Locating Roadside Wellness Centers using a Greedy Heuristic

Name student: Thomas Wieringa
Student ID number: 461169
Supervisor: L. van Rijn
Second assessor: A.P.M. Wagelmans

Abstract

Truck drivers in Africa have a lifestyle which puts them at great risk of catching HIV and various other diseases. This also puts the communities they are in direct contact with at risk. Non-governmental organization North Star Alliance battles this problem by installing Roadside Wellness Centers (RWCs) along various busy truck stops. To evaluate possible locations three objectives are combined. The literature presents a mixed-integer programming formulation to solve the problem. This formulation is not suitable for large instances, because the problem is NP-hard. To solve large instances this research proposes a Greedy Heuristic. This heuristic solves a large instance in 94.6% less time with a optimality gap of 0.00%

Bachelor Thesis Econometrics and Operations Research

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics
Erasmus University Rotterdam
the Netherlands

July 7 2019

Contents

1 Introduction

Africa is a very vast continent with large distances between populated areas. To supply these populated areas truck drivers need to travel long distances. Because of that drivers live a very monotone life while they are separated from their spouses and cultural norms for long periods of time Morris and Ferguson [2007]. Their working environment leads to irresponsible and dangerous sexual behaviour, having unprotected intercourse with many different partners [Morris and Ferguson, 2007]. Due to this these truck drivers and their partners have a very high risk of catching HIV and other sexually transmitted infections [Gatignon and Van Wassenhove, 2008]. Malaria and Tuberculosis are also linked to these long range truck drivers. Another reason for the high presence of diseases among truck drivers is that they are very often secluded from the traditional health care infrastructure [Morris and Ferguson, 2007]. Because of this seclusion truck drivers very often struggle to find adequate treatment.

In 2005 non-governmental Organization (NGO) North Star Alliance started to battle this problem by opening its first Roadside Wellness Center (RWC) in Malawi, Southern Africa. Currently there are 33 centers in service in 14 Sub-Saharan countries. These centers provide health care services to the drivers which pass by while also helping the communities that are in direct contact with the truck drivers, including sex workers. These centers provide general primary care. Next to that they give education and counseling to prevent disease. In addition they test and treat patients infected diseases. North Star Alliance wants to be able to reach as much truck drivers as possible. So they place RWCs at border posts, transit towns, ports or other high throughput locations like truck stops or gas stations to ensure as much truck drivers as possible can benefit from them. In the coming years North Star Alliance wants to build even more of these RWCs.

When deciding on the location of these facilities a complex problem emerges. The amount of possible locations is very large and there are different objectives to be accounted for. First of all it is important that as much people as possible can be reached. Maximizing the total amount of people serviced maximizes the amount of people which they can administer 'single shot' health care services. These single shot services include packs of condoms or HIV tests. These services provide patients with a basic level of health care access. It is also important to provide continuous access to as much people as possible. Drivers have continuous access if they are sufficiently close to a RWC during all of their trip. This is important for treatment adherence which in turn benefits disease prevention and treatment [Starfield et al., 2005] [De Vries H., 2014b].

De Vries H. [2014b] propose a mixed-integer programming formulation to locate a number of RWCs while optimizing the patient volume criterion and the continuous access criterion. Next to this they choose an optimal set of services for each RWC. Núñez Ares et al. [2016] then tackle the same location facility problem but with an additional equity criterion. This equity criterion tries to make sure every driver is treated equally independent of how busy the route they are driving is. The equity criterion also adds more complexity to the problem. De Vries H. [2014b] show their problem is NP-hard which makes solving times to optimality very long. The added complexity of the equity criterion makes these solving times even longer. Long solving times can restrict swift real-world implementation and decision making. As a possible solution to these long solving times I propose a Greedy Heuristic which solves problems considerably faster while retaining near-optimal solutions.

This paper has the following structure. In Section 2 I first review the literature written about similar problems and the health care context. Section 3 defines the patient volume, continuous access, and equity criteria. Section 4 contains a small summary of the instances I use for evaluating the methods. After that section 5 contains the methods used. Section 5.1 I outline the Direct MIP formulation [Núñez Ares et al., 2016]. Section 5.2 contains the Greedy Heuristic. After that I show the results in section 6. Section 7 contains conclusion and discussion.

