Reinvent the Toilet Challenge

Validation of the transport logistic concept

Final Report

18.04.2012

This report has been written by Rafael Schmitt as term paper for his studies at ETH Zürich

Author:

Rafael Schmitt (schmittr@ethz.ch) **schmittrjp@gmail.com**

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

IfU
Institut für Umweltingenieurwissenschaften
Institute of Environmental Engineering

Chairs of Urban Water Management

General Comments on Master's Theses

As the Chairs of the Urban Water Management at ETH Zurich, we try to assign Masters Theses topics to our students that are related to current practical problems. The results of these Theses often attract wide interest. This motivates our students and us.

Our task is to evaluate the work of the Masters students and, concomitantly, their abilities. However, we take no responsibility for the content of the Theses.

Although the students take their work seriously, they are not yet experienced engineers. The results of their Theses are valuable. In practice, they should only be used, however, if the primary intent of the work as a Thesis project has first been accepted. Neither the Masters students, nor the Chairs of Urban Water Management take responsibility for what could occur if the work is applied into practice. This application is not done by ETH, Eawag or the Professor of Urban Water Management.

The students' work and accomplishments should not be underestimated. A lot of effort, time and commitment has gone into their work, and these efforts deserve mentioning. We cite Masters Theses in publications and technical reports in the following way:

As John Sample (2003) shows in his project work...
Based on the field campaign data of Jon Sample's (2003) project work...

List of literature:
Sample, John (2003): Investigating denitrification processes at the wastewater treatment plant XY, Masters' Thesis at the Department of Environmental Science, ETH Zurich

For questions regarding Masters Theses, please contact the authors directly or the Assistants of the Professor of Urban Water Management. They can forward to you the contact information of the academic advisors of the Masters students and of any other involved persons.

ETH Zurich
Chairs of Urban Water Management
Institute Environmental Engineering (IfU)
HIL G31.2
Wolggang-Paulo-Str.15
8093 Zürich

Phone +41 44 633 3073,
Email: sww@ifu.baug.ethz.ch

Zurich, June 26th, 2014

Prof. Dr. Eberhard Morgenroth

Abstract

Providing the urban poor with safe and affordable sanitation is a major challenge in informal settlements due to a lack of funding and appropriate technologies. The Bill and Melinda Gates Foundation (BMGF) organizes a competition to develop new technical and operational approaches towards this problem. Rules of competition include an ambitious cost limit (0.05 \$ p^{-1} d^{-1}). The Swiss Federal Institute of Aquatic Science and Technology (EAWAG) enters this competition with a system that combines innovative source separating toilets with internal water recycling and a community level plant for excreta processing. In order to guarantee long term functioning and community acceptance, a reliable and financially sustainable transport concept for accruing excreta is central. To provide proof of concept for the proposed transport system a numerical model was developed. Uncertainties in input variables are considered using a Monte Carlo Analysis (MCA). Routes for the serviceperson were generated using a MCA based on data derived from a Geo Information System (GIS) and remote sensing data of four informal settlements. The impact of several input parameters on the overall system capacity and cost was assessed. Results indicate that efficiency of the service person is most crucial to the system and that it can be operated in a wide range of spatial setups. The applied methodology is applicable to design and manage a wide range of logistic related services (e.g. municipal waste collection) in informal settlements based on cost and efficiency criteria.

Contents

1. Introduction

GIS based optimization of operations is a standard procedure for providers of transport services in industrialized countries. Several studies also indicate a major potential for increasing the efficiency of collection or delivery systems by route optimization methods in developing and transitioning countries. Nevertheless, these studies focus on city-scale planning. No literature is available on quantitative planning or optimization of transport systems in informal settlements where transport services are often provided by individuals or small companies. Accordingly, no literature exists on how to model logistic systems in informal settlements to quantitatively plan new or to improve existing service systems.
The concrete example of EAWAG's proposal to the Bill and Melinda Gates Foundation's "Reinvent the Toilet Challenge", which strongly relies on a transport system for human waste, is the base for a newly developed approach towards the quantitative assessment of logistic systems in informal settlements. A numerical model was developed in order to quantify the performance of the proposed transport system and to prove its feasibility. To cope with major uncertainties regarding system design and economic parameters, a Monte-Carlo Tool was integrated into the model to evaluate possible ranges for system capacity and costs.
Travel distances are supposed to have a major influence on the capacity of the transport system as well as on its costs. Relevant structures (road network, houses) were mapped out for four informal settlements in India and Africa. Remote sensing data served as base for a Monte Carlo simulation of service rounds. For the selected settlements the dependency of travel distances on user density (user/ha) was determined in order to simulate impact of increasing system propagation.
In the view of the global target of the "Reinvent the toilet challenge" a regression analysis was applied to spatial data and simulation results in order to quantify key dependencies between parameter values and system performance beyond concrete settings.

2. Logistics concept

The proposed sanitation system consists of shared toilet facilities and a community based treatment unit referred to as Resource Recovery Plant (RRP). Each RRP will have a vehicle and a service person to collect excreta from connected facilities. The number of facilities connected to one RRP will be dependent on the capacity of the logistic system. User interfaces consist of source-separating toilets, which are shared between two families. Each facility will contain two user interfaces and serve four families accordingly. With a mean family-size of 5 persons one facility will serve 20 persons. A minimum capacity of 500 persons (25 facilities) is planned for the RRP.

Toilets are movable and rented by the users rather than bought. Thus, they can be removed in case of payment default or tenancy changes. Facilities will be emptied by a service person twice a week. The service person is equipped with a small motorized vehicle for traveling and excreta transport. During a service round, the service-person will travel from house to house until its capacity is depleted. After unloading at the RRP a new round will commence until all scheduled facilities are serviced (Figure 1). Logistics of excreta must not cost more than 0.015 $/p/d (one third of the total budget of 0.05 $/p/d).

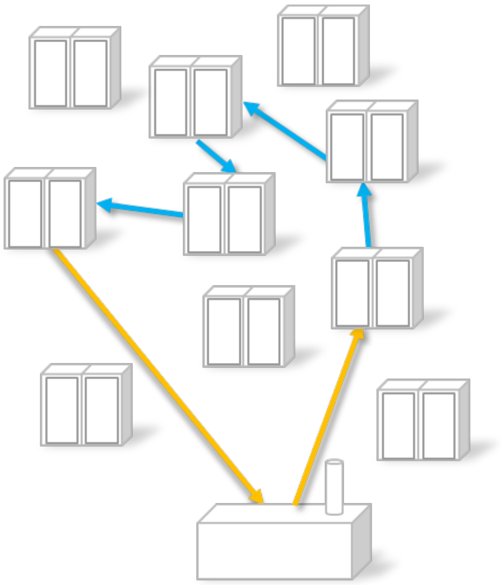


Figure 1: Representation of a service round. A service round consists of two distinct elements:

1. The travel distances to and from the RRP (orange)
2. The travel distances in between facilities (blue)

3. Areas under study

To quantify travel related costs and travel times the numerical model is dependent on spatial data to calculate lengths of service rounds. These data were derived from four informal settlements with different geographic contexts and settlement structures. Based on available field experience, selected study areas cover the required range of population densities and settlement patterns. 3 study sites are located in the slum district Kyebando-Kisalosalo in Uganda (0°21’05.52’’N/32°35’05.52’’E) while the fourth is based on informal settlements in the city of Raipur, India (21°13'56" N / 81°35'33"E) (compare Appendix A 1.).

4. Methods

4.1. Numerical Model

The model to simulate costs and capacity of the transport system is based on Microsoft Excel in combination with a Monte-Carlo Analysis tool (Simulacion 4.0). This tool allows for considering the major uncertainties in system parameters which are either immanent to the system (e.g. fluctuating excreta accumulation, fuel prices and working hours) or due to not yet completed technical design (e.g. sludge accumulation in the internal water recycling, service time to empty a facility). The MCA-tool was also used to determine the distribution of travel distances. Input parameters for the generation of travel distances were based on a GIS analysis of four concrete informal settlements.
Modeling of required travel distances is obviously a crucial part for the numerical modeling of a transport system. Nevertheless, there are major differences to common spatial optimization problems. First, the exact position of system elements (RRP, facilities) is not known prior to system implementation. Second, the system is supposed to be highly dynamic, with new users being connected and repositioning of facilities. Accordingly, modeling of travel distances is on purpose not based on predefined, optimized service rounds. Rather, the aim is to assess ranges of travel distances and related costs and system capacity.
Therefore, the model makes use of distance distributions which were measured using the GIS. Accordingly, a method for the determination of distance distributions was developed based on the proposed logistics concept (chapter 2).

4.1.1. Digitalization of areas under study

Spatial data collection was based on satellite images of the four selected study areas using a GIS-Software (Esri ArcGIS 10.0). Satellite imagery was derived via BingMaps©. Based on these images housing structures and the road network were digitized. According to their size, visible housing structures were assigned one or several households. To calculate travel distances between points in the study area, the road network was digitalized in four classes: paved roads, large (unpaved) roads, small (unpaved) roads and footpaths. While roads of the first three categories can be traveled on with motorized vehicles, paths are only accessible by foot.

4.1.2. Determination of distances between system elements

In a first step, it is assumed, that all households are potential positions for facilities. Then, a certain number was randomly selected from the population of all households in order to represent facilities. Several suitable RRP-positions were selected in each study area. From the RRP positions the GIS automatically selects the position which minimizes travel distances between RRP and facilities. In the proposed service concept twice weekly service is scheduled for each facility. Service on demand is also possible if the holding capacity of a facility is depleted in advance. It is assumed, that the service-person will not follow a distinct and optimized service round to service the facilities. The service person will rather plan his service round based on e.g. personal experience, preference or road conditions. This implies that the facilities serviced on a certain day will not be direct neighbors. The number of facilities that require service on a certain day depends on the total number of facilities connected to the RRP. In general, each facility is serviced twice a week. Thus the number of facilities which require service is calculated according to:

Equation 1

$$n_{req} = 2 * \frac{C_{RRP}}{d}$$

n_{req} : Facilities to be served [d^{-1}]
C_{RRP}: Capacity of the RRP [*facilities*]
d : work days [$d\ week^{-1}$]

The probability that a facility requires service on a certain day is

Equation 2

$$p = \frac{n_{req}}{C_{RRP}}$$

n_{req} : Facilities to be served [d^{-1}]
d : work days [$d\ week^{-1}$]

Combining Equation 1 and Equation 2 allows calculating the absolute value of p assuming 6 work days per week:

Equation 3

$$p = \frac{2 * \frac{C_{RRP}}{d}}{C_{RRP}} = \frac{2}{6} = 0.33$$

Thus, the probability that a facility does not require service is $1\text{-p} = 0.67$. The probability *q* that there is no facility scheduled for service within the next *i* facilities is

Equation 4

$$q_i = (1 - p)^i$$

n_{req} : Facilities to be served [d^{-1}]
d : work days [$d\ week^{-1}$]

For $i > 8$, q_i becomes smaller 5%. Thus, for all facilities, there is a 95% probability that there is a full facility within its next 8 neighbors. Accordingly distances from each facility to its next 8 neighbors were measured with the GIS. The number of facilities was varied in each setting to simulate different penetration rates. As population densities vary between the study areas different user densities result, where user density is defined as

Equation 5

$$\rho_{user} = \rho_{pop} * P$$

ρ_{user}: user density [user ha^{-1}]
ρ_{pop} : population density [inhabitants ha^{-1}]
P: Penetration rate [% of inhabitants connected to the system]

A distinct empirical distance distribution for RRP-facility and facility-facility travels resulted in each setting in dependence of the user density. Analytical distributions were fitted to the determined empirical distributions distances using Matlab. According to determined distribution parameters, the MCA-tool generated input values for the simulation of travel distances in the model (Figure 3).

4.2. Structure of the numerical model

As described in section 2, a service round consists of 2 distinct route elements. The RRP-facility and facility-facility distances (Figure 1). The number of facilities serviced per round depends on the capacity of the vehicle and the weight of excreta in the facilities. It was assessed, how many facilities can be serviced by the selected vehicle before its capacity is depleted. Therefore, a MCA over 6240[1] work days (20 years) was used simulating the excreta content of 25 individual facilities (for parameter values refer to Table 2). Results indicate that the capacity of the service vehicle (2 wheel tractor) is with a 97 % probability 4 facilities/round (Figure 2). Its capacity is thus considered to be constantly 4 facilities/round[2]. As a consequence, a service round for the 2-wheel tractor consists of 2 RRP-facility and 3 facility-facility (compare Figure 3) distances. The distance of a service round is:

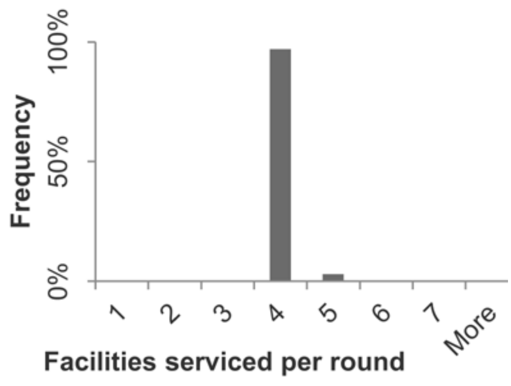


Figure 2: Based on expected distributions of excreta accumulation in the facilities, a Monte Carlo Simulation was used to evaluate the service capacity of the vehicle. In over 97 % the capacity of the vehicle is four facilities per round.

