BSc Thesis Applied Mathematics

Determining water well locations

Anna Jannie Gesink

Supervisors: Jasper de Jong and Theo Kleinendorst

January 23, 2025

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

I would like to thank Theo Kleinendorst and everyone at Acacia Water for making this thesis possible for me. I appreciate your introduction to the subject of hydrology and water wells, a whole new world for me. You have helped me greatly by thinking along, your suggestions and our discussions. We may have had to change the plan a couple of times but I am happy with what we ended up with. I have had a great time at the Acacia office and am very grateful for how everyone there has welcomed me.

I would like to thank Jasper de Jong for supervising this thesis. Our weekly meetings have been of great help in developing my mathematical model and shaping the assignment. I am thankful for our great discussions on the matter and for guiding me when I got stuck. After our meetings I was always full of new ideas to try which gave a great enthusiasm boost to continue working.

Lastly I would like to thank my friends, family and house mates for their support and continually asking how it is going, which made me pause for a second and appreciate what I had done. Overall this thesis has been a good experience in which I learned a lot, thanks to all of you.

Determining water well locations

Anna Gesink*

January 23, 2025

Abstract

In drought-struck Ethiopia water wells are a vital part of a safe drinking water system. Currently, determining new water well locations is done by hand based on suitability maps made by Acacia Water. In this paper we describe two methods to determine approximate water well locations. First we explore an approximation method where one well is placed such that drilling and pipeline costs are minimal. We find configurations to connect 50%, 70% and 100% of the population to this well. Secondly, we explore an approximation method that places multiple wells such that costs are minimal, also to connect 50%, 70% and 100% of the population to a well. This is done by regarding the problem as an Uncapacitated Facility Location Problem (UFLP). We saw that for connecting 100% of the population it is better to place multiple wells, for connecting 50% and 70% of the population this depends on the situation and it is recommended to run both algorithms and choose the cheapest option.

Keywords: water wells, Uncapacitated Facility Location Problem, UFLP, suitability maps

*Email: a.j.gesink@student.utwente.nl

Contents

1 Introduction

In drought-struck countries, such as Ethiopia, the need for safe and reliable water sources is of great importance. Although a great part of Ethiopia has a greater annual precipitation than the Netherlands, the rain is seasonal and drinking surface water brings risk of disease. Therefore, ground water is the preferred source of drinking water [3], [12], [14].

To access ground water new water wells need to be dug. At Acacia Water [1] so called Suitability Maps are made. These maps show where it is more and less likely for a digging to lead to a successful well and hence which areas are most suitable for digging.

These maps are made by combining six different map layers, being Permeability, Recharge, Lineaments, Land cover, Stream proximity and the Topographic Wetness Index (TWI). The TWI is a measure of how wet an area is. For this it takes the water supply from up-slope catchment area's and the down slope water drainage into account [11].

In a workshop with experts each of these six layers is given a weight. The weights sum up to one and each map layer's values are scaled between 0 and 1. Then for each pixel of each input layer the value is multiplied with its respective weight and with all of these summed the suitability is obtained, scaled between 0 and 1. An example of such a map can be seen in Figure 1.