2 Literature Review

The focus of this paper is providing and improving health care access to truck drivers. The setting in which this problem is solved is Sub-Saharan Africa. In Sub-Saharan Africa providing adequate health care is in itself already a large challenge. Most health care related problems in Africa deal with scarcity of resources. For instance Brandeau et al. [2003] and Deo et al. [2015] analyze the optimal allocation of different types of health care services. White et al. [2011] wrote an overview containing multiple uses of OR methods in developing countries. From their overview it is clear that there is more and more research in the OR field to battle diseases like HIV in developing countries.

The problem studied in this paper is essentially a facility location problem, since the main focus is selecting which facilities should be opened. Outside of the context of developing countries much has been written about locating facilities. An application of a facility location problem that is similar to this problem is the problem of locating a hospital. For example Mehrez et al. [1996] and Bennett et al. [1982] solve different hospital location problems with single objectives. Dökmeci [1979] and Calvo and Marks [1973] both propose facility location problems with multiple objectives. This problem differs from these problems by nature of the problem. In all of these problems distance from patient to facility is always a part of the objective. This comes from the fact that for locating a hospital all patients can be considered stationary, because their distance from the facility does not change much over time. In this problem this is different because the patients served in the RWCs are moving along different routes. Therefore this problem belongs to the class of 'flow interception' problems seeing that potential locations are optimized based on flow of patients or truck drivers in this case.

Because some routes are long, a single facility on a route does not necessarily cover all truck drivers on the route. De Vries H. [2014b] claim this is because some health care services require continuous access to really benefit patients. Continuous access means that drivers along the route are always within a certain critical time from the nearest RWC. Continuous access is important because it gives the truck driver the possibility to receive treatment or screening on a regular basis. Next to that it provides them with the possibility of quick treatment when complications arise. Overall having continuous access to health care services increases the likelihood of adherence to treatment which in turn strongly decreases the disease transmission and progression [Starfield et al., 2005].

I can further differentiate by naming the problem as a multi-coverage flow interception problem. In the literature there are two examples of this problem with the same kind of health care context as our problem: [De Vries H., 2014b] and [Núñez Ares et al., 2016]. The former proposes a MIP formulation for choosing locations for RWCs out of a given set of possible locations and furthermore decide on which health care services should be offered at the chosen locations. They optimize based on patient volume and continuous access. The latter proposes a new equity criterion to be added to the objective. An equity criterion aims to reduce the difference in coverage between different routes. Truck drivers very often drive the same route for long periods of time. So an equity criterion results in a more equal treatment for every individual. At the authors propose four different ways of measuring equity. These include the sum of absolute deviations

(SAD), the mean absolute deviation (MAD), the minimum effect (ME) and the Gini-coefficient (GC). The authors select the SAD measure based on the relative low solving complexity and good solution quality. The authors also propose a set partitioning formulation. Finally a column generation algorithm to solve larger instances. Apart from this column generation approach there is no heuristic available in the literature to solve the problem with the three objectives, continuous access, patient volume and equity.

However there a bit more is written about the similar refueling problem, where the location of refueling stations is optimized. Lim and Kuby [2010] proposes a Greedy adding, Greedy adding with substitution and Genetic algorithm for this problem. The problem the authors are solving in their paper differs from the current problem because coverage is measured in terms of distance and it lacks an equity criterion. Boccia et al. [2009] also propose different heuristics including a Greedy heuristic for solving different general flow intercepting problems.

3 Problem Description

The main studied in this paper is selecting a given amount of locations where a RWC should be located out of given set of possible locations. In addition to this set of possible locations there is a set of currents locations where a RWC is located. To formalize these sets I use the following notation conform [Núñez Ares et al., 2016]. KP and KC are the sets of possible and currents RWC locations respectively. The union of these sets is K and is the set of all locations and is indexed by k. x_k is the binary decision variable with value 1 if a RWC is opened and 0 if no RWC is opened at location k. The amount of locations to be selected from KP is called p. The quality of a solution is determined by the patient volume criterion, the continuous access criterion and the equity criterion. These criteria are denoted by Z_{PV},Z_{CA} and Z_{EQ} respectively. With these criteria the objective function becomes as in equation 1.

$$\max \quad w_{PV}Z_{PV} + w_{CA}Z_{CA} + w_{EQ}Z_{EQ} \tag{1}$$

In this function w_{PV},w_{CA} and w_{EQ} represent the relative weight of the criteria. The following subsections give explanation of and motivation for these criteria.