Equation 6

$$d_{j,Round} = 2 * d_{Fac-RRP} + \sum_{i=1}^{3} d_{i,fac-fac}$$

$d_{j,Round}$: Distance traveled during a distinct service round j [m]

$d_{Fac-RRP}$: RRP-Facility distance [m]

$d_{i,fac-fac}$: Facility-Facility distances [m]

For each service round values for $d_{j,Fac-RRP}$ and $d_{i,fac-fac}$ were generated separately using the MCA-tool. Time required for one service round is given by:

Equation 7

$$t_{j,Round} = \frac{d_{j,Round}}{v} + t_{service}$$

$t_{j,Round}$: Required travel time during service round j $[\mathrm{h}]$

$t_{service}$: Time to empty and service facilities and to dispose of extracted excreta at the RRP $[\mathrm{h}]$

v: Mean velocity of the vehicle $[\mathrm{km/h}]$

The number of service rounds a service-person can complete per day is limited by its working hours.

Equation 8

$$\sum_{j=1}^{n} t_{j,Round} \leq t_{work}$$

n: Completed service rounds per day [-]

t_{work}: Daily maximum working hours [h d^{-1}]

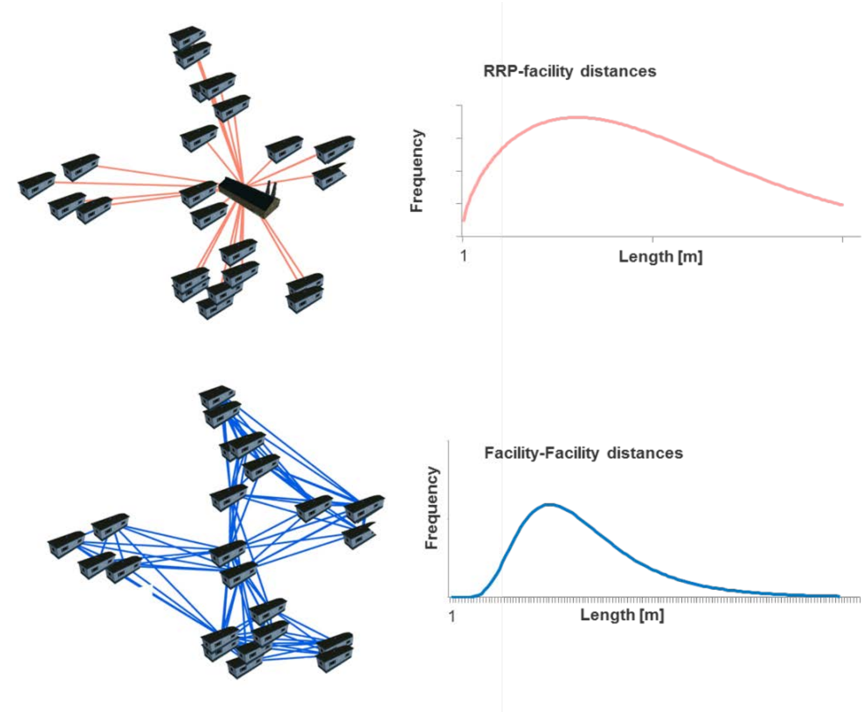
---

[1] =52 weeks/year * 6 workdays/week * 20 years

[2] If further information concerning the distribution of excreta accumulation becomes available, this assumption should be verified.

Normally, though there is not enough time for a new, complete service round, some service time remains at the end of the workday. The model also calculates how much facilities can be serviced in the remaining time.

GIS



NUMERICAL TRANSPORT LOGISTIC MODEL

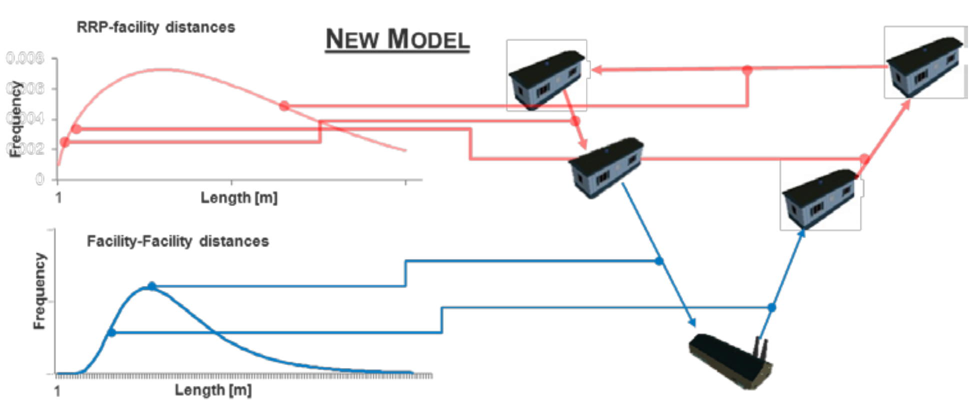


Figure 3: Modeling of round distances in the numerical model. Distributions of RRP-facility and facility-facility distances are determined using a GIS (top). A service round is simulated by separately generating values for each distinct route element (bottom).

Table 1: Input parameters for the transport logistics model as used throughout this report.

1) Distributed inputs	Mean ± Stdev.
1a) Product Inputs	
Urine [kg p^{-1} d^{-1}]	1.14 ± 0.27
Faeces [kg p^{-1} d^{-1}]	0.23 ± 0.16
Sludge accumulation [kg p^{-1} d^{-1}]	0.08 ± 0.008
1b) Socio economic	
Effective working hours [h d^{-1}]	8 ± 1.6
Fuel costs [$ L^{-1}]	1.6 ± 0.32
2) Constant inputs	**Value**
Work days [d $week^{-1}$]	6
Cost of labor [$ p^{-1} d^{-1}]	5
Interest rate [% yr^{-1}]	6
Worker/vehicle	1
Vehicle lifespan [yrs]	10
Max. vehicle capacity [kg]	500
Vehicle speed [km/h]	5
Fuel Demand [l/h]	2
Purchase Price [$]	1700
Maintenance [% of purchase price yr^{-1}]	10
Percentage of facilities located on paths [%]	61[3]

[3] Average value of the four areas under study (compare chapter 5.1)

4.3. Regression analysis of spatial data

To provide general proof of concept beyond concrete study areas, a regression technique was used to analytically describe system behavior. For each setting and user density a characteristic mean round length was calculated from the measured distances. A regression analysis of user density ($\boldsymbol{\rho_{User}}$) [user ha^{-1}] vs. characteristic round length was done. User density as explanatory variable was selected rather than total user number because it considers catchment area and system penetration as well as local population density. This analysis was performed in order to assess if a) user density is a suitable explanatory variable for travel distances and b) to assess to which degree travel distances are dependent on user density and setting. In order to include also cost respectively capacity in the analysis two analysis steps were performed.

Step 1 - Round lengths vs. user density: $\boldsymbol{d_{Round}} = \boldsymbol{f}(\boldsymbol{\rho_{User}})$

As each service round includes the same number of facilities (see chapter 4.2) a mean round distance can be calculated for study area and user density:

$$\bar{d}_{Round} = 2 * \bar{d}_{Fac-RRP} + 3 * \bar{d}_{fac-fac}$$

Facility-facility distances

The user density (and accordingly the facility density) has a major influence on travel distances. In addition, facility-facility distances also depend upon the settlement pattern, e.g. separation of an informal settlement into several “pockets” and the structure of the road network. To represent different user densities in the four areas under study, the number of connected households was varied between 100 and 1000 (500 – 5000 user) in each area. The resulting facility-facility and RRP-facility distances were measured.

RRP-facility distances

With around 500 users/RRP and a user number of 500-5000 p the number of RRPs would vary between 1 and 10. The determination of RRP-facility distances thus required assumptions on the expansion strategy of the RRP organization. Two scenarios were considered:

Scenario A: Flexible RRP location with movable RRPs

The RRPs are not fix but e.g. installed in movable containers. As soon as the capacity of the existing RRPs is depleted, a new RRP is brought into the study area. The original catchment area is divided into two distinct sub-catchments and each RRP is located as centrally as possible in its catchment area. The more RRPs are constructed, more and ever smaller catchments will result.

Scenario B: Fix RRP location

While expanding the user number and the number of RRPs, the original RRP-position remains the same. Thus, either the capacity of the original RRP is increased or new RRPs are located just next to the initial RRP. Scenario B can also be conceived as worst case scenario in settlements where there is no space available for the construction of several small RRPs (due to a lack of open area, road access, safety regulations etc.).

While for scenario B only facility-facility distances decrease with the number of connected users, scenario A will lead in addition to a decrease in RRP-facility distances (For a visualization compare Appendix A 2). Of course, also intermediate scenarios are conceivable, where the original RRP position remains at its position, but each new RRP is located at a new position. So far, there is no tool available for automated clustering of facilities; therefor the decrease in RRP-facility distances is simulated based on geometric assumptions [4](cf. Appendix A 3).

Step 2.: Costs and Reliability vs. Round distances: $\boldsymbol{c, r = f(d_{Round})}$

Modeled travel costs and reliability were correlated with generated round lengths. Combining results from step 1 and step 2 leads to

Equation 9

$$\mathbf{c, r = f\left(d_{Round} = f(\rho_{User})\right) = \ f(\rho_{User})}$$

Thus, performance in terms of cost as well as reliability can now be expressed in dependence of user density ($\boldsymbol{\rho_{User}}$) only.

4.4. Single stance sanitation system

Single stances might be required in settings where double stance toilets are either not feasible due to spatial restrictions, existing superstructures with limited volume or acceptance issues. A facility will not serve 4 families (2 toilet stances) but only 2 families (1 stance). Effects of the increased number of service events on system costs and capacity will be evaluated.

[4] The required functions for direct clustering and RRP-facility locations will be includes in the next version (10.1) of ArcGIS which is to be released soon

5. Results

5.1. A spatial anthology of selected informal settlements

The consideration of spatially diverse settlements to provide generic proof of concept is a key element of this report. Results of the spatial analysis of the four informal settlements under study indicate that selected settlements cover a variety of critical states such as low population density, very high population density, poor road access and discontinuous settlement areas (Table 2). Figure 4 shows the spatial distribution of characteristics. For all settings, except Raipur, Ward 68 the number of required RRPs exceeds the number of suitable locations (Table 2). The most extreme situation in this regard is the dense area in the Kyebando-Kisalosalo settlement, where only one suitable location is available (Table 2) (for detailed maps of the areas refer to appendix A 1).

Table 2: Characteristics of selected informal settlements

SETTING	SETTLEMENT AREA [HA]	TOTAL POPULATION [P]	POPULATION DENSITY [P HA^{-1}]	REQUIRED RRPS/ AVAILABLE POSITIONS	% OF HOUSES ON PATHS	PATH LENGTH MEAN/MAX [M]
KYEBANDO KISALOSALO COMPLETE	56.5	10678	189	22 / 16	61	25.8/66.5
KYEBANDO KISALOSALO LOW DENSITY	28.9	3700	128	8 / 9	46	34.1/177
KYEBANDO KISALOSALO HIGH DENSITY	6.0	3120	520	7 / 1	83	26.6/68
RAIPUR WARD 68	9.8	3163	320	7 / 7	55	21.4/81