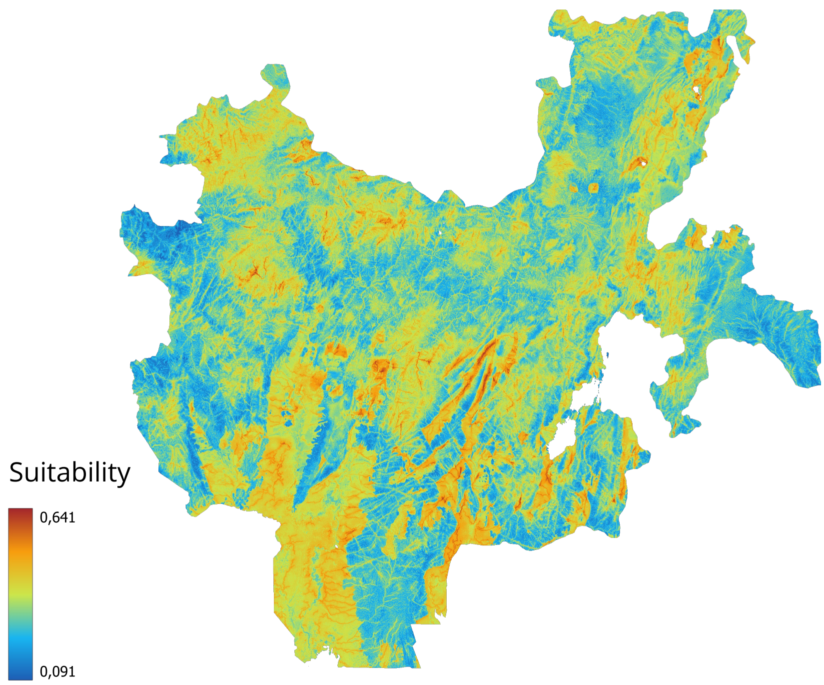


Figure 1: Suitability map of SNNP Region in Ethiopia

These suitability maps are used to find areas where a successful well digging is likely. In these area's then a detail study with field work and discussion with local stakeholders is done, after which the final well location is decided.

For these management and planning purposes it is desired to know how these suitability scores relate to the probability of a successful well. In this report we will see that with the currently available data no such direct relation can be found, and we will explore how digging locations could be determined if we assume the suitability to be equal to the probability. This is done by minimizing the total cost of digging and pipelines. Two approximation algorithms are implemented. One that finds one location and connects it to a given percentage of the population, and one that finds an arbitrary number of locations such that the total costs are minimal. We will call this the multiple well algorithm, for which post-processing is applied to obtain the costs when asked to connect to only a given percentage of the population. For both of these two algorithms finding the exact minimal costs is not possible, and hence they will be approximated.

As it is unknown what the costs in euros of a well are we will assume this cost to be C and calculate the costs of the pipeline expressed in C. For example, if for the cost of one well ($1C$) one can buy 5000 meters of pipeline, one meter of pipeline costs $\frac{1}{5000}C$.

(A) SNNP Region in Ethiopia [5]

(B) Ethiopia [4]

Figure 2

This study will focus on the former Southern Nations, Nationalities and People Region (SNNP Region). This region lies in the South West of Ethiopia, and has recently (2023) been split up in four new regions. The geographical placement of this region can be seen in Figure 2.

We find well locations for each of ten assigned priority area's, each of which is a Woreda, comparable to a municipality. Since each Woreda is governed locally and collaboration is complicated it has been decided to determine well locations for each Woreda separately and independently of the other Woreda's.

It is important to note that even if water is found, it is not necessarily good water. The water quality wildly varies and even salt water might be dug up. Since there are currently no satisfactory maps that show the water quality in an area we could not take this into account for the maps made in this report.

In Figure 3 we can see the ten Woreda's and their respective names and colours can be seen in Table 1:

Woreda	Colour	Area(km^2)	Woreda	Colour	Area(km^2)
Basketo SP	bright green	411	Menit Shasha	light blue	1602
Bilate Zuria	pink	272	Salamago	red	4572
Kucha	dark purple	987	South Ari	yellow	1638
Maji	light purple	4658	Uba Debre Tsehay	orange	754
Melekoza	dark blue	943	Wulbareg	dark green	398

