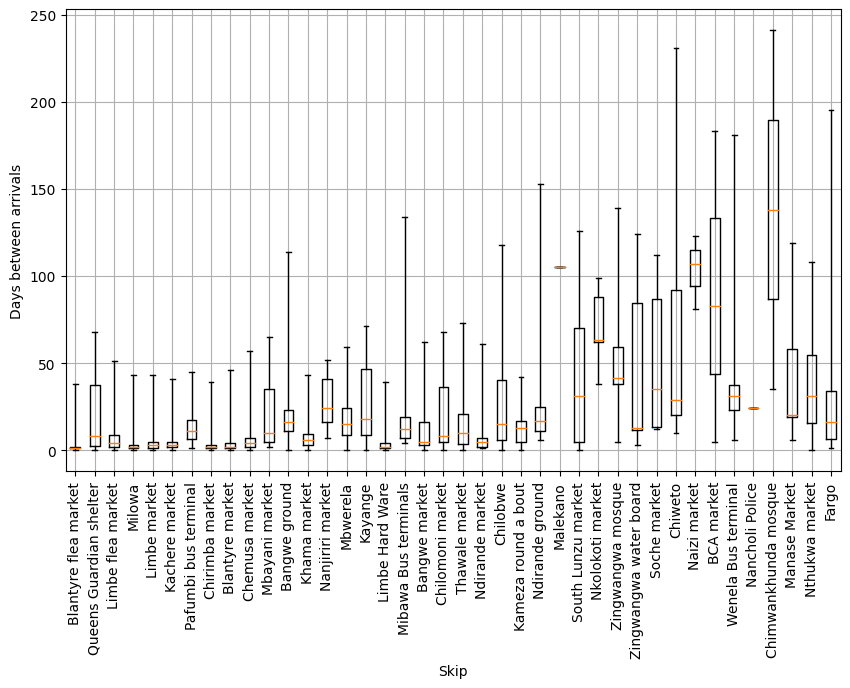


Figure 13 Box plots of number of days between arrivals for each skip. The boxes represent the range between the quantiles Q1 (0.25) and Q3 (0.75). The whiskers extend to the minimum and maximum gaps.

### Extrapolating filling rates

As previously mentioned, the first dataset, though useful in estimating the filling rates, is limited to twelve skips. Those skips are also concentrated in one area of Blantyre, as seen in Figure 1. This makes it difficult to extrapolate to other areas in the city. It is attempted to use characteristics from the two datasets to get filling rates estimates for each skip. Since the skips in the filling dataset and the areas from the dump logs do not match, they are aggregated as shown in Table 2. Chigumula, though in the original list of skips, does not appear in the dump arrival logs. This despite the fact that skip filling data indicates the inorganic skip was emptied 4 times in the period. The organic skips are not considered since they should be processed at the composting facility.

Table 2 Aggregation of skips and skip areas and number of data points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Aggregate area | Skips from filling data | # ramps | Area from dump logs | # arrivals |
| Bangwe | Bangwe\_inorganic\_1  Bangwe\_inorganic\_2 | 23  16 | Bangwe ground  Bangwe market | 14  31 |
| BCA | BCA\_inorganic\_1  BCA\_inorganic\_2 | 6  1 | BCA market | 4 |
| Naizi | Naizi\_inorganic\_1 | 15 | Naizi market | 4 |
| Chigumula | Chigumula\_inorganic\_1 | 4 | none |  |

Since it is known that not all Bangwe ground and Bangwe market skips are included in the skips filling data, the emptying events (which are the “bottom” dates described in 2.2.1) should all fit within the arrival events described in the Mzedi dump logs. Figure 14 contradicts this notion for Bangwe, however. It shows many skip emptying events do not match with arrivals at the dump. Furthermore, there are noticeable clusters of emptying events and dump arrivals that do not match one another. In Figure 17, the proportion of emptying events matching arrival events is plotted, with a “padding”. An emptying event is said to match if it is within a period of time defined as the padding. For example, if a ramp event occurs on November 13th, and the padding is set to 2, the emptying event and arrival will match if there is an arrival on November 13th, November 12th or November 11th. As padding is increased, more events match, but with a median time between arrivals of just a few days, events get confounded. Furthermore, for Naizi and BCA, the maximum matching for reasonable padding is 14% and 0% respectively.