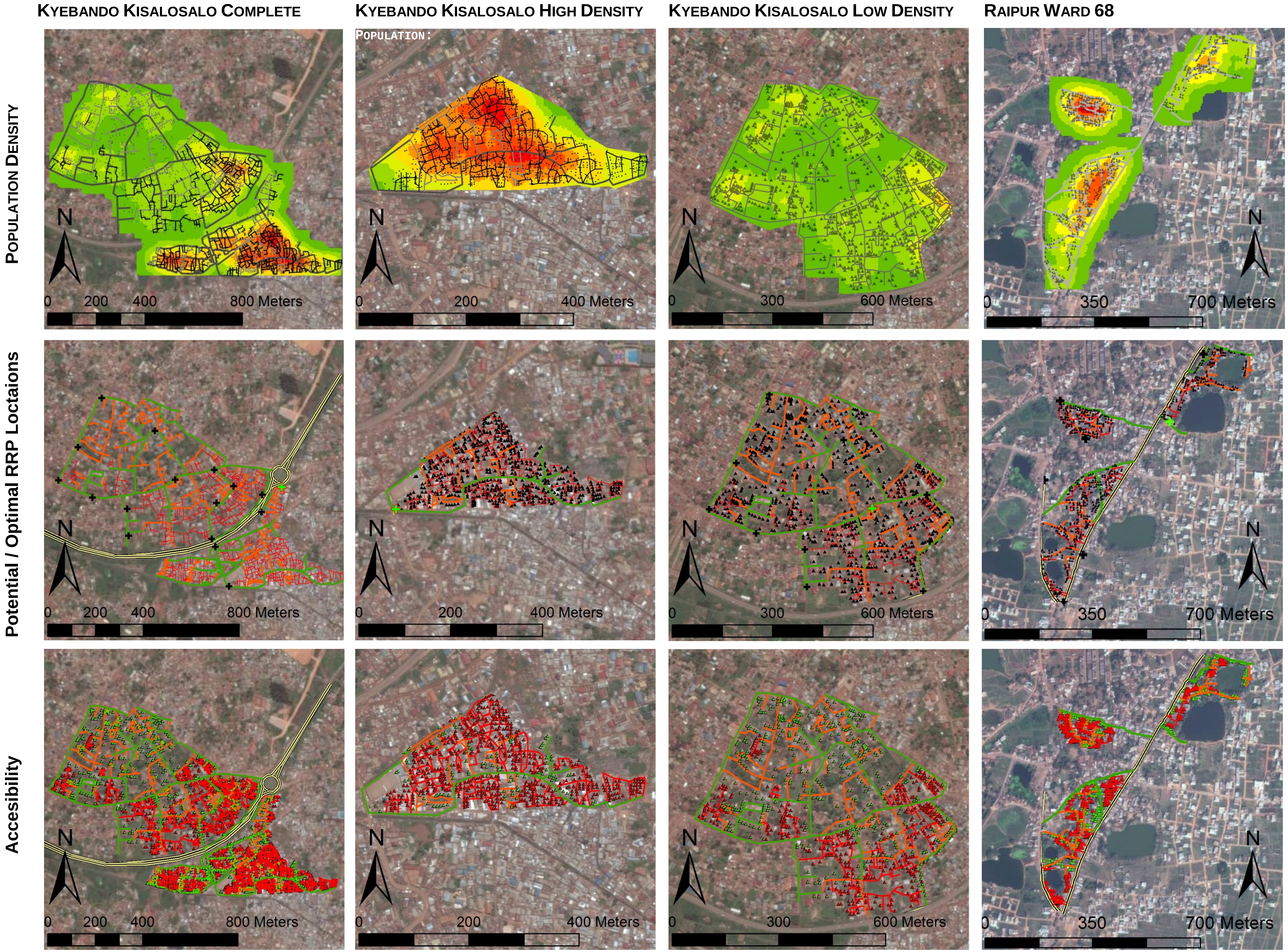


Figure 4: Spatial overview over most relevant attributes of areas under study. Kyebando Kisalosalo (High Density) and Raipur are considered especially critical, the former due to its very high population density, low accessibility and low number of suitable RRP-locations the latter for its discontinuous settlement pattern.

LEGEND

POPULATION DENSITY

Population density [p/ha]

- 0 - 5
- 5 -100
- 100 - 200
- 200 - 300
- 300 - 400
- 400 - 500
- 500 - 600
- ▲ Households
- Streets

Potential / Optimal RRP Loctaions

Road Network

- Paved
- Large
- Small
- Path
- Optimal RRP Position (scenario B)
- Potential RRP Positions

Accesibility

Facilities

- Inaccessible
- Accessible

5.2. Regression analysis

5.2.1. Regression analysis of spatial data

RRP-Facility and facility-facility travel distances were assessed for the four selected areas and varying user densities. As expected, the analysis reveals a strong anti-correlation between user density and the facility-facility distances. For low user densities there is a major difference in between the settings. For Raipur, Ward 68 the highest facility-facility travel distances were found. Due to the fragmented settlement structure the service round includes travels between the distinct slum pockets. The lower the user density, the less facilities are located in the individual pockets and the more travels in between pockets are required. The difference between "Kyebando-Kisalosalo Low density" and "Kyebando-Kisalosalo Complete" is not significant (p=0.42, Wilcoxon Rank Sum Test). In contrast, the difference of these two settings with the "High Density" setting is significant (HD/Complete: p=0.0017, HD/LD: $p=5.8*10^{-4}$). This indicates that for the settings in the Kyebando-Kisalosalo the population density has a major influence on the facility-facility distances. Most likely, this can be explained, as the different settings in Kyebando Kisalosalo show also different spatial characteristics, e.g. in terms of the road network. The higher the population density, the lower is the fraction of houses that are directly accessible via roads. In the "Kyebando-Kisalosalo HD" setting only a minority of facilities has direct road access, but those with road access are located basically along one major road. For the other two settings, the road network is much more branched, which increases the travel distances between the accessible facilities (compare appendix A 1). For all settings it was shown that facility-facility distances converge for user densities above around 50 users/ha.

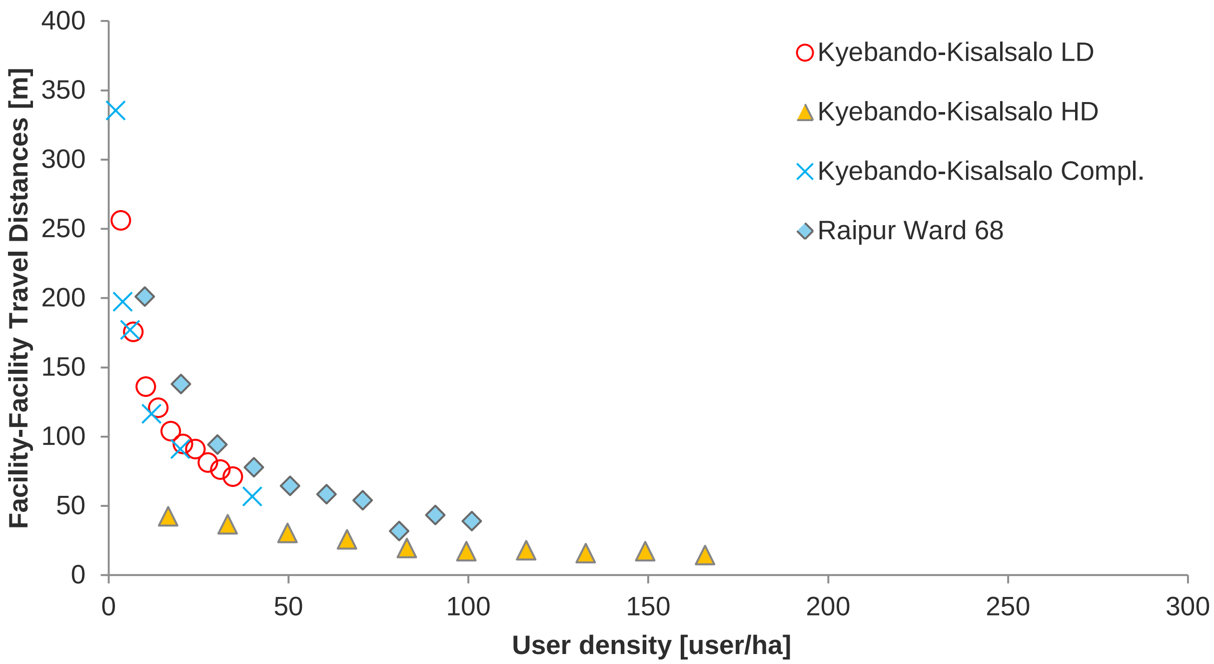


Figure 5: Spatial analysis reveals the strong anti-correlation of facility to facility travel distances for all areas under study. For all areas under study an increase in user density results in a decrease in mean facility- facility distances.

Mean round lengths were calculated according to scenarios A and B for the different settings. For scenario B, where additional RRP capacity is enlarged at the initial RRP position, total travel distances decrease asymptotically towards the RRP-facility distance (compare Figure 6). Assuming, that new users are connected homogenously in the study area an increase in user numbers will lead only to a limited decrease in RRP-facility travel distances. There is not only a major decrease in overall travel distances for scenario A, in addition the smaller catchment areas for individual RRPs will decrease variability in RRP-facility distances (Figure 6).

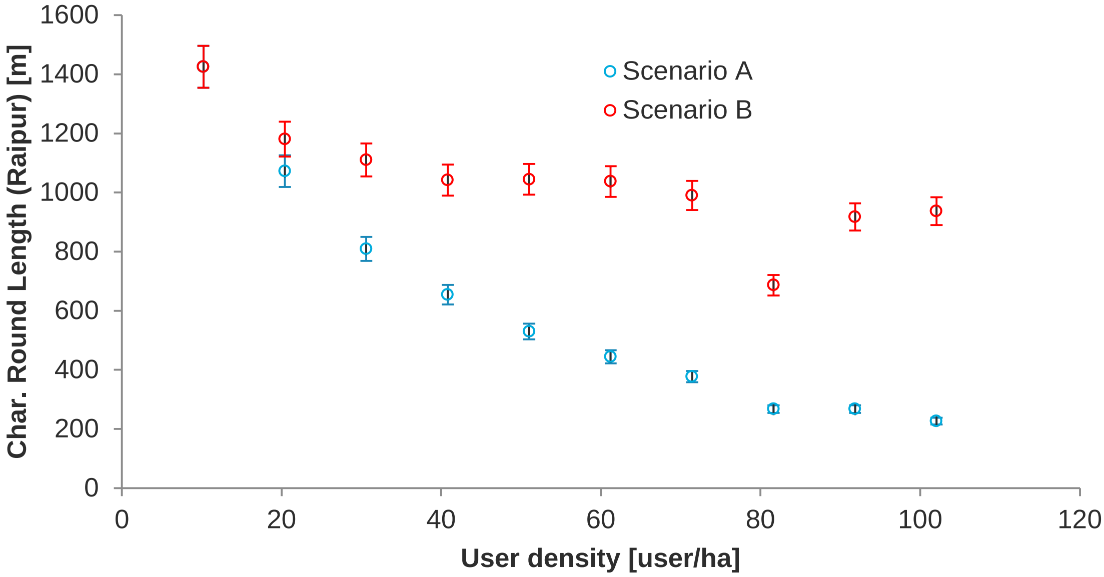


Figure 6: RRP-facility distances for scenario a and b (fixed vs. flexible RRP positioning), based on the informal settlement in Raipur, India (error bars indicate 1 standard deviation).

A power law regression was applied to the characteristic round length data. Results indicate a high explanatory power of the power law regression (Scenario A: R^2=0.786, Scenario B: R^2=0.77) (compare Figure 7). Figure 7 also demonstrates the effect of optimized RRP positioning in Scenario A. Although there exist differences in required travel distances in between the settings, the regression is able to explain a major part of the

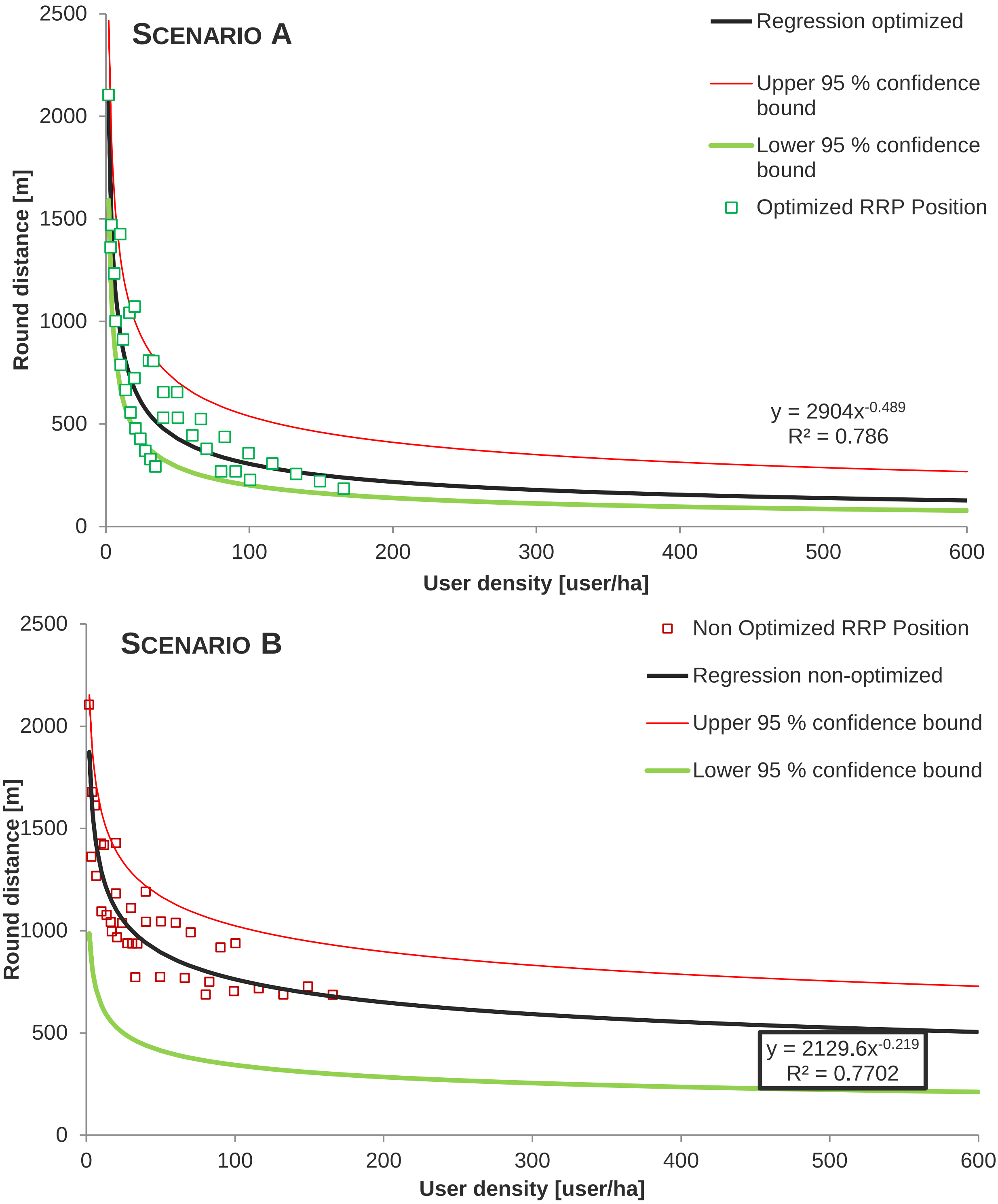


Figure 7: Comparison of characteristic round distances for scenarios a and b. For scenario A, increasing user density results in a pronounced decrease in round distances and distance variability in comparison to scenario B.