3.1 Patient Volume

The patient volume criterion is defined by the following function.

$$Z_{PV} = \sum_{k \in KP} d_k x_k \tag{2}$$

Here d_k is the expected volume daily volume of patients visiting location $k \in KP$. This criterion is intended to maximize the total amount of patients reached by the selected RWCs. Maximizing the amount of people serviced can be motivated by the nature of some services the RWCs provide. Some of these health services are 'single shot'. Examples of these services are testing for HIV or providing a pack of condoms. For these services it is important that as much people as possible are reached. The patient volume criterion encourages selecting RWC locations with high expected volumes of visitors.

3.2 Continuous Access

The continuous access criterion is defined by the following function.

$$Z_{CA} = \sum_{q \in Q} f_q c_q \tag{3}$$

Here the set Q, indexed by q, contains every truck route used by truck drivers. To each q I relate a $K_q \subset K$ which contains an ordered set of RWC locations which are on the route. These locations are ordered first to last visited. The amount of truck drivers using route q is denoted by f_q. The last variable c_q is defined as the coverage score of route q. This coverage score measures the level of sufficient access of route q given the chosen locations $\boldsymbol{x}$. The coverage score is calculated as is (4). Inside $g(.)$ is the fraction of time a driver is safe. A driver is safe if he is within a critical time limit from the nearest RWC. The coverage score is then calculated with piece wise linear function $g(.)$. This function is displayed in Figure 1. This function has two parameters: μ_1 and μ_2 which are the breakpoints of the function.

$$c_q = g\left(\frac{\sum_{j=1}^{m-1} \min\left(t_{(K_q^{\boldsymbol{x}}(j), K_q^{\boldsymbol{x}}(j+1))}, \tau\right)}{T_q}\right) \tag{4}$$

Equation (3) measures the total level of sufficient access. The motivation behind this criterion is that for some health services it is crucial that they are performed continuously. So optimizing this criterion gives a 'continuous access' to as much people as possible.

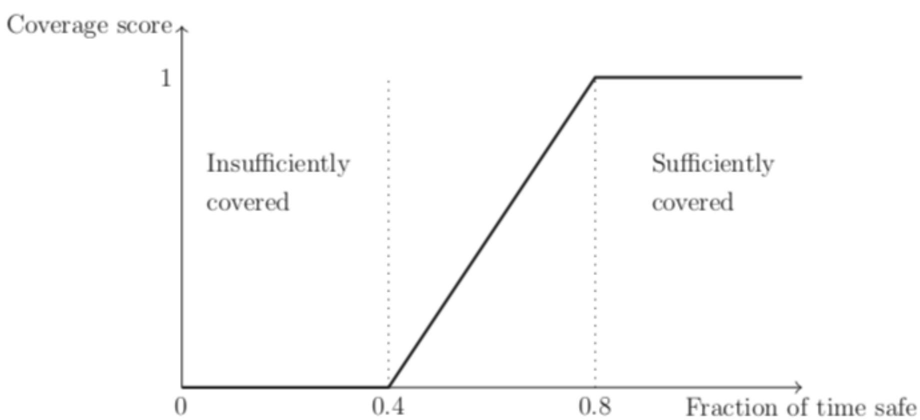


Figure 1: Coverage score as a piece-wise linear function of the fraction of time safe, with $\mu_1 = 0.4$, $\mu_2 = 0.8$. source: Núñez Ares et al. [2016]

3.3 Equity

The equity criterion is defined by the sum of absolute deviations (SAD).

$$Z_{EQ}^{SAD} = - \sum_{q_1, q_2 \in Q | q_2 > q_1} |c_{q1} - c_{q_2}| f_{q_1} f_{q_2} \tag{5}$$