TABLE 1: Woreda's and their respective colours

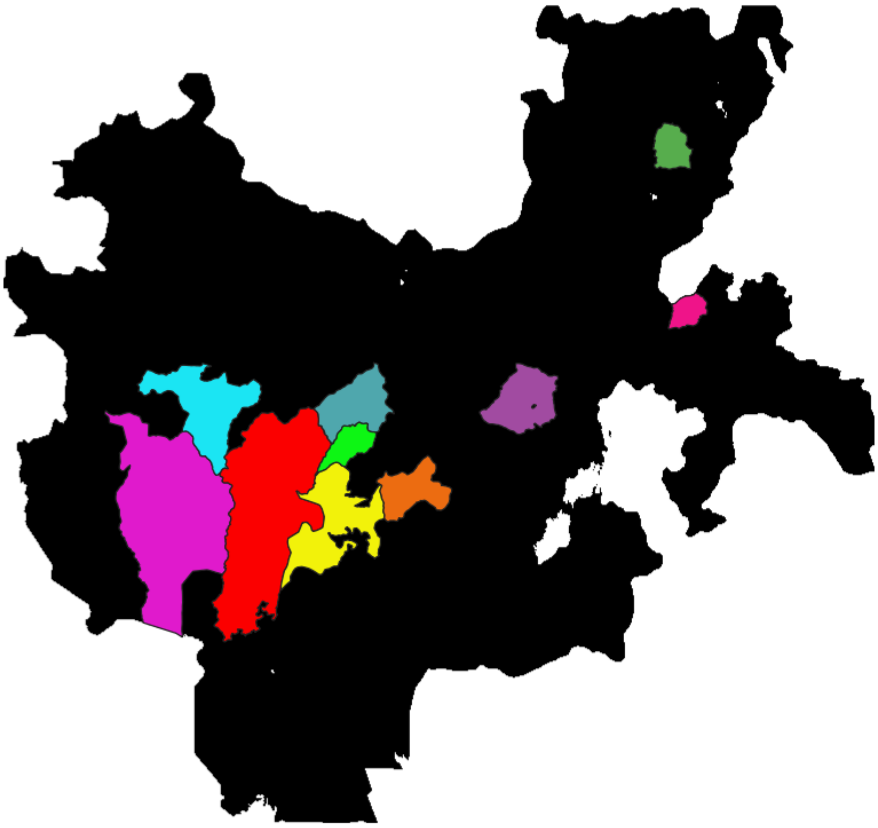


FIGURE 3: Project priority area's of SNNP Region in Ethiopia

2 Theoretical Background

The problem to place multiple wells may be regarded as a Facility Location Problem. In the Facility Location Problem the placement of facilities, such as warehouses to supply stores, in order to minimize the cost of satisfying the demand in an area is dealt with. Generally there are fixed costs for opening facilities and connecting them with the user. In the warehouse case, these might be the costs of building the warehouse and the costs for trucking the supplies to the stores (the users). In the Facility Location Problem we may differentiate between the capacitated (CFLP) and uncapacitated (UFLP) case. In the capacitated case there is a maximum of users a facility can supply to, whereas this is not true for the uncapacitated case [7].

Formally, we may define the UFLP as follows [10]:

In the Facility Location Problem, we have a set F of n_f facilities and a set C of n_c users. For every facility $i \in F$, a nonnegative number f_i is given as the opening cost of facility i. Furthermore, for every facility $i \in F$ and user $j \in C$, we have a connection cost c_{ij} between facility i and user j. The objective is to open a subset of the facilities in F, and connect each user to an open facility so that the total cost is minimized.

The UFLP has been shown to be NP-hard [7]. Hence if we assume $P \neq NP$, which is generally believed to be true [6], no solution can be found in polynomial time. We will therefore want to find approximation algorithms to solve the UFLP. We may compare different algorithms by their approximation factor, that is, the ratio between the costs found by the algorithm and the costs in the optimal (theoretical) solution.

In [15] the authors found a polynomial-time algorithm for the UFLP that finds a solution with cost within a factor of 3.16 of the optimal. This is the first constant performance guarantee known for this problem. Gradually, better approximation factors have been found. An approximation factor of 1.61 was found by [10] and a factor 1.5 was found in [2], which was then improved to 1.488 by [13] by combining the two aforementioned papers. The approximation factor 1.488 is the best known thus far.

In [9] the authors prove that no better approximation than 1.463 can be found, assuming $NP \notin DTIME[n^{O(\log\log n)}]$. Here $DTIME[f(n)]$ is the class of decision problems that can be solved in deterministic time $O(f(n))$, where n is the input size.

As NP is the class of problems that can be verified in polynomial time and $DTIME[f(n)]$ is the class of problems that can be solved in polynomial time we may interpret the assumption $NP \notin DTIME[n^{O(\log\log n)}]$ to mean that there exist problems that can be verified in polynomial time, but cannot be solved in $n^{O(\log\log n)}$.

3 Determining probabilities

The first part of this project consist of an attempt to relate a (measure of) probability and suitability. The following analysis has been performed in QGIS (Version 3.34.12), an open source software for visualization and analysis of spatial data.

Two input layers have been used: a pixel map with suitability per 100 by 100 meters, scaled between 0 and 1 (see Figure 1) and a map containing points representing water wells, with a total of 9262 water wells, of which the yield in L/s is known for 1433 wells, see Figure 4.