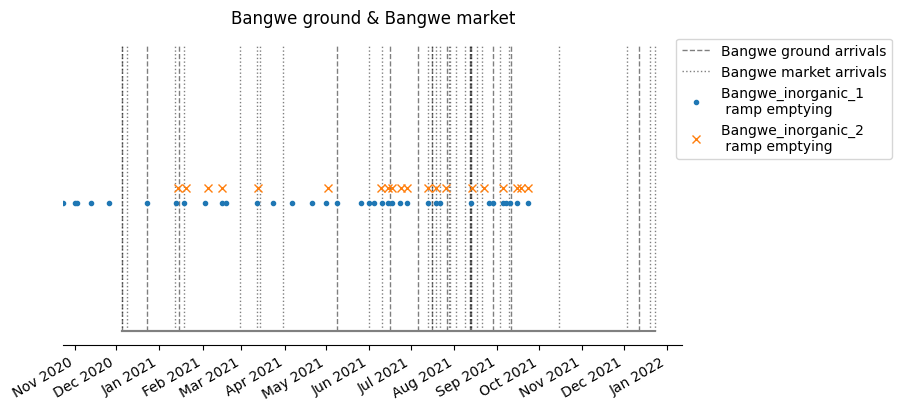


Figure 14 Timeline of Bangwe arrivals at Mzedi dump and emptying events of select skips

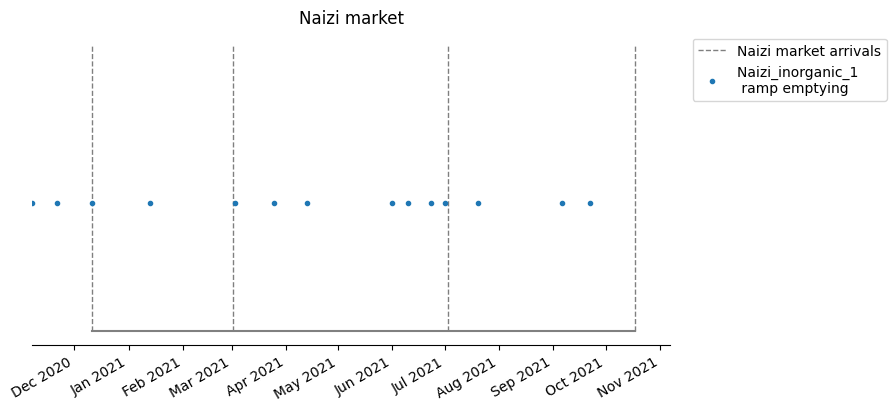
`

Figure 15 Timeline of Naizi market arrivals at Mzedi dump and emptying events of select skips

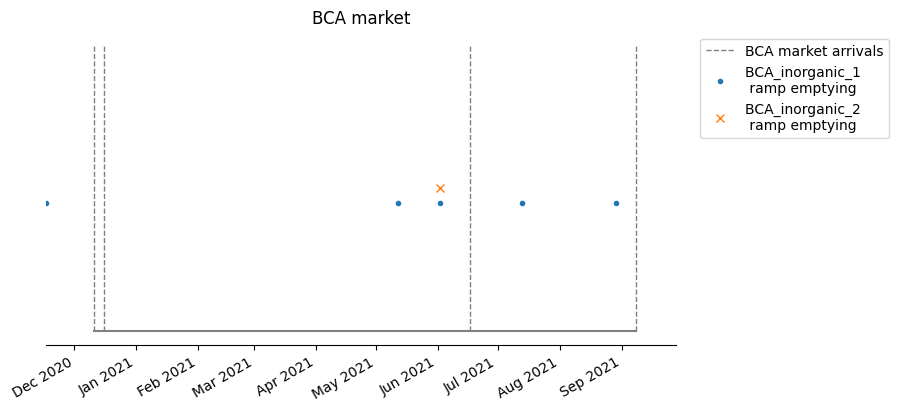


Figure 16 Timeline of BCA market arrivals at Mzedi dump and emptying events of select skips