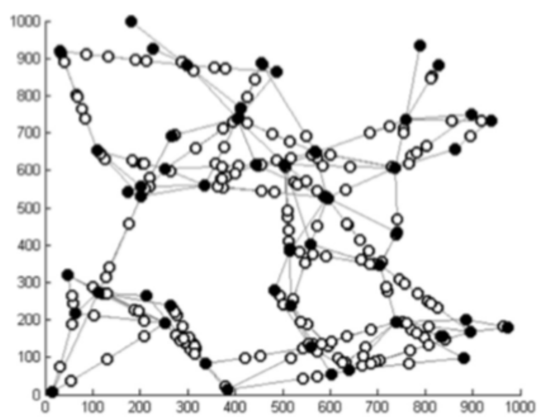
The value of this criterion increases as differences between coverage scores of different routes decrease. So maximizing this criterion reduces differences in coverage scores between routes. The motivation to include equity as a criterion is that a lot of truck drivers drive the same route for multiple years. Because of this every truck route can be seen as a different population. If only patient volume and continuous access would be accounted for then routes with a relatively low amount of drivers could easily be neglected in favour of more busy routes.

4 Data

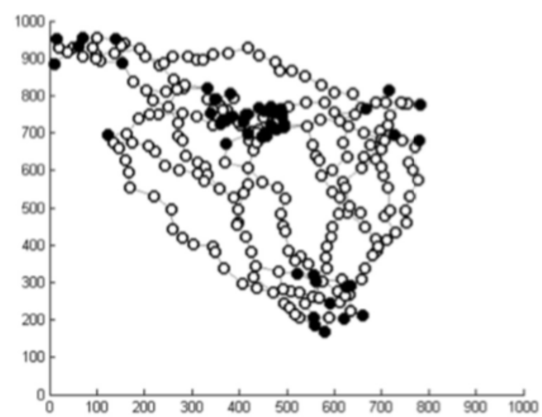
For this problem I use two different types of instances, clustered and non clustered. Both sets are made out of generated data. The exact method used for this random network generation can be found in Núñez Ares et al. [2016]. Both sets contain the following variables:

- K set of locations
- $KP \subset K$ set of possible locations
- $KC \subset K$ set of current locations
- Q set of routes
- f set of truck drivers per route
- d set of expected daily visitors per location
- An ordered list of locations linked to each route
- An ordered list of travel times linked to each route

For the non clustered instances different sizes are available. The amount of locations in KP ranges from 150 to 5000 and the size of Q from 75 to 2400. The clustered data sets all have the same size, KP and Q both have 200 values. All instances have the following notation: rxp$y$$z$. In this notation x indicates the number of routes, y indicates the number of possible RWC locations and z is a c or n to indicate whether the instance contains clustered or non-clustered data. In the clustered data sets there are a lot of location close to each other in a cluster. The non clustered sets have locations which are much more evenly distributed. Figure 2 shows plots of two random instances which show the difference between a clustered and non-clustered.



(a) Non-clustered instance



(b) Clustered instance

Figure 2: Examples of randomly generated instances. Black points represent current RWC locations and white points represent potential RWC locations. source: [Núñez Ares et al., 2016]

5 Methodology

This section propses the methods to select RWCs based on patient volume, continuous access and equity criterion. Section 5.1 presents a MIP formulation. Section 5.2 presents a Greedy Heuristic.

5.1 MIP

The first method to select RWCs based on the three criteria used is a MIP formulation which can be solved exact by a solver. I use the direct formulation [Núñez Ares et al., 2016].The objective is equal to (1), but with small change for programming the piece-wise linear function. Furthermore a small error in the formulation by Núñez Ares et al. [2016] is corrected in the formulation stated below. Namely it contains $(\Delta^+_{q_1q_2} - \Delta^-_{q_1q_2})$ to model the absolute difference of the coverage scores of routes q_1,q_2 in the Equity part of the objective. This is incorrect and should be $(\Delta^+_{q_1q_2} + \Delta^-_{q_1q_2})$.

$$\min \quad w_{PV}\sum_{k\in K} d_k x_k + w_{CA}\sum_{q\in Q} f_q c_q - w_{EQ}\sum_{q_1\in Q}\sum_{\substack{q_2\in Q\\ q_2>q_1}} \left(\Delta^+_{q_1q_2} + \Delta^-_{q_1q_2}\right) f_{q_1} f_{q_2} \tag{6}$$

$$\text{s.t.} \qquad c_q = \lambda_{3q} + \lambda_{4q} \qquad q\in Q \tag{7}$$

$$\lambda_{1q}0 + \lambda_{2q}\mu_1 + \lambda_{3q}\mu_2 + \lambda_{4q} = \frac{1}{T_q}\left(\sum_{k\in K_q}\sum_{l\in K_{kq}} i_{klq}\min\{t_{kl},\tau\}\right) \qquad q\in Q \tag{8}$$