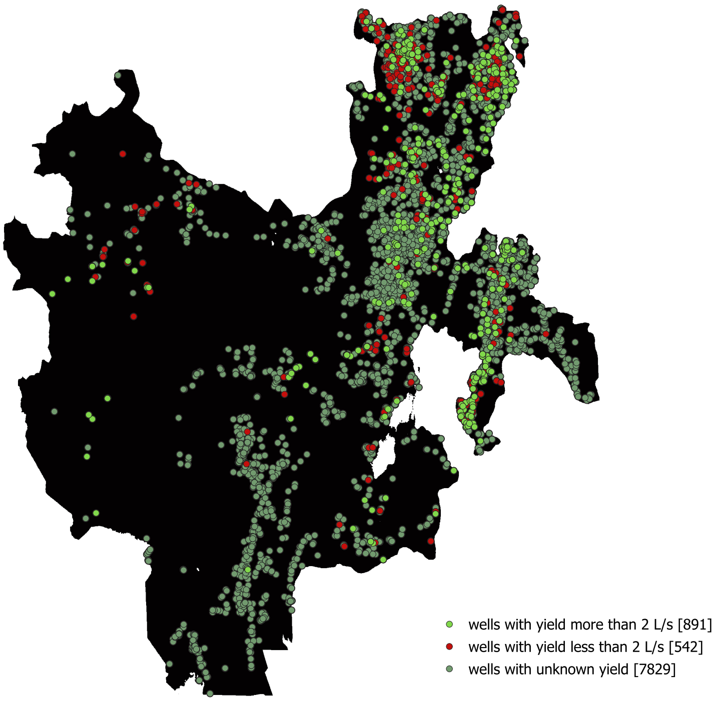


FIGURE 4: Well distribution of the SNNP Region

To be able to compare the point map with wells and pixel map with suitability all maps have been resampled to a 5 by 5 km grid. This resulted in two maps, where each 5 by 5 km square represents the average suitability of the pixels in that square and a count of the number of wells in that square.

Since the wells are quite sparse in some regions a 5 by 5 km grid has been chosen. It was decided that this was a good balance between relating population to the available wells (a bigger grid) and retaining spatial information on population and suitability (a small grid).

It has been considered to correct the number of wells in each 5 by 5 km square for the population, since we may assume that more drillings are done where the population is bigger. Hence a certain area with plenty of wells will appear very suitable for drilling, when there really is only a large number of wells because a large number of attempts has been made since the population is large. However, no clear relation between number of wells and population per 5 by 5 km square can be obtained, as can be seen in Figure 5 and is further confirmed by the R-squared value of 0.2, which is much less than 1. Hence no proper correction for the number of wells and the population was found.

For a probability one usually divides the number of successes by the total samples. In this case however, we have no data available that states the number of failed wells or

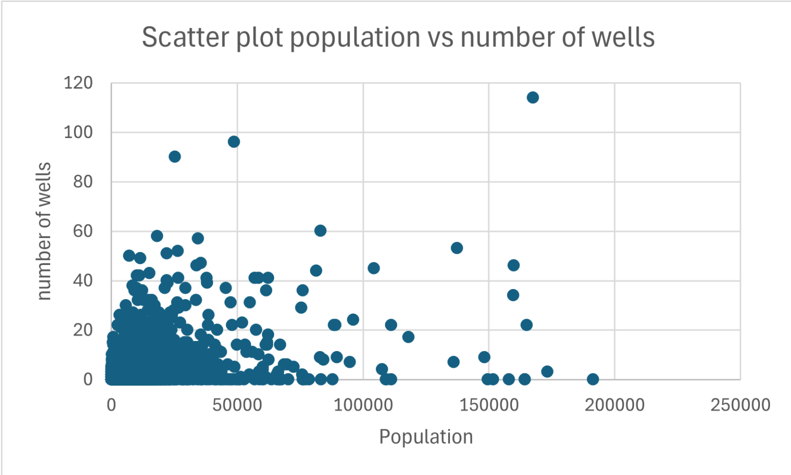


FIGURE 5: Scatter plot of population vs number of wells in each 5 by 5 km square

drillings. To provide a measure of unsuccessfulness we will take all wells with yield less than 2 L/s as unsuccessful. There are situations in which a well with less than 2 L/s is still sufficient, for example as a hand pump. We will assume here, however, that the yield is a measure of how successful the well is in terms of digging a new one at that location.