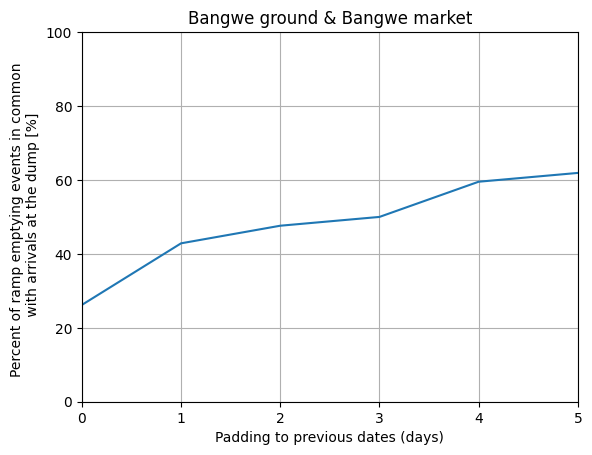


Figure 17 Percentage of skip emptying events matching arrivals at the dump for Bangwe area, with padding to allow for error

Possible reasons for this mismatch are:

* Inaccuracies in the ramps fitting method
* Measurement errors in the skip filling data
* Issues with data logging at Mzedi dump, such as not all arrivals being logged

## Data analysis insights and estimates

Several conclusions are drawn to build the optimization model:

Service:

* The service operates at reduced capacity on Sunday.
* There is high variability in collection periods for both skips that are serviced on a low and high frequency.
* Skips are left overfull for a considerable amount of time.

Waste generation:

* Filling rates range from around 0.04 up to at least 0.8 skip/day for inorganic waste.

To estimate:

* Filling rates
  + Measured filling rates for known skips
  + Filling rates is Q1 divided by the number of identically identified skips

Do:

* How is currently operated cost + estimated overflowing
* Estimated cost of operating without overflow
* Change to parameters such as
  + Adding a skip
  + Adding a vehicle
  + Separating waste (organic and inorganic)
  + Add transfer station
  + Sensitivity to labour cost, fuel cost

# Optimization model

The problem is formulated as a series of features, assumptions, objective and constraints.

The optimization is done over one week, but taking into account skips that fill at a rate where they are emptied at a period of more than a week. Up to two collections are possible per day.

## Methods

The optimization problem is an integer program with most variables constrained to binary. There are N stages.

### Scenario creation and selection

A number of computations are done before the optimization, which reduces the number of constraints and variables, which in turn reduces the computation time.

The variation of filling rates during the week is important in assigning schedules. Therefore, filling rate vectors have a length of 7\*2 = 14 for each skip. A number of possible skips schedules are generated. These are referred to as scenarios. For each period of the week, the scenario vector is assigned either a 1 or a 0. A 1 indicates the skip is serviced in that period. The scenarios are generated for collection intra-weekly (once a week up to 6 times a week) and extra-weekly (once every two, three and four weeks), and for each a combination of morning and afternoon collections. The difference between the smallest and longest gap between collections may not be larger than 1 day, which constrains the collection scenarios to be regular (i.e. for a 4 period scenario, [1 0 1 0 1] is possible while [1 1 1 0 0] is considered too irregular). The extra-weekly scenarios are assigned a cost divisor, which, in the objective function, is multiplied by the collection costs of a skip it is assigned to, in order to give the average cost of collection over several weeks.

Next, each skip, is assigned its feasible scenarios. The filling rates of each skip is combined with each scenario, and if a scenario does not result in overflowing, it is added to the possible scenarios for this particular skip. Additionally, scenarios which result in the skips being underutilized (i.e. emptied when it is below a certain threshold), are not considered to reduce the number of possible scenarios and therefore the size of the problem.

There are also scenarios where additional skips are added. These are considered for a particular skip if it has above a certain number of collections in a week or if no scenarios with no additional skips fit.

### Problem formulation

Variables

The top-level decision variables are scenario selection for each skip, number of vehicles operating on each period and the first skip to be serviced on each round. Consequent variables are the number of added skips (which results from the selected scenarios), the number of vehicles on any one day and the number of crews needed (the maximum number of vehicles used in the same period).