$$\lambda_{1q} + \lambda_{2q} + \lambda_{3q} + \lambda_{4q} = 1 \quad q\in Q \tag{9}$$

$$\lambda_{1q} \le z_{1q} \quad q\in Q \tag{10}$$

$$\lambda_{2q} \le z_{1q} + z_{2q} \quad q\in Q \tag{11}$$

$$\lambda_{3q} \le z_{2q} + z_{3q} \quad q\in Q \tag{12}$$

$$\lambda_{4q} \le z_{3q} \quad q\in Q \tag{13}$$

$$z_{1q} + z_{2q} + z_{3q} = 1 \quad q\in Q \tag{14}$$

$$c_{q_1} - c_{q_2} = \Delta^+_{q_1q_2} - \Delta^-_{q_1q_2} \quad q_1,q_2\in Q, \quad q_2>q_1 \tag{15}$$

$$x_k = 1 \quad k\in KC \tag{16}$$

$$\sum_{k\in KP} x_k = p \tag{17}$$

$$\sum_{l\in K_q} i_{klq} = x_k \quad k\in K_q, \quad q\in Q \tag{18}$$

$$\sum_{l\in K_q} i_{klq} = 1 \quad k\in KC_q(1), q\in Q \tag{19}$$

$$\sum_{k\in K_q} i_{klq} = x_l \quad l\in K_q, \quad q\in Q \tag{20}$$

$$\sum_{k\in K_q} i_{klq} = 1 \quad l\in KC_q(m), q\in Q \tag{21}$$

$$x_k\in\{0,1\} \quad k\in K_q, \quad q\in Q \tag{22}$$

$$\Delta^+_{q_1q_2}, \Delta^-_{q_1q_2} \ge 0 \quad q_1,q_2\in Q, \quad q_2>q_1 \tag{23}$$

$$i_{klq}\in[0,1] \quad k\in K_q, \quad l\in K_{kq}, \quad q\in Q \tag{24}$$

$$\lambda_{iq} \geq 0 \quad i \in \{1,2,3,4\}, \quad q \in Q \tag{25}$$

$$z_{iq} \in \{0,1\} \quad i \in \{1,2,3\}, \quad q \in Q \tag{26}$$

Constraints (7)-(14) define c_q as in (4), with a piece-wise linear function. Constraint (15) makes it possible to implement the absolute value in the SAD part of the objective (5). Constraint (16) depicts the current RWC locations. Constraint (17) sets the amount of new RWC locations to p. Constraints (18)-(21) set the i_{klq} variables to the correct value. Finally (22)-(26) define the decision variables.

5.2 Greedy Heuristic

The facility location problem is strongly NP-hard [Núñez Ares et al., 2016]. Therefore solving the MIP with a solver can take a long time for large instances. A greedy heuristic could be able to solve the problem faster while also maintaining solution quality. The algorithm works as follows. It starts by initializing the current and possible locations. After that it always adds the location which renders the best objective. It stops when p locations are added to set KC. The pseudo code of the main algorithm is shown in Algorithm 1. Algorithm 2 shows the calculation for the coverage vector c in the objective. In this algorithm $selectedLocationsOnRoute(q, \boldsymbol{x})$ returns a vector with the locations which have an open RWC on route q given $\boldsymbol{x}$. $traveltime(KJ1, KJ2)$ returns the travel time between locations $KJ1$ and $KJ2$. Finally $piecewise(frac)$ is the function $g(.)$ as in (4).