5.2.2. Regression analysis of system parameters

In addition to travel distances, a variety of factors influences the system's capacity. In order to identify the parameters with the most significant impact on the system and to quantify their impact on cost and capacity regression analyses were used. Four parameters were selected as explanatory variables:

[1] Mean travel time/round
[2] Service time (time to empty one facility and to dispose of the related products at the RRP. Travel time on paths to inaccessible facilities is not included)
[3] Effective working hours
[4] Travel distances on footpaths.

Correlation between capacity and the four explanatory variables is shown in Figure 8. The coefficient of correlation between characteristic route length and capacity is near zero. Higher correlations are found for efficiency indicators (effective working hours, service time) and travel distance on paths.

The low correlation between time per round and capacity can be explained by the characteristics of the service system: First, the velocity of the vehicle is relatively high (4 km/h) second, only service 4 facilities need service per round, thus rounds are rather short. Both factors keep required travel distances and travel times relatively low (compare also Figure 7).

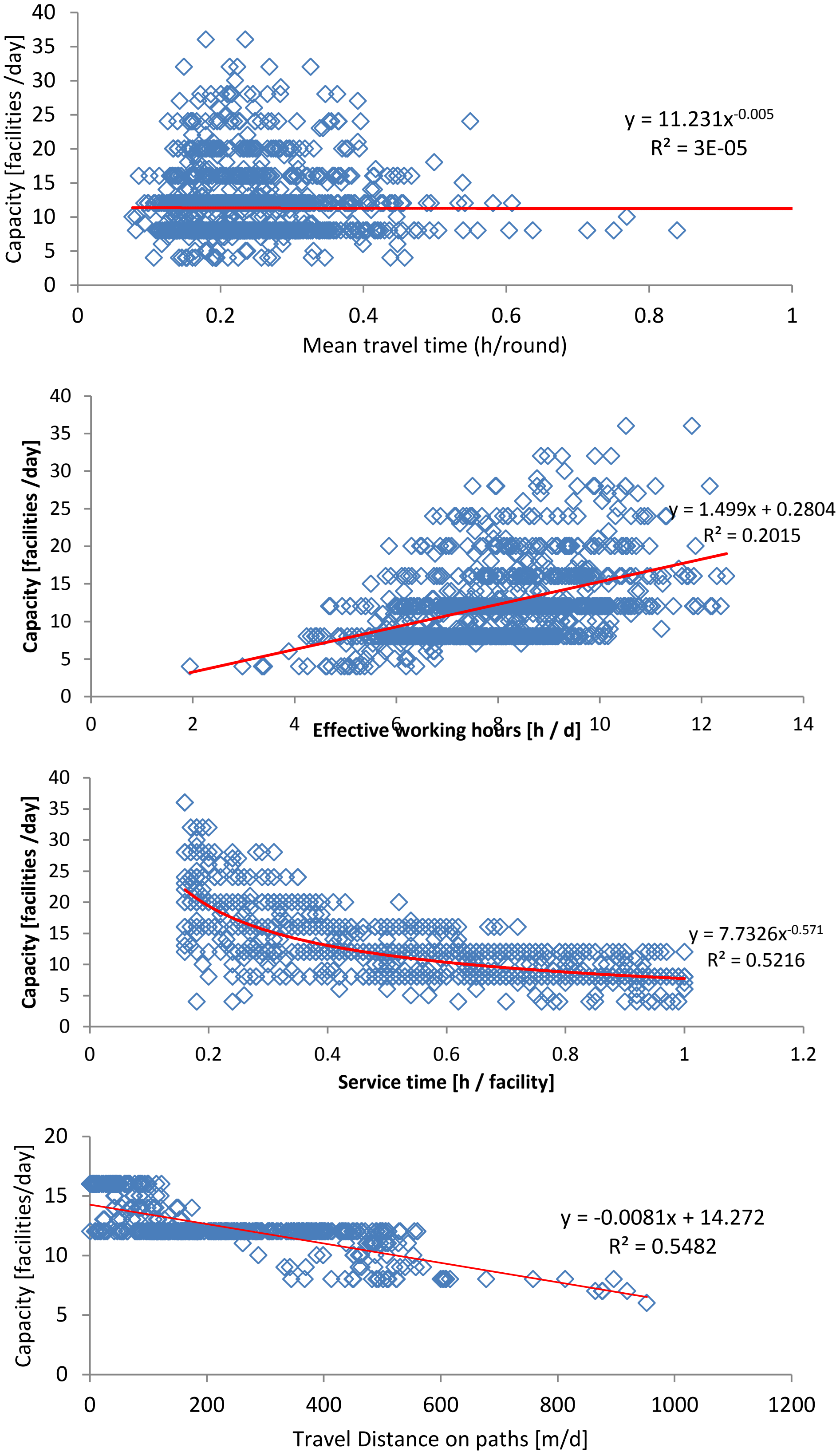


Figure 8: Regression analyses for the capacity of the small two wheel tractor

For on-path distance a closer analysis reveals that for a specific setting (Kyebando-Kisalosalo) there is a very high correlation between user density and on-path travel distance. At the same time, a comparison between results for Kyebando-Kisalosalo and Raipur shows that results are very site specific (Figure 9). Raipur's ward 68 shows a relatively low distance on paths for its relatively high user density. The clear divergence between results for Kyebando-Kisalosalo and Raipur's Ward 68 indicates that the impact of path distances should be assessed site specific and not interpolated from user density.

Results of the regression analysis indicated the parameters with the highest impact on the system's performance. The parameters route length[5], service time and work time were further used for a multivariate regression analysis between the capacity of the service system and two selected parameters. Though showing a high correlation, distance on paths was not considered for the reasons mentioned above.

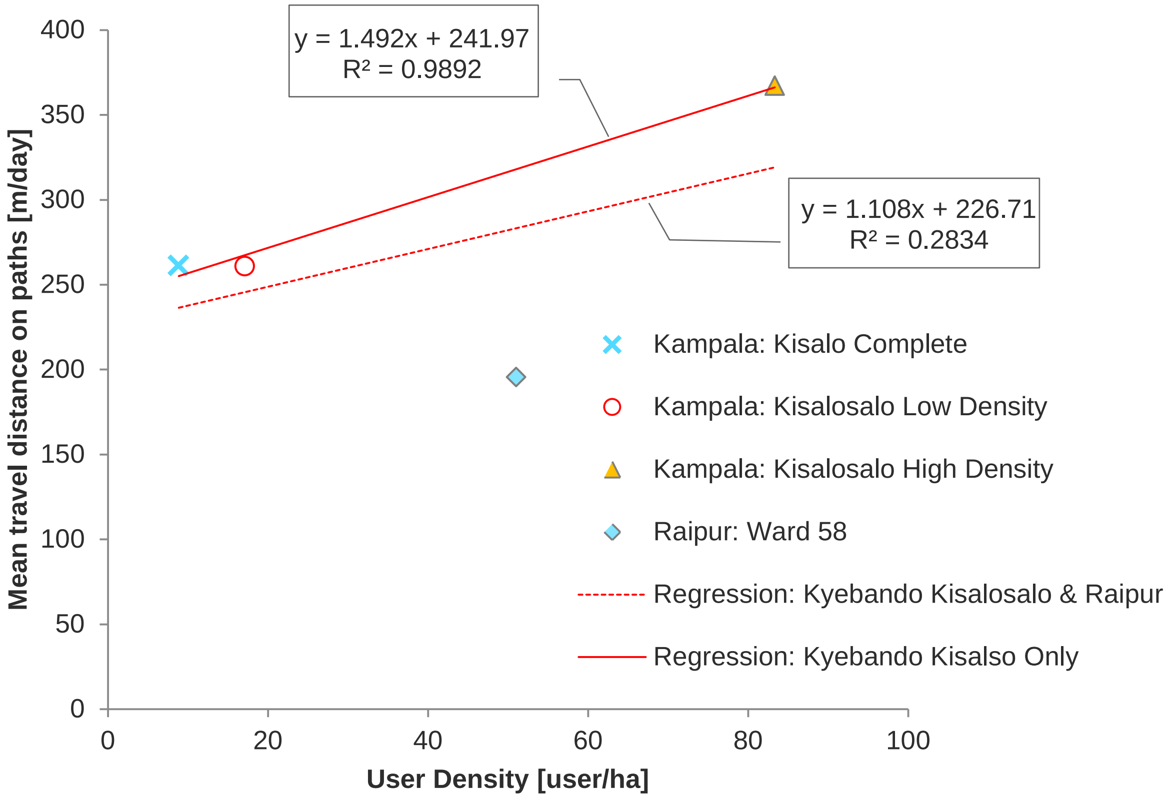


Figure 9: Comparison of on-paths travel distances vs. user density. While travel distances on footpaths are highly proportional to user density in all three settings in the Kyebando-Kisalosalo, the length of footpaths in Kyebando-Kisalosalo is relatively low.

5.3. Determination of RRP size and transport cost based on transport capacity

5.3.1. Bivariate regression

Based on basic assumptions of the business-model of EAWAG's proposal, each RRP has one service person and one vehicle available for servicing facilities. Accordingly, the capacity of the transport system will also limit the maximum number of facilities that can be connected to the RRP. An increasing number of user per RRP decreases logistic costs per user as logistic costs are divided between more users. As shown above, service time as well as effective daily working hours are key drivers for the capacity of the logistic system.

A bivariate regression was used in order to graphically represent dependencies between key parameters and the cost and capacity of the transport system. The bivariate regression will result in 3-dimensional surfaces rather than in 2 dimensional lines (Figure 10: Bivariate regression of RRP capacity vs. service time per facility and user density. A 3-D scatterplot is used to visualize raw data (simulation results) (top). A surface represents the best fitting regression (bottom).

[5] Despite its low impact on the capacity, route lengths was selected as parameter in further analysis as it is directly influencing fuel demand and accordingly energy costs.

).

Two criteria were used to select the best fitting regression: The correlation coefficient R^2 and a visual inspection of residuals. While the former allows for a general evaluation of the goodness of fit, the inspection of residuals allows for an evaluation of total deviation between regression and model results (Figure 11). For displayed sample data, the use of a cubic instead of a linear leads not only to an increase in the correlation coefficient but also to a decrease in absolute residual values (deviation between fit and model results).

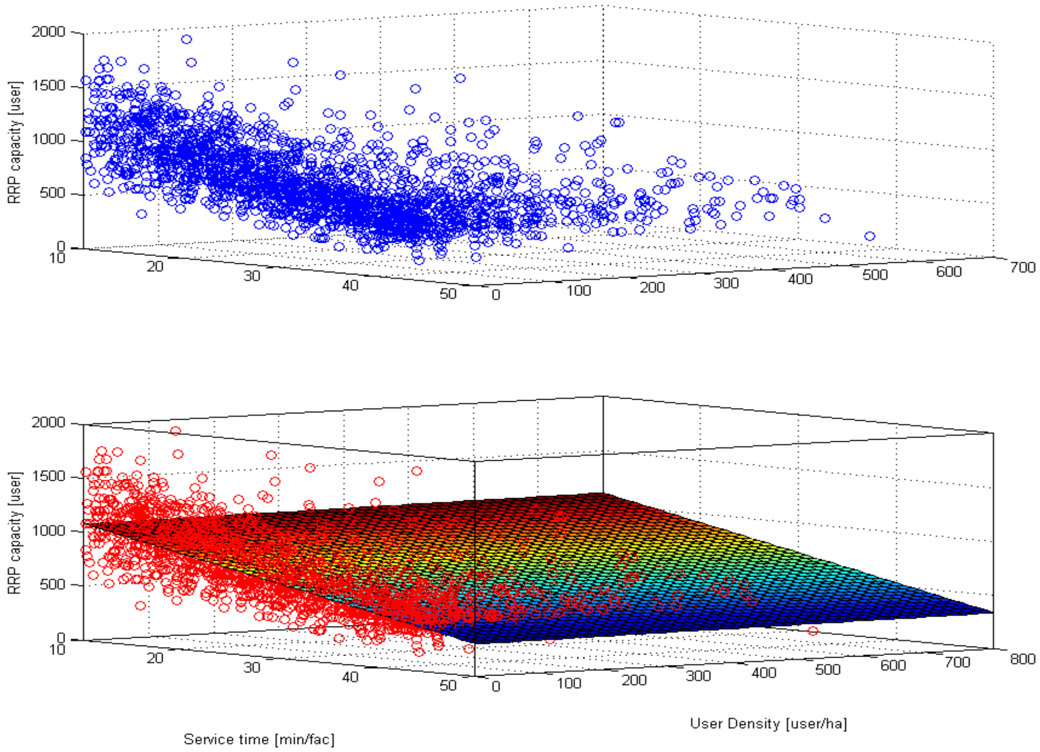


Figure 10: Bivariate regression of RRP capacity vs. service time per facility and user density. A 3-D scatterplot is used to visualize raw data (simulation results) (top). A surface represents the best fitting regression (bottom).