The variables are: scenario selection, period servicing, selection of first skip in period, number of vehicles, number of vehicles operating per day, period operation, number of additional skips.

Parameters

Parameters are set at the beginning of the problem.

|  |  |  |  |
| --- | --- | --- | --- |
| Stage | Variable | Default value | Unit |
| Pre-processing | Average speed | 30 | km/h |
| Days in optimization horizon | 7 | days |
| Periods per day | 2 | periods/day |
| Underfull threshold | 0.6 | n/a |
| Possible additional skips for each existing one | 2 | skip/skip |
| Number of weekly collections above which additional skips are considered | 4 | collections/week |
| Default skip filling rate where missing | 0.2 | skip/day |
| Wage structure (1 if constant wage per day, 0 if dependent on number of vehicles out per period) | 0 | n/a |
| Optimization problem parameters | Labour cost of operating a truck for one period (or labor costs if the wage structure is 1) | 4.5 | USD/day |
| Cost of 1km of travel | 0.5 | USD/km |
| Maximum number of additional skip | ------ | USD/skip |
| Maximum travel time per period | 4 | hours/period |
| Optimizer parameters | Duality gap (tradeoff between computation time and proximity to optimal solution) | 3 | % |

Objective function

The objective function adds all the costs of operation. It assumes optimal operation of the system. To calculate the distance travelled, it assumes empty skip roll-on-roll-off operationThe costs are: distance costs (xit and fit), labor costs (based on number of vehicles operated on each day), (capital costs).

Constraints

Number of hours per period per vehicle. Number of additional vehicles and skips.

### Problem solving

* Features:
  + 28 days planning period
  + Morning and evening collection
    - The two useful cases are:
      * Afternoon collection on Saturday and morning collection on Monday
      * Multiple collections per day
    - Heuristically assign sets of schedules to skips and add skips
      * Assume that a double collection is always done on the same day for both skips in an area
    - Both cases are specific
  + No emptying on Sunday (and Saturday)
* Assumptions:
  + Constant filling rate over the planning period
  + Constant speed (i.e. time of travel is directly proportional to distance)
* Objective: Minimize cost of capital, labour and of operation in addition to overflowing costs:
  + Cost of capital:
    - Number of vehicles
    - Number of extra skips
  + Cost of labour: based on number of days operating
  + Operational costs: based on distance travelled
  + (Overflowing costs: based on amount overflowed)
* Constraints:
  + Visits following a regular pattern
  + **Emptying schedule preventing overflowing** (with some overflow allowable subject to high costs)
  + Cost constraints
  + Daily operation constraints (max 8h of travelling per day)

Currently, the optimization algorithm is implemented in MATLAB, and solved by Gurobi through YALMIP. The points in bold in the previous list have been implemented for now. It is to note that the mathematical formulation of the problem involve many equality and inequality constraints that do not show in the above logistical problem.

Figure 20 gives the schedule representation of a first optimization formulation. The implication of the daily cost is obvious in the fact that many skips are scheduled on the same days. This is also caused by the lack of time constraints for working days, which would limit this. Additionally, the filling rates in this case are randomly generated in a range between 0-0.5. The upper limit of 0.5 is because the model still does not offer any flexibility around overflowing (which should have an associated cost), and the collections are assumed to be at the same time every day (instead of morning/afternoon collections).

Additionally, the problem assumes one vehicle only, and that all collections are done by bringing an empty bin and exchanging it (meaning as many empty bins as vehicles). Since there are multiple trucks, the formulation will have to be adapted.

Finally, this is formulated as a one stage optimization problem, focusing on operation. However, the number of trucks, additional skips will become decision variables and the problem will become a multi-stage problem. This may increase complexity drastically and require dynamic programming.