Algorithm 1 Greedy Heuristic

$x_k \leftarrow 1 \qquad k \in KC$
$x_k \leftarrow 0 \qquad k \in KP$
$c \leftarrow |KC| + p$
while $|KC| \leq c$ **do**
  $largest \leftarrow 0$
  $indexchoice \leftarrow 0$
  for $k \in KP$ **do**
    $x_k \leftarrow 1$
    $obj \leftarrow objective(\boldsymbol{x})$
    if $obj > largest$ **then**
      $largest \leftarrow obj$
      $indexchoice \leftarrow k$
    end if
    $x_k \leftarrow 0$
  end for
  $x_{indexchoice} \leftarrow 1$
  $KC \leftarrow KC \cup k$
  $KP \leftarrow KP \setminus k$
end while

Algorithm 2 $c(\boldsymbol{x})$

```
for q ∈ Q do
    KXQ ← selectedLocationsOnRoute(q, x)
    KXQ ← KXQ \ KXQ(last)
    sum ← 0
    for i = 1 : |KXQ| do
        KJ1 ← KXQ(i)
        KJ2 ← KXQ(i + 1)
        time ← traveltime(KJ1, KJ2)
        minT ← min(sum, Theta)
        sum ← sum + minT
    end for
    frac ← sum/T(q)
    c(q) ← piecewise(frac)
end for
return c
```

5.3 Computational Experiments

For both the MIP and Greedy Heuristic I solve the model with the following parameters. $\mu_1 = 0.4$, $\mu_2 = 0.8$, $\tau = 100$ and $p = 20$. I solve the MIP with IBM ILOG CPLEX Optimization Studio in Java. I also program the Greedy Heuristic in Java but this does not make use of CPLEX. The computer I run these progams in has a 2 GHz Intel Core i7 processor and 8 GB of RAM running on macOS 10.14.5.

I compare both solving speed and solution quality of the MIP and Greedy Heuristic. First I solve ten r75p150n instances for both of the models. These instances contain 75 routes and 150 possible RWC locations. The locations in these instances are non-clustered. Next to that I solve three r200p200n instances and two r200p200c instances. Both of these instances contain 200 routes and 200 possible RWC locations. The difference is that they contain respectively non-clustered and clustered RWC locations.

Finally I use both methods to solve one large instance with 500 routes and 1,000 possible RWC locations to compare the ability to solve large instances.

6 Results

This section contains the results of both the Direct MIP formulation as well as the Greedy Heuristic. First I analyze the solving times for the different instances mentioned in section 5.3. I compare the results of the MIP solutions from CPLEX with the solutions of the Greedy Heuristic, which I refer to as GH from now on. Table 1 shows the solving times in seconds. Besides the solving times it also shows a Δ which is the difference between MIP and GH, calculated as follows: $MIP - GH$. The last column contains the difference in percents. This percentage has the following calculation: $\frac{\Delta}{MIP} * 100$.

Table 1 shows that that the GH algorithm is almost always faster than the MIP formulation for the r75p15n instances. It is always faster for the larger instances. It can also be seen that the GH algorithm obtains a solution using on average 66.36, 83,07 and 67.92 percent less time than the MIP algorithm for the

Table 1: Solving times in seconds for different instances

	MIP	GH	Δ	%
r75p150n1	41.59	9.066	32.52	78.20
r75p150n2	7.26	8.263	-1.00	-13.82
r75p150n3	22.24	8.514	13.73	61.72
r75p150n4	39.49	7.322	32.17	81.46
r75p150n5	10.64	9.134	1.51	14.15
r75p150n6	5.66	8.44	-2.78	-49.12
r75p150n7	32.09	8.728	23.36	72.80
r75p150n8	47.12	8.826	38.29	81.27
r75p150n9	42.23	8.264	33.97	80.43
r75p150n10	6.22	9.08	-2.86	-45.98
avg	25.45	8.56	16.89	66.36
r200p200n1	1,302.66	210.90	1,091.76	83.81
r200p200n2	570.49	173.19	397.30	69.64
r200p200n3	1,193.65	181.85	1,287.19	87.62
avg	1,022.27	188.65	833.62	81.55
r200p200c1	423.57	162.99	260.58	61.52
r200p200c10	609.58	156.50	453.08	74.33
avg	516.58	159.75	356.83	67.92
r500p1000n	16,878.74	911.25	15,967.49	**94.60**

r75p15n, r200p200n and r200p200c instances respectively. The table also shows that both methods perform better on the clustered instances than the non-clustered. For the largest instance the difference in solving time is also the largest. For this instance the GH uses 94.60 percent less time than the MIP.