The number of wells in each 5 by 5 km square can then be subdivided in three classes. Yield more than 2 L/s, the 'successful wells' (891 wells), yield less than 2 L/s, the 'unsuccessful wells' (542 wells) and no known yield (7829 wells). We will assume that of the wells with no known yield the same ratio between successful and unsuccessful applies as for the existing wells for which yield is known, that is, a $\frac{891}{891+542} = 0.62$ success rate.

With this data a (measure of) probability has been calculated for each square as follows:

$$\text{Probability} = \frac{(\#\ \text{yield} \geq 2) + (\#\ \text{yield unknown}) * 0.62}{(\#\ \text{yield} \geq 2) + (\#\ \text{yield} < 2) + (\#\ \text{yield unknown})}$$

We may now have a look at the relation between the probability and the suitability in each square.

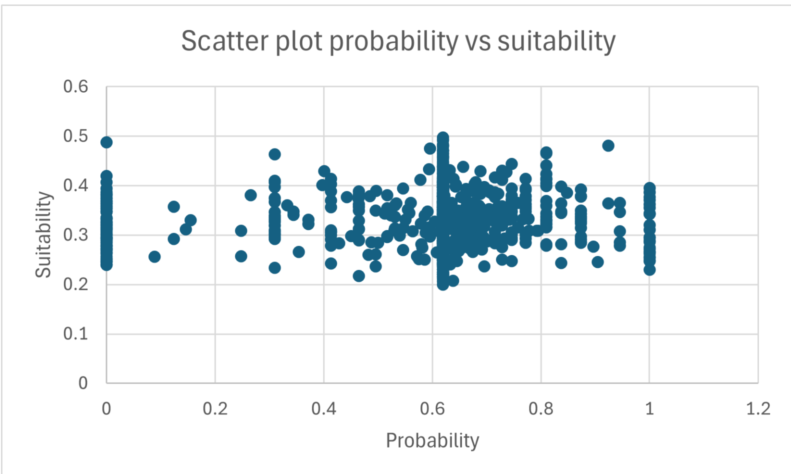


FIGURE 6: Scatter plot of probability vs suitability

As can be seen from the scatter plot in Figure 6 there appears to be no relation between suitability and probability. This is further confirmed by the R-squared value of 0.00349, which is much smaller than 1. The notable line for probability 0.62 can be explained by the data and the chosen method. All the squares with just one well of which the yield is unknown get a probability of 0.62:

$$\text{Probability} = \frac{0 + 1 * 0.62}{0 + 0 + 1} = 0.62$$

We may also note these wells in Figure 4 in dark green.

We will now have a look at the results when the unknown data is left out and we use the simpler formula as follows:

$$\text{Probability} = \frac{(\#\ \text{yield} \geq 2)}{(\#\ \text{yield} \geq 2) + (\#\ \text{yield} < 2)}$$

We again scatter against the suitability to obtain the scatter plot in Figure 7. We can see that no relation can be obtained, which is also shown by the R-squared value of 0.0118, which is much smaller than 1.

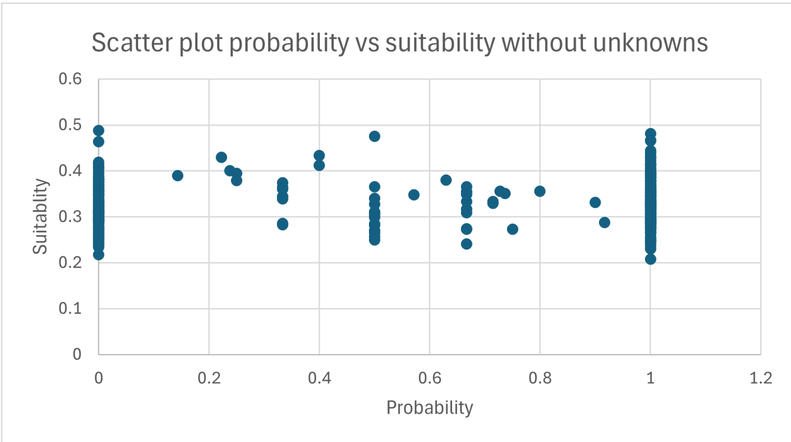


FIGURE 7: Scatter plot of probability vs suitability, with unknown data left out

Hence, with the currently available well data we cannot find a relation between the suitability and the probability. It is advised to extend the data set with failed wells and obtain more data from the current data sparse region. Then a proper relation between the probability and suitability could possibly be obtained.