A) R^2 = 0.92

Linear: $f(x,y) = p_{00} + p_{10}*x + p_{01}*y$

B) R^2 = 0.98

$f(x,y) = p_{00} + p_{10}*x + p_{01}*y + p_{20}*x^2 + p_{11}*x*y + p_{02}*y^2$

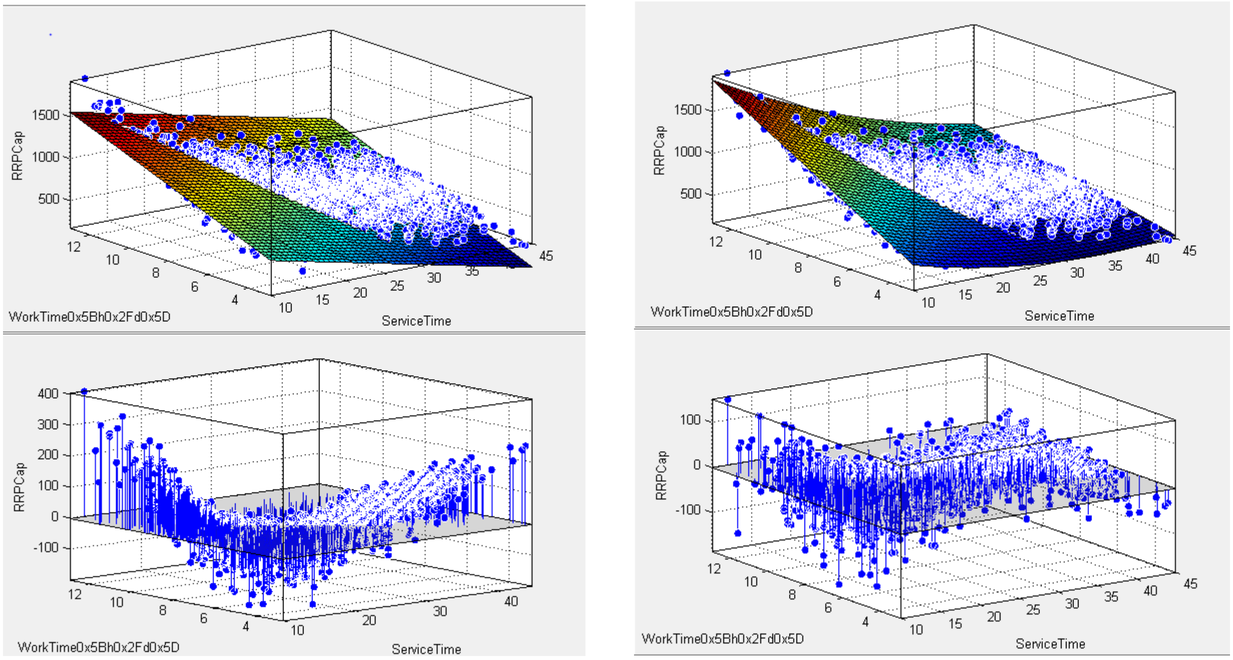


Figure 11: Comparison of the two criteria to select best fitting regression equations. Not only the R^2 value is considered but also the visual inspection of residuals indicates the goodness of fit. The cubic regression (B) decreases the deviation between simulation results and the fit in comparison to a linear regression (A). Residuals are largest for the cubic fit for very low service times, were the maximal difference simulation results and the regression equation reaches around 100 users

For better readability, the 3-dimensional plots were transformed in 2-D plots where z-values are represented by a color code (compare e.g. Figure 12 and Figure 13). The two plots allow to visually asses the dependency of RRP capacity respectively cost per user in dependence of service time, user density and work time.

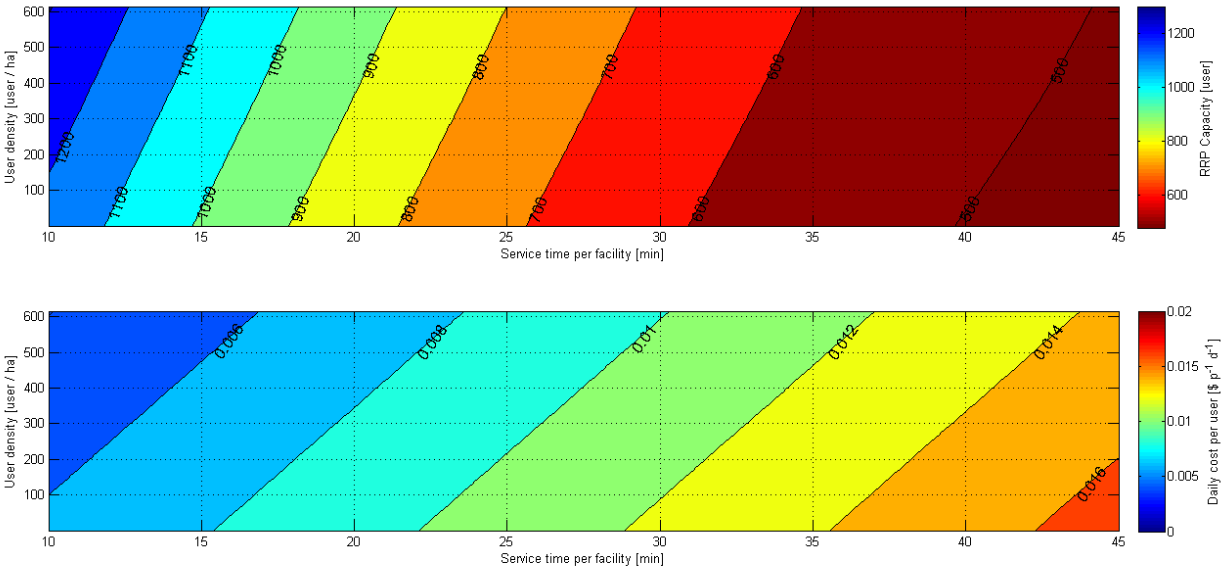


Figure 12: Capacity (top graph) and costs (bottom graph) in dependence of user density and service time per facility.

5.3.2. Results of Bivariate regression

The proposed method for data analysis allows for a straight-forward assessment of interdependencies in between system parameters as well as for estimating system parameters for concrete settings.

As to be expected from the mono-variate regression (Figure 8) there was only a weak correlation of user density (and accordingly round distance) with the total capacity of the system (Figure 12 top). Nevertheless, user density has an influence on the cost per user as increasing travel distances lead to an increase in energy costs (Figure 12, bottom). Concrete values indicate that the system is feasible within the given cost constraints (0.015 \$/p/d) for a wide range of user density and service times. The area, where logistic costs exceed 0.015 \$/p/d ($t_{service}$>40 min) are not feasible at all, because RRP capacity is below the minimum required RRP size of 500 p. Service time can only be estimated as there is no final design of collection vehicle and facility. A service time in between 20 and 30 minutes is considered feasible at this stage of the project. Due to the high correlation between service time and RRP capacity, the uncertainty concerning service time leads to a rather wide range of feasible RRP capacities. Assuming 144 user/ha (the average value for all four areas under study for 50% penetration) the RRP needs to be designed for a load of 650 to 850 users. Related logistic costs are in the range of 0.008 \$/p/d (0.8 cent/p/d) to 0.0115 \$/p/d (1.15 cent/p/d).

Work productivity (effective working hours) is a key issue for the design of the system, both in terms of cost and reliability (Figure 13). Besides the effective working hours, work productivity is likely to also influence time required to empty the facilities. Work productivity must not fall below 6 hrs/d, otherwise the transport system is not feasible within the given cost and capacity requirements. In order to reach the capacity mentioned above (650 to 850 p/RRP), an effective work time of 8 hrs/d is required. Proper incentives for workers are thus a key issue to implement an efficient service system.

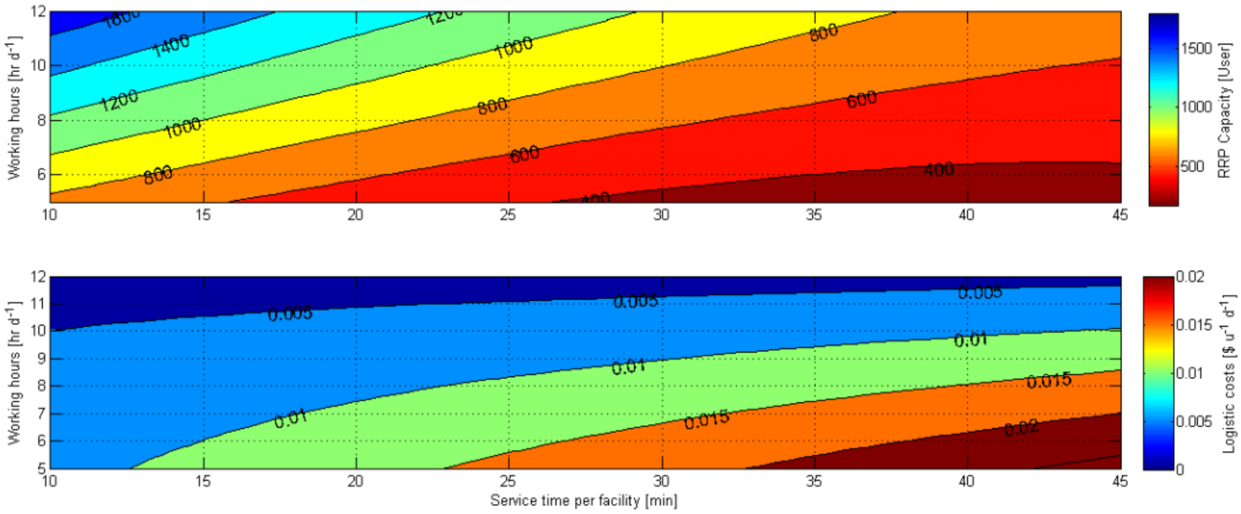


Figure 13: Capacity (top graph) and costs (bottom graph) in dependence of effective working hours and service time per facility.

5.3.3. Validation of bivariate regression

Bivariate regression and related plots allow for a much faster assessment of system characteristics than explicit modeling, especially if spatial data are not available. It was assessed how much information is lost in the process of transformation from model results. Thus how accurate contour plots are in comparison with numerical modeling. As an example, values derived from the contour plots were compared with modeling results for Raipur, Ward 68. In the model, working hours between 6 and 8 hrs/d and 20 to 30 min/facility service time were used. The contour plot indicates in an RRP capacity between 450 and 1100 users (Figure 14) for these value ranges.

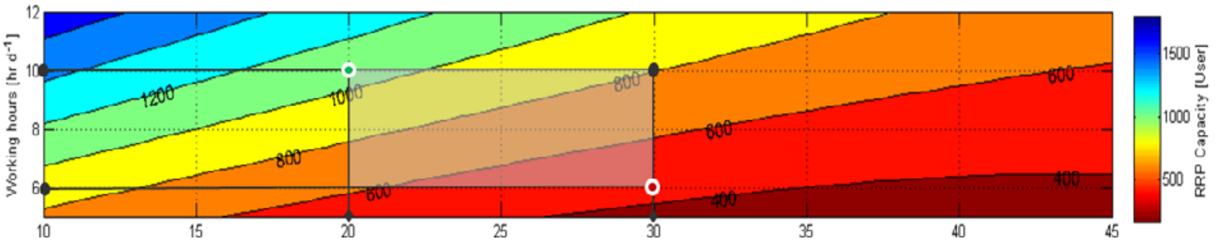


Figure 14: Estimation of ranges of capacity for RRP capacity. The green dot indicates the maximum, the red dot the minimum expected value for RRP capacity. The grey area indicates the expected range of capacity

A comparison of these results with results of the numerical model for Raipur, Ward 68 shows that the contour plots preserved most of the information. The contour plot, indicating a RRP capacity of 450 to 1100 users covers 90 % of the variability predicted by the numerical model (Figure 15).