Table 2: Objective values for different instances

	MIP	GH	Δ	%
r75p150n1	21,714.81	21,550.94	163.87	0.75
r75p150n2	21,961.88	21,961.88	0.00	0.00
r75p150n3	24,353.88	24,104.16	249.72	1.03
r75p150n4	21,343.47	21,259.51	83.95	0.39
r75p150n5	28,830.06	28,396.40	433.66	1.50
r75p150n6	26,476.29	26,476.29	0.00	0.00
r75p150n7	24,911.57	24,889.69	21.88	0.09
r75p150n8	24,282.23	24,282.23	0.00	0.00
r75p150n9	23,310.69	23,310.69	0.00	0.00
r75p150n10	24,407.58	24,407.58	0.00	0.00
avg			95.31	0.38
r200p200n1	45,574.62	45,496.60	78.02	0.17
r200p200n2	53,741.66	53,741.66	0.00	0.00
r200p200n3	51,696.16	51,696.16	0.00	0.00
avg			26.01	0.06
r200p200c1	60,846.40	60,481.29	365.10	0.60
r200p200c10	61,273.07	60,966.56	306.51	0.50
avg			335.81	0.55
r500p1000n	87,710.98	87,710.69	0.30	**0.00**

Table 2 states the objective values that belong to the final solutions of both the MIP and GH. The GH algorithm produces solutions with average optimality gaps below 1 percent. Namely 0.38, 0.06 and 0.6 percent for the three types of instances respectively. Also noteworthy is the quality of the large instance solution. For this solution the GH gives a solution 0.00 percent lower than the MIP solution while it reached the solution in 94.60 percent less time.

7 Conclusion & Discussion

This paper examines the problem of choosing a set number of locations for RWCs in Sub-Saharan Africa. The quality of a solution is determined by three criteria: Patient volume, Continuous access and Equity. Maximizing these criteria gives a solution which balances between serving as much patients as possible and providing the best coverage possible to each individual. I used the MIP formulation from previous research to model this problem. But because numerical experiments show that this problem is strongly NP-hard this formulation is not suitable for solving large instances. Therefore I propose a Greedy Heuristic. This Heuristic solves instances considerably faster than the MIP formulation. Although the solving times are much lower the solution quality does not differ much. The objective value decreases with on average less than 1% with respect to the MIP solution. The strength of the Greedy Heuristic is best illustrated with the result of the large instance. This instance was solved in 94.60 percent less time while with an optimality gap of 0.00%

The results are very good given the simplicity of this Heuristic. Numerical experiments show that Greedy Heuristics can be good at solving problems like the Set Covering problem and the knapsack problem. The problem studied in this paper can be seen as comparable to the knapsack problem since the task is selecting a subset of locations ,or "items", which maximize the objective. Although a Greedy Heuristic works well in a lot of cases for a knapsack problem, it is still possible that it returns a solution far from optimality. This can be the case if a item is selected early on which prevents other, more optimal, items to be selected later on. Relating this to the problem studied in this paper it seems that there are not many choices of RWC locations which prevent more optimal solutions later on.

In this research I also found some substantiations for the claim that the notation of Núñez Ares et al. [2016] could be incorrect. A reason for these doubts comes from the Heuristic. In the Greedy Heuristic I use the same objective as stated in the MIP but it lacks the need of MIP specific programming tricks to model the absolute value. Because of this it can exactly match equation (1). With this objective function I can do the following test. First I solve an instance with the MIP as Núñez Ares et al. [2016] propose. After that I calculate the objective again with the objective function from the heuristic. These objectives do not match. When I fix the mistake in the MIP and do the test again the objectives do match. Because this is a small mistake it could very well be that the authors did not make it in their implementation. But their results may indicate the opposite. The results of the Direct formulation in Núñez Ares et al. [2016] show solving times which are more than 10 times shorter compared to the ones I have. Nevertheless their machine is not that much faster compared to mine. When solving the same problems with the error added in the solving times reach comparable figures. Although I do not have access to the exact same instances as the authors used it could still be a good idea to verify the results the authors stated.

While the results of the Greedy Heuristic look good there are also a few aspects which could be improved. First of all there could be more testing on more and even larger instances.

Second the heuristic provides close to optimal solutions but the optimality gap is not constant. There could be examples of instances where the GH renders solutions which are far from optimal. It is possible that a case like this is present in further testing. Greedy Heuristics could in theory render very sub optimal solutions, because they choose the best option in each stage without looking back to check if a certain choice results in a wrong direction. A local search which looks back could render better objectives but it would require more research.