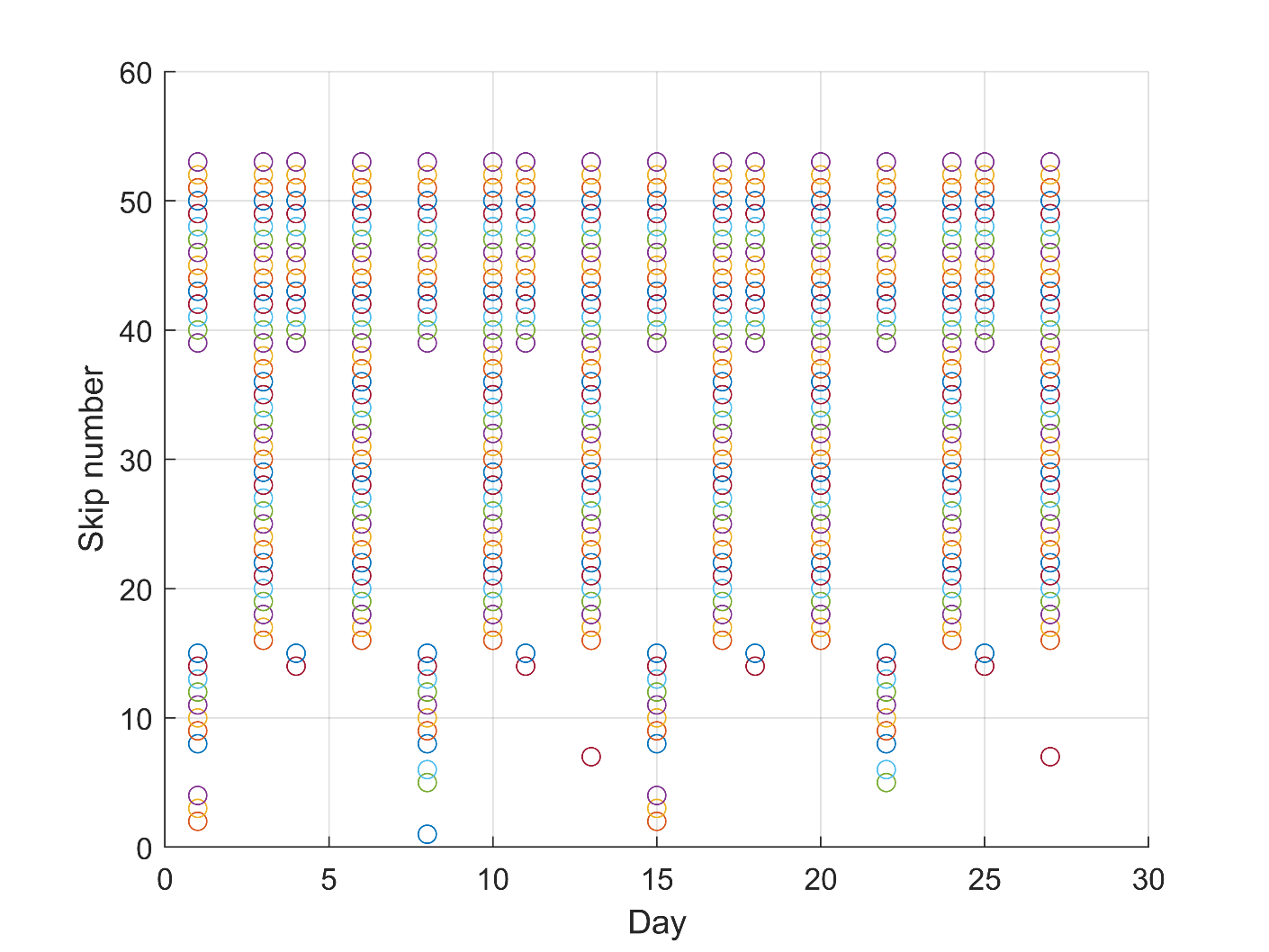


Figure 20 Schedule diagram (first version, more work will be done on operational planning representation)

## Results

# Conclusions

## Capital expenditure

## Operational planning

# Recommendations for future work

## Skip-level data

* **More complete data**, with a more robust scale, two measurements per day and a more spatially and categorically representative set of skips.

## Current organizational and operational state

* Current operational protocols
* More specific constraints
* Accurate costs analysis
* Measured time of travel between skips
* Stakeholder analysis

# References