Nevertheless, when designing a system, the worst case parameter values indicate the real capacity of the system. Designing a system to the mean parameter values will result in an insufficient capacity. Based on Figure 14, a mean capacity of 700 to 750 users could be expected. Figure 15 clarifies that such a design is not valid. The cumulative percentage indicates that the mean capacity would only be sufficient for 750 users with a probability of 60 % or that there is a 40 % probability of failure. The high probability of insufficient capacities would lead to an accumulation of un-serviced facilities over several days (Figure 16). When using the contour plots, worst case values have to be used for the parameters to derive valid design values (analog the red-dot in Figure 14). For the concrete case, only a RRP capacity of 450 users is feasible, which is below the required minimum capacity. Nevertheless, it was demonstrated with this example, that using worst-case parameter values results in an excellent performance of the system (very low failure rates, Figure 17).

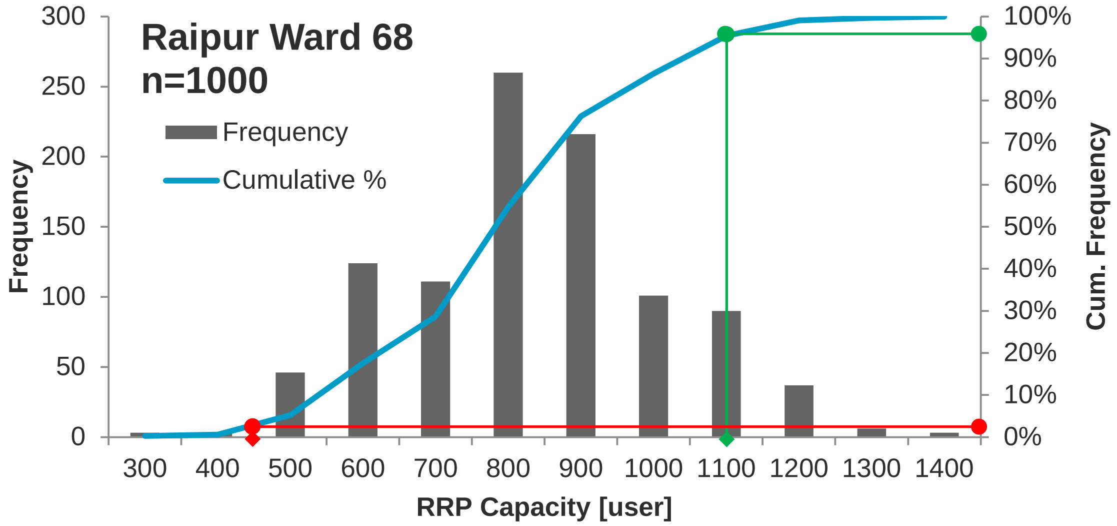


Figure 15: Comparison of results of the numerical model and values derived from a contour plot (Figure 14). The expected range of capacity based on the contour plot (between 450 and 1100, red and green square) covers around 90 percent of the variation of capacity predicted by the model (see red and green dot on the cumulative frequency axis).

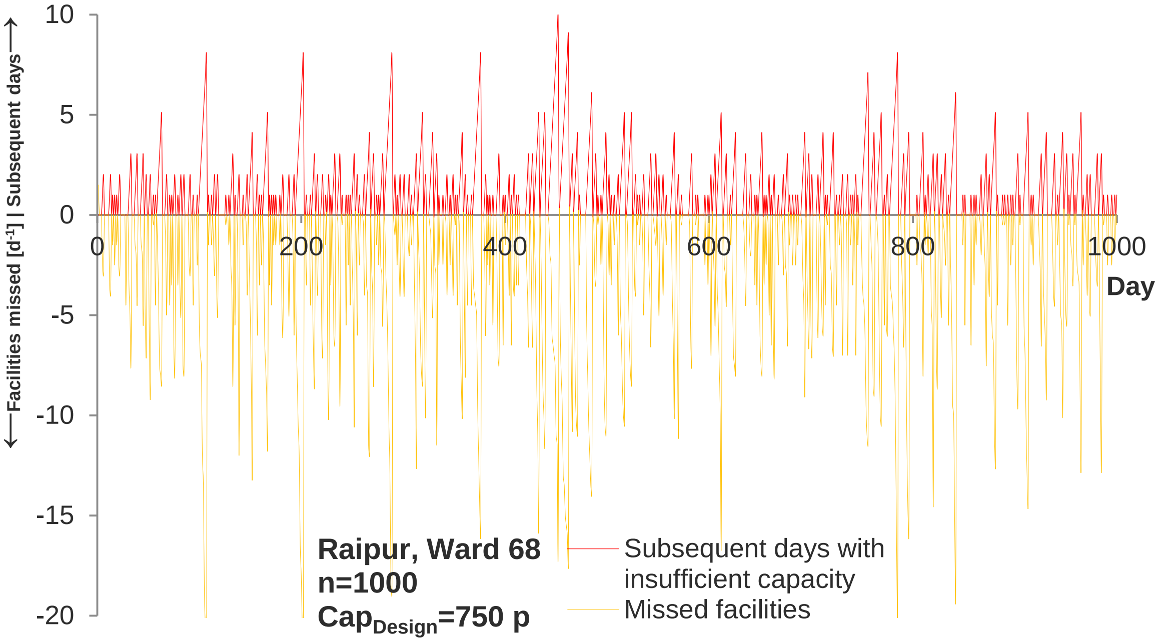


Figure 16: Time series of capacity of the transport system in Raipur, Ward 68 (20-30 min service time, 6-10 working hours / day). Designing the capacity of the RRP to the mean value derived from the contour plot (750 user, compare Figure 14) would lead to a massive accumulation of under-capacity (yellow: number of unserviced facilities per day) over extended time-spans (red: maximum service delay [days]).

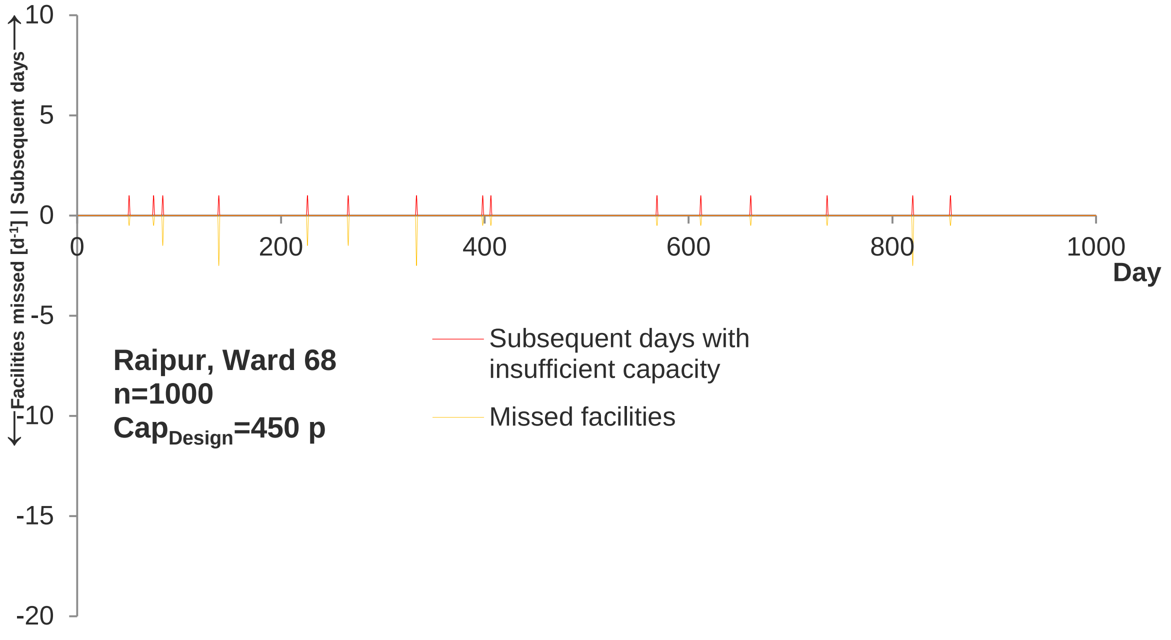


Figure 17: Time series of capacity of the transport system in Raipur, Ward 68 (20-30 min service time, 6-10 working hours / day). Designing the RRP capacity based on the contour plots with worst-case values for working hours and service time leads to an excellent failure tolerance of the system. Maximal service delay (red) is one day. Facilities are designed to offer sufficient additional service capacity for at least one day, so there is no risk for overflows.

5.3.4. Application of the proposed method to concrete settings

The proposed method based on the contour plots rather than on explicit modeling allows for a direct visual comparison of system capacity for concrete settings. In this example 3 Parameters are considered: user density, working hours and service time. For all four settings a penetration of 50% is assumed. There are most likely differences in work productivity in between Africa and India. Residents from the Kyenbando-Kisalosalo area considered 8 working hours per day realistic (C. Lüthi, personal communication). For India literature indicates up to 9.5 working hours for similar services (female waste-pickers in informal settlements) [6]. Service time is assumed to be 20 minutes for all settings. The assumptions are compared in Table 3.

Table 3: Assumptions for the graphical comparison of transport cost and capacity in the four different settings.

SETTING	POP. DENSITY [P HA^{-1}]	USER DENSITY (50% PENETRATION) [USER HA^{-1}]	WORKING HOURS [HRS D^{-1}]	SERVICE TIME [MIN FACILITIY^{-1}]
KYEBANDO KISALOSALO LD	128	64	8	20
KYEBANDO KISALOSALO HD	520	260	8	20
KYEBANDO KISALOSALO COMPL.	189	95	8	20
RAIPUR, WARD 68	320	160	9.5	20

As service time is constant, a direct regression of user density versus work time and cost respectively capacity is used (Figure 18). Results visualize how local differences in user density and work productivity result in different RRP capacities and transport costs. Especially because of the long working hours in Raipur, the RRP capacity can be increased for up to 1000 users at costs of around 0.6 cent/user/day. For the settings in Kyebando-Kisalosalo, there is no major difference in cost or capacity related to their different user densities. This corroborates that the proposed sanitation system is feasible even with low user densities.

[6] Huysman, M. (1994). Waste picking as a survival strategy for women in Indian cities. *Environment and Urbanization* 1994 6: 155.

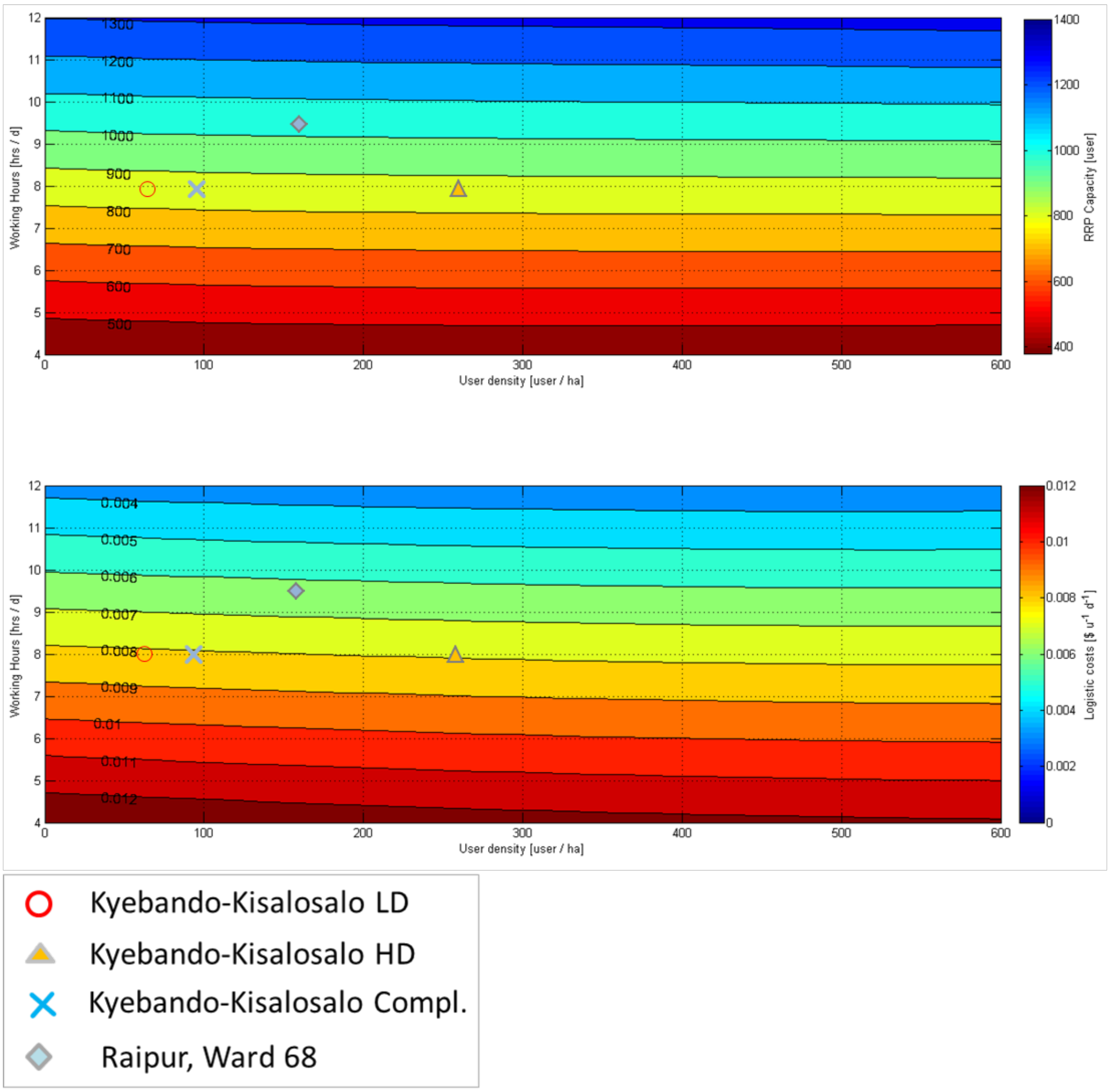


Figure 18: Visual assessment of transport cost and capacity based on the user density and the working hours for the four selected settlements.