Furthermore this research only covers computer generated instances. If I could obtain data of actual truck routes in Sub-Saharan Africa as well as real possible and current RWC locations, I could compare the 'real world' performance versus the computer generated instance performance. In this paper I did not validate if the computer generated instances are representative for the real world case. So it could be that the algorithm does not work well for the real world case.

8 Appendix

Complementary to this paper there is a .zip file. This file contains the following Java classes which were used in this research.

- Data.java: This class stores an instance of data.
- Greedy.java: This class contains the Greedy Heuristic.
- Greedyrun.java: This class runs the Greedy Heuristic
- Model.java: This class constructs the MIP forumulation in CPLEX.
- MIPrun.java: This class runs the MIP formulation

References

Vivienne L. Bennett, David J. Eaton, and Richard L. Church. Selecting sites for rural health workers. *Social Science Medicine*, 16(1):63 – 72, 1982. ISSN 0277-9536. doi: https://doi.org/10.1016/0277-9536(82)90424-5. URL http://www.sciencedirect.com/science/article/pii/0277953682904245.

Maurizio Boccia, Antonio Sforza, and Claudio Sterle. Flow intercepting facility location: Problems, models and heuristics. *Journal of Mathematical Modelling and Algorithms*, 8(1):35–79, Mar 2009. ISSN 1572-9214. doi: 10.1007/s10852-008-9098-5. URL https://doi.org/10.1007/s10852-008-9098-5.

Margaret L. Brandeau, Gregory S. Zaric, and Anke Richter. Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis. *Journal of Health Economics*, 22(4):575 – 598, 2003. ISSN 0167-6296. doi: https://doi.org/10.1016/S0167-6296(03)00043-2. URL http://www.sciencedirect.com/science/article/pii/S0167629603000432.

Alberto B. Calvo and David H. Marks. Location of health care facilities: An analytical approach. *Socio-Economic Planning Sciences*, 7(5):407 – 422, 1973. ISSN 0038-0121. doi: https://doi.org/10.1016/0038-0121(73)90039-6. URL http://www.sciencedirect.com/science/article/pii/0038012173900396.

Van de Klundert J J & Wagelmans A P M De Vries H. The roadside healthcare facility location problem. 2014b.

Sarang Deo, Kumar Rajaram, Sandeep Rath, Uday S Karmarkar, and Matthew B Goetz. Planning for HIV screening, testing, and care at the Veterans Health Administration. *Operations research*, 63(2):287–304, 2015.

V F Dökmeci. A multiobjective model for regional planning of health facilities. *Environment and Planning A: Economy and Space*, 11(5):517–525, 1979. doi: 10.1068/a110517. URL https://doi.org/10.1068/a110517.

A Gatignon and L Van Wassenhove. Paving the road to healthy highways: a partnership to scale up hiv & aids clinics in africa. *Fontainebleau: INSEAD*, 2008.

Seow Lim and Michael Kuby. Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *European Journal of Operational Research*, 204(1): 51 – 61, 2010. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2009.09.032. URL http://www.sciencedirect.com/science/article/pii/S0377221709006717.

Abraham Mehrez, Zilla Sinuany-Stern, Tal Arad-Geva, and Shemuel Binyamin. On the implementation of quantitative facility location models: The case of a hospital in a rural region. *The Journal of the Operational Research Society*, 47(5):612–625, 1996. ISSN 01605682, 14769360. URL http://www.jstor.org/stable/3010013.

Chester N Morris and Alan G Ferguson. Sexual and treatment-seeking behaviour for sexually transmitted infection in long-distance transport workers of East Africa. *Sexually transmitted infections*, 83(3):242–245, 2007.

Núñez Ares, Harwin De Vries, and Dennis Huisman. A column generation approach for locating roadside clinics in Africa based on effectiveness and equity. *European Journal of Operational Research*, 254(3): 1002–1016, 2016.

Barbara Starfield, Leiyu Shi, and James Macinko. Contribution of primary care to health systems and health. *The Milbank Quarterly*, 83(3):457–502, 2005. doi: 10.1111/j.1468-0009.2005.00409.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0009.2005.00409.x`.

Leroy White, Honora Smith, and Christine Currie. OR in developing countries: A review. *European Journal of Operational Research*, 208(1):1–11, 2011.