5.3.5. Assessment of single stance sanitation systems

As an additional scenario, the impact of single stance facilities on system costs and capacity was evaluated. The introduction of single stance toilets has a major impact on the system. So far, only the worst-case scenario of a single-stance-only system was evaluated, but of course also intermediate solutions are conceivable with a mixture of single and double stance toilets. According to model results, the introduction of single-stance facilities has a major impact on the system. The RRP capacity is highly decreased in comparison to the double-stance-only system. Logistics is feasible if working hours above 8 hrs/day and if a service time of below 15 minutes per single stance facility can be reached.

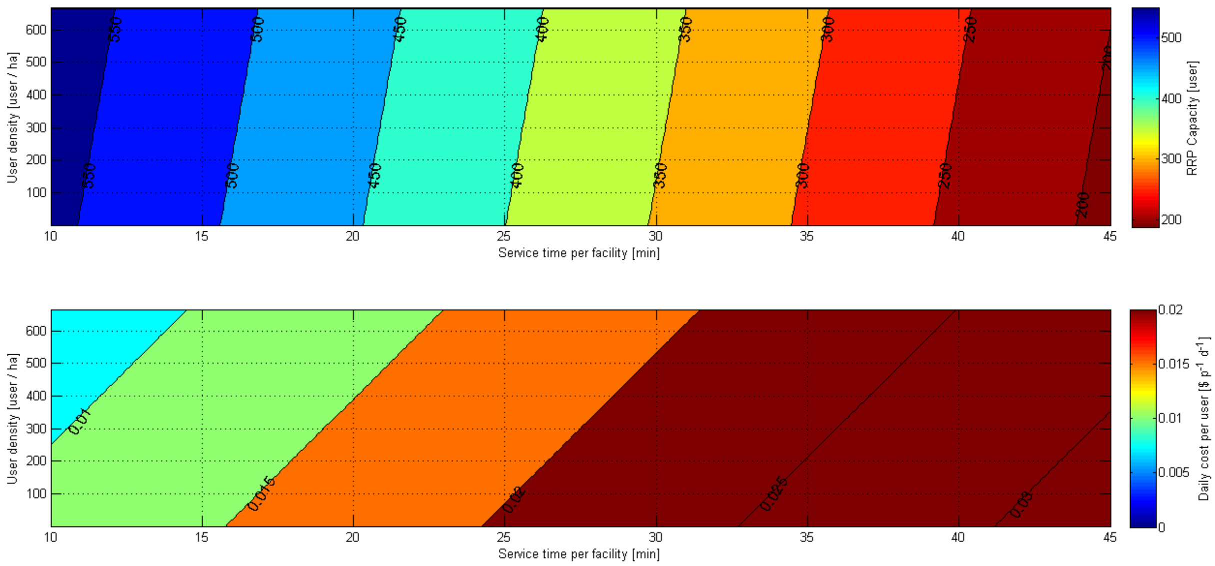


Figure 19: RRP capacity (top) and cost per user (bottom) for a sanitation system with 100 % single stance facilities in dependence of service time and user density.

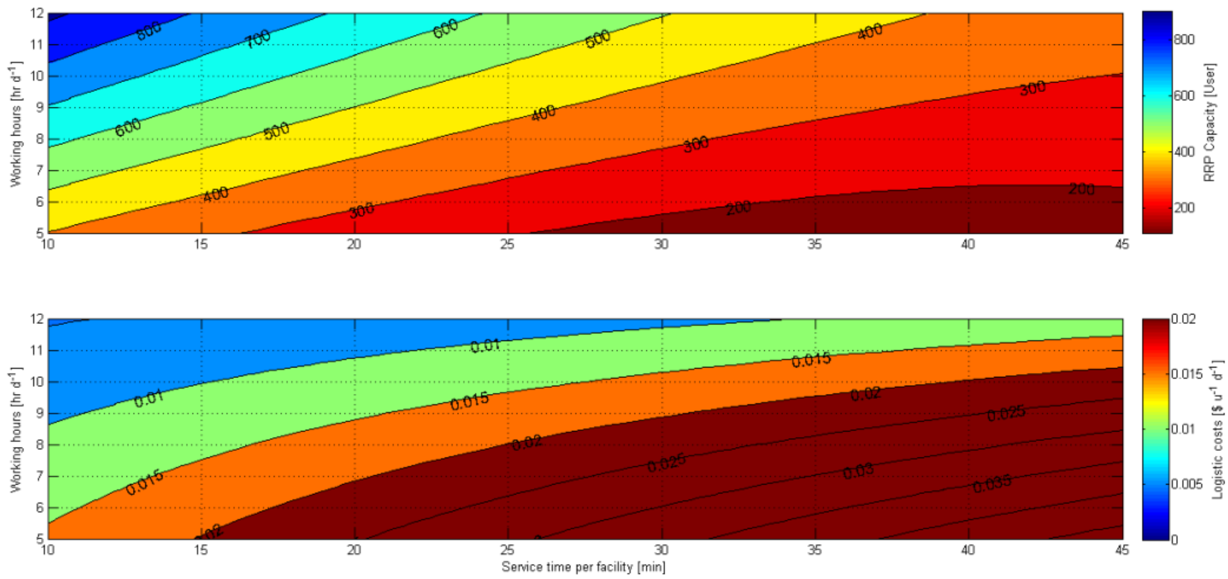


Figure 20: RRP capacity (top) and cost per user (bottom) for a sanitation system with 100 % single stance facilities in dependence of service time and working hours.

6. Conclusion

The transport system is a key element of the sanitation system proposed by EAWAG. This work proves that the required transport system is feasible within the given cost constraints and provides a high reliability of service. Transport logistics is feasible for facilities shared between twenty persons (double stances) as well as for facilities shared between 10 persons (single stances).

The numerical model in combination with the Monte Carlo Analysis allows to quantify cost and capacity of the service system. The proposed small vehicle offers the required capacity at competitive costs. The contextualization of the proposed service system is possible with the presented GIS methodology.

With a regression analysis, key impact factors on the system's performance were identified and their impact was quantified. It was shown that the spatial setup of the system is only of minor importance for the system performance. Accordingly, the proposed transport system can be evaluated without a prior GIS analysis. Work productivity and a design of toilet facilities that allows for rapid emptying were identified as key drivers for system performance.

Based on simulation results, generic design diagrams were developed for system based on results of a bi-variate regression. Comparison of results from explicit numerical modeling with design diagrams indicates a very high accordance between the numerical model and design diagrams.

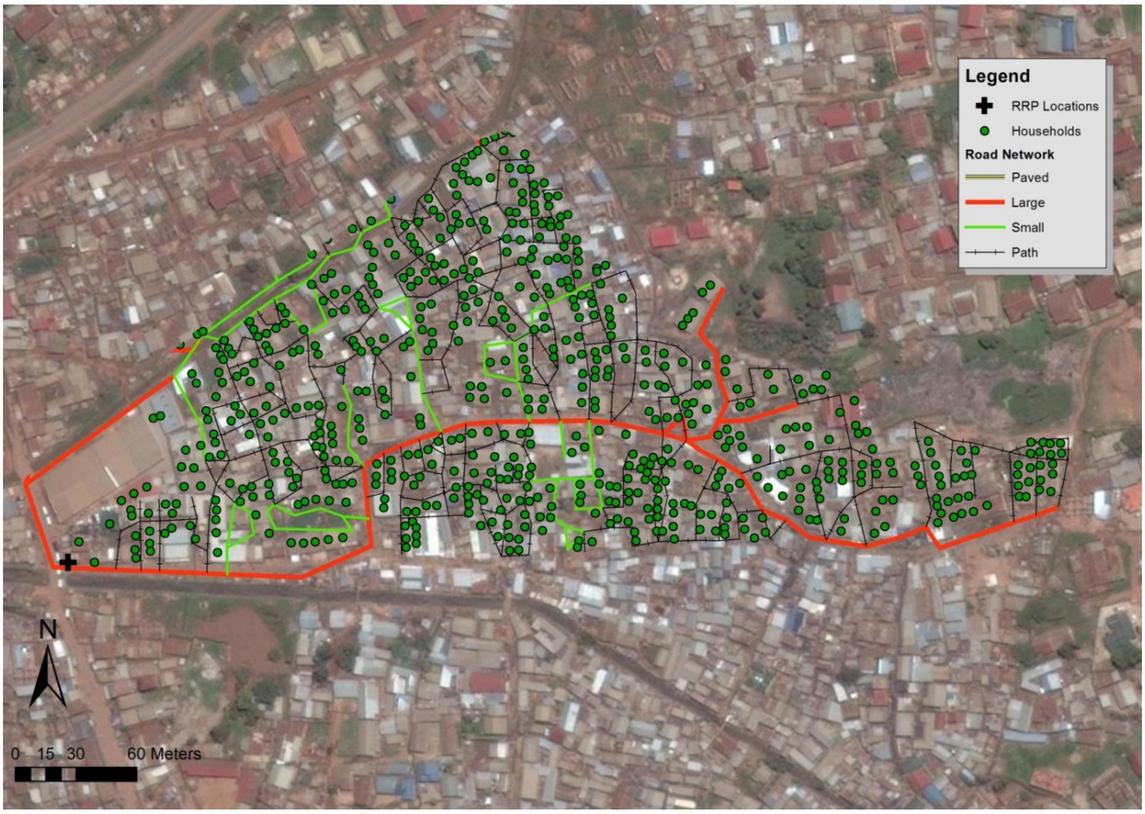
Though the method was developed and applied to a concrete logistics problem it has wider applicability. The proposed GIS analysis method allows for a contextualization of the numerical model. Where vehicle speed is relatively low in comparison to travel distances, a quantification of required travel distances will be of major importance. The numerical model allows for a detailed assessment of transport performance and consideration of uncertainties via the integrated Monte Carlo Analysis.

So far, there is no method available for quantitative assessment or optimization of transport based services in informal settlements. The presented methodology can be used for a wide range of analysis and optimization tasks for transport based services in informal settlements.

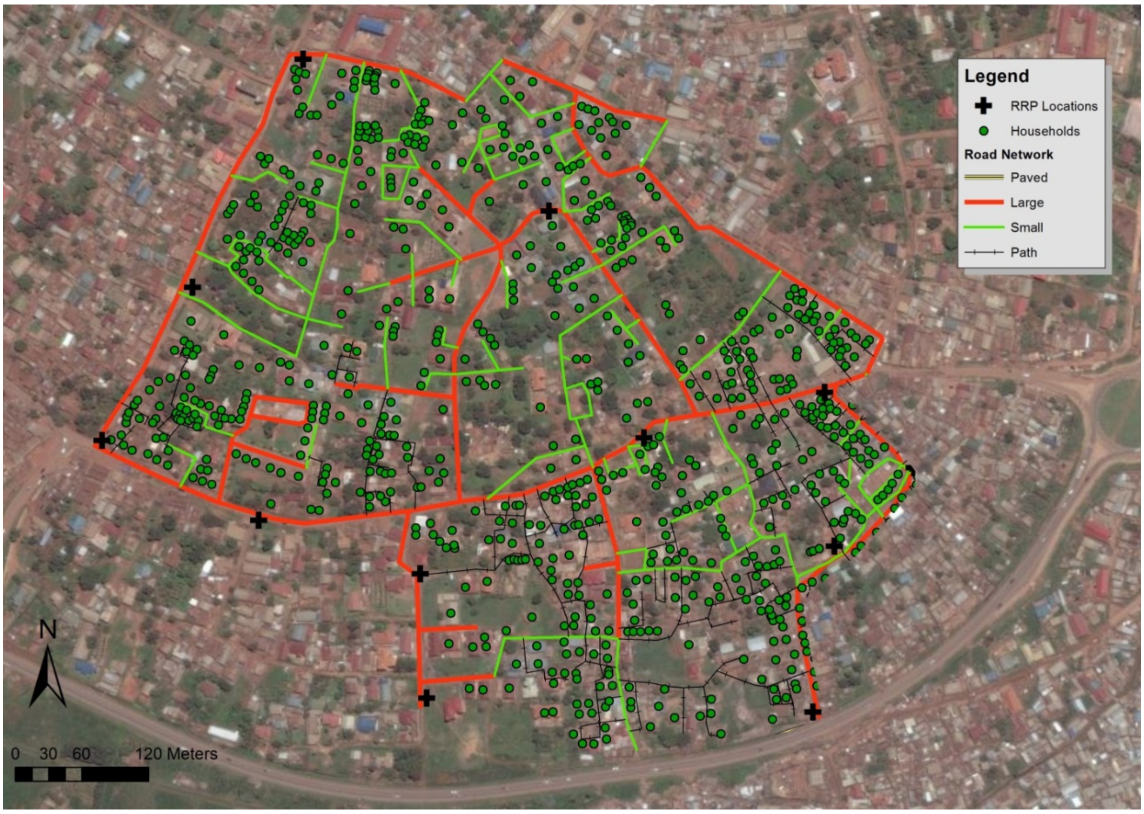
A. Appendix

A 1. Maps of the areas under study

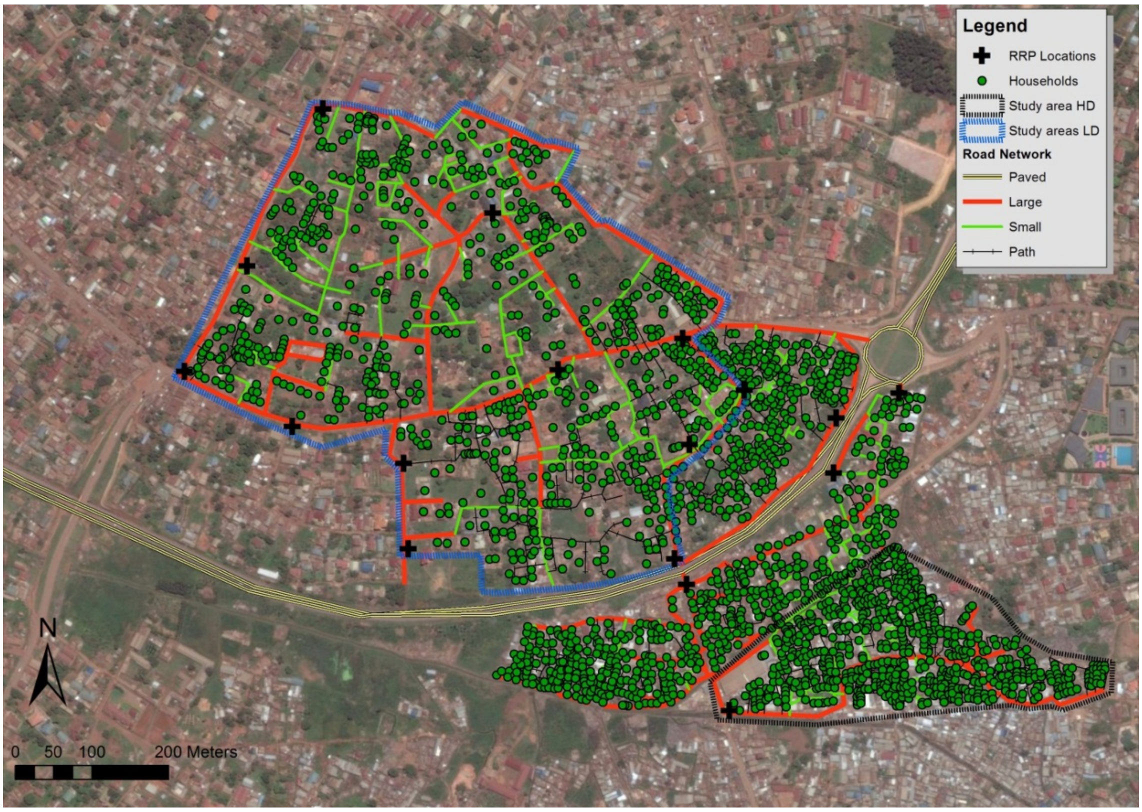
KYEBANDO KISALOSALO - HIGH DENSITY



KYEBANDO KISALOSALO - LOW DENSITY



KYEBANDO KISALOSALO - COMPLETE



RAIPUR WARD NO. 68

Figure 21: Satellite images (Bing Maps, 2012) of the 4 areas under study and digitalized features.

A 2. Comparison of scenario A and B

Different growth strategies for the RRP-Organization are shown in Figure 22. Initially, 500 users (25 facilities) are connected to a central RRP, whose position is selected out of several candidates using the GIS. Figure 22 shows the expansion of 500 to 1500 users under scenario A and B. Scenario A allows for dynamic repositioning of the RRPs which decreases required travel distances. Scenario B assumes a fixed RRP position: Either new RRPs are constructed alongside the existing ones or the capacity of the initial RRP is increased.

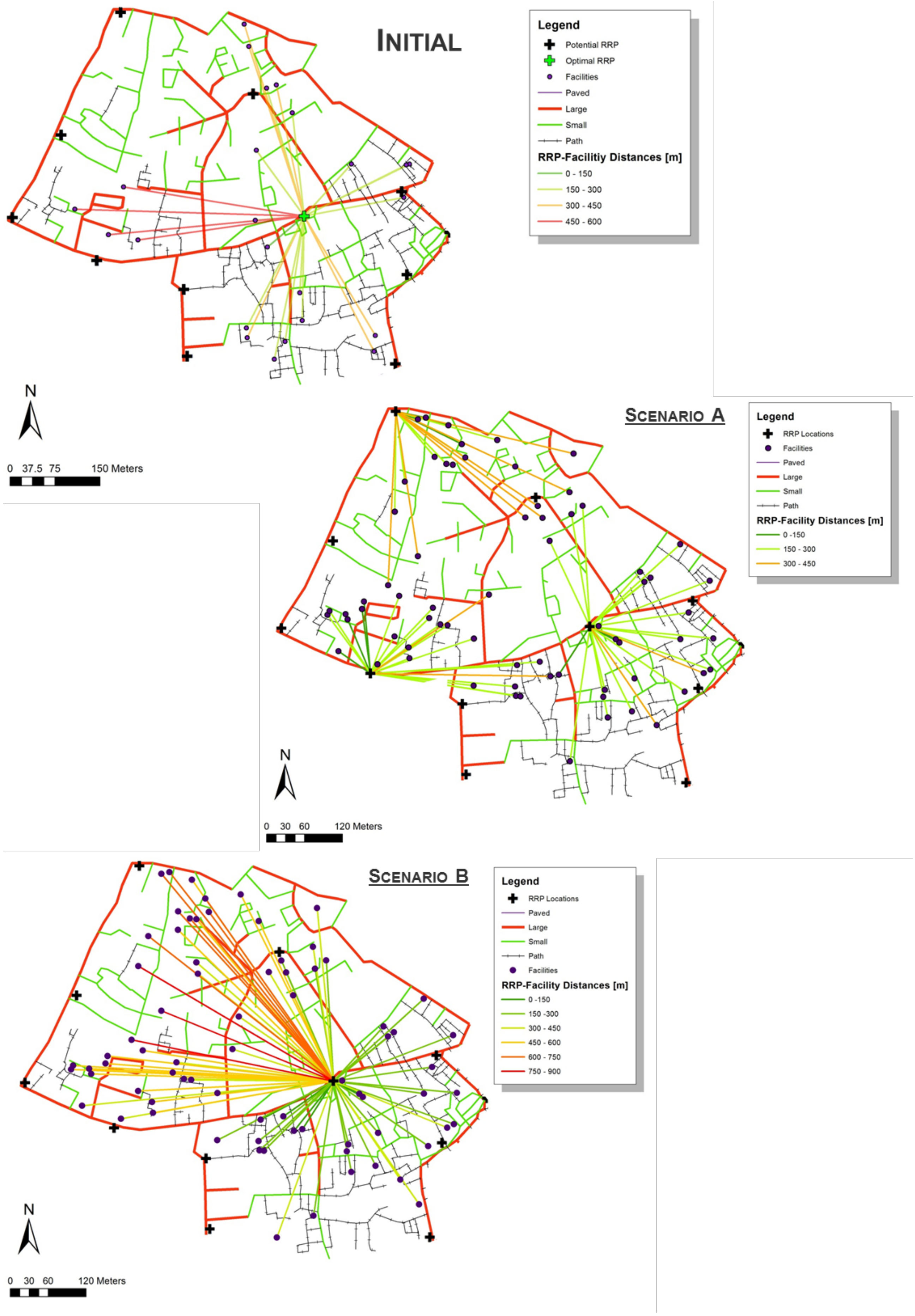


Figure 22: Initial state and growth strategies for the RRP-Organization from 1 RRP (25 facilities) to 3 RRPs (75 facilities) (based on the Kyebando Kisalosalo - low density area).

A 3. Calculation of RRP-facility distances under scenario A

So far, there is no tool available for formation of facility clusters in ArcGIS based on on-road distances (clustering is so far only possible based on air-line distances). Thus, the decrease of RRP-facility distances under scenario A was simulated based on the following geometric assumptions: The catchment is described by a rectangle. Assuming a centrally located RRP, the maximum distance from the RRP is described by

$$c = \sqrt{\frac{a^2}{4} + \frac{b^2}{4}}.$$

If the area is divided as soon as a new RRP is constructed, the original area will be cut in two parts, thus either a or b is divided by 2 (Figure 12). The new maximum distance in the area is only the fraction r of the initial maximum distance. r can be described analytically

$$r = \frac{0.5 * \sqrt{a^2 + 4b^2}}{\sqrt{a^2 + b^2}}$$

The numerical value of r is dependent on the length ratio of a and b. r was calculated for 10^5 random combinations of a and b. the resulting average of r was 0.77. If $\bar{d}_{RRP-Fac,0}$ is the mean RRP-facility for the first RRP, the mean RRP-facility distance for the i^{th} next RRP is given by

$$\bar{d}_{RRP-Fac,i} = \bar{d}_{RRP-Fac,0} * r^i$$

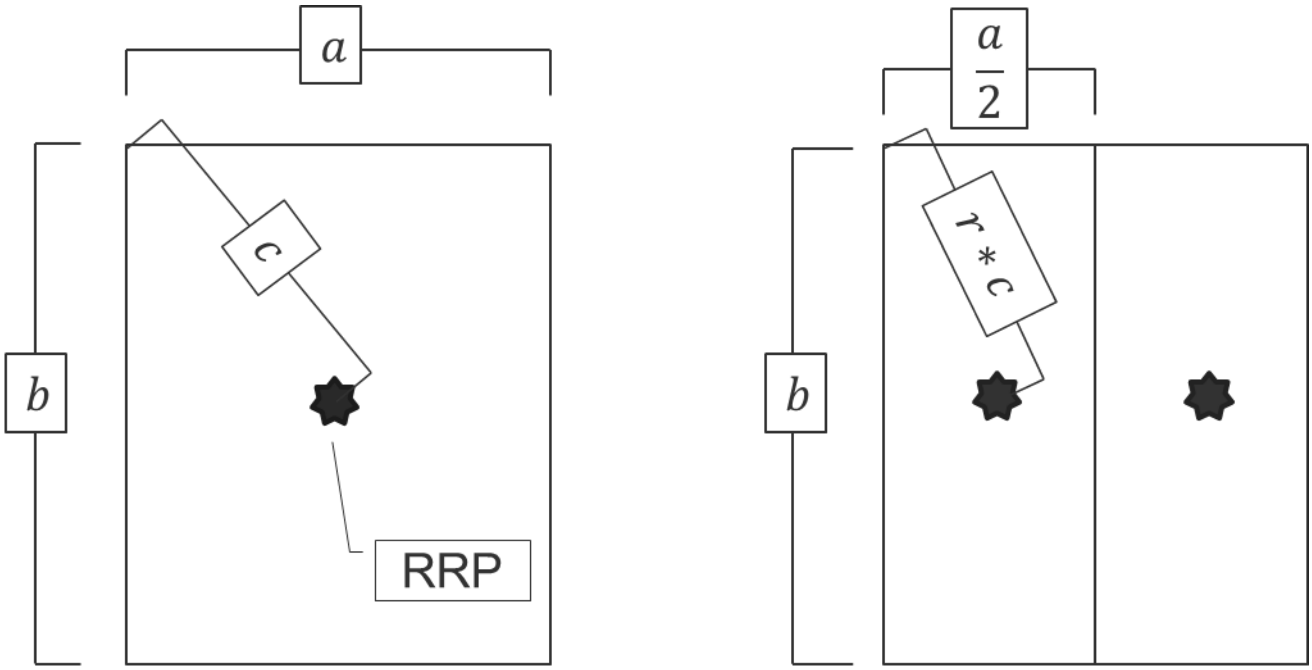


Figure 23: Geometric assumptions for RRP-Facility distances during system growth under scenario A.