4 Mathematical Model

In the previous section we discovered the current data was not sufficient for relating probability and suitability. In the following sections we will assume the suitability is equal to the probability. As probability and suitability are both scaled from 0 to 1 a later found relation between them will then easily be applied. With this assumption we shall find a location for a new well by minimizing the total costs, consisting of drilling costs, pipeline cost and hydraulic costs to account for elevation differences between the well and the village. It should be noted that this is done by an approximation and that we hence are not guaranteed to find the exact minimal costs.

We will describe two methods. The first method places only one well, the second method places multiple wells. We do this to find an optimum between drilling/construction costs and pipeline/hydraulic costs. One can imagine that one well leads to large pipeline and hydraulic costs, whereas multiple wells lead to large drilling costs.

Since it is in practice not usual to place one well for the entire population of a Woreda we calculate the costs to connect the well to 50%, 70% and 100% of the population of the Woreda. As we want to compare this to the method of placing multiple wells we will also calculate the costs for connecting the multiple wells to 50%, 70% and 100% of the population.

Formally, we may describe the problem of finding a well location with minimal costs as follows:

Let x be the new well and let V denote the set of villages connected to well x. If we now let T be the set of all possible locations in the Woreda we have that $x \in T$ and $V \subseteq T$.

We have the following two cost functions:

The expected drilling costs, $f_d(x) = \frac{1}{S(x)} \cdot C$, where $S(x)$ is the suitability at x and C the cost of a well drilling/ construction. Since we assume that the suitability is equal to the probability, one can consider $\frac{1}{S(x)} = \frac{1}{Probability}$ which is of course equal to the expected number of drillings. We will assume the costs of drilling are not dependent on the location.

The pipeline costs, $f_p(x, V) = r_p \cdot C \cdot \sum_{i \in V} d(x, i) + r_e \cdot C \cdot \sum_{i \in V} HP(x, i)$, where r_p is the ratio between the costs of a drilling and a meter pipeline. For example, if for the price of one drilling one can buy five kilometers (5000 m) of pipeline then $r_p = \frac{1}{5000}$. Furthermore, $d(x, i)$ is the distance in meters from x to village $i \in V$.

This distance should also account for differences in elevation. That is, if x has lower elevation than village i, we add the costs of hydraulic power (HP) needed [16]. We multiply this with the ratio r_e between the costs of a drilling and a kWh of energy. We do not take the power needed to pump the water from the water level to ground level into account here, as that may be regarded as part of the drilling and construction cost C, as this is dependent on the well.

Let $h = \text{elevation(village } i) - \text{elevation}(x)$

$$HP(x, i) = \begin{cases} \frac{\rho \cdot g \cdot h \cdot Q \cdot hours}{\eta \cdot 1000} & h \geq 0 \\ 0 & h < 0 \end{cases}$$

where $HP(x, i)$ is the hydraulic power in kWh, ρ is the density of water, $1000 \frac{kg}{m^3}$, g is the gravitational constant $9.81 \frac{m}{s^2}$, h is the difference in elevation in m, Q is the flow rate in $\frac{m^3}{h}$, $hours$ is the time period to calculate the power over in h, η is the pump efficiency (no unit), and divide by 1000 to obtain the Hydraulic Power in kWh.

We have the following constraint:

Let P denote the total population of the Woreda and p_i the population in village i. Then $g_c(x, V) = \sum_{i \in V} p_i$ denotes the number of people connected to x.

Hence, to find the location with minimal cost we want to solve the following minimization problem.

$$\begin{array}{ll} \text{minimize} & f_d(x) + f_p(x, V) \\ \text{subject to} & g_c(x, V) \geq u \cdot P \\ & x \in T \\ & V \subseteq T \end{array}$$

where u is the percentage of the population that is desired to connect to well x.

We may adapt this such that multiple wells can be placed. Let $X = \{x_1, ..., x_n\}$ denote the chosen well locations and let V_{x_i} denote the cities connected to well x_i. Then, we let $\mathcal{V} = \cup_{x_i \in X} V_{x_i}$ be the set of all chosen villages.

Hence to find the locations with minimal total cost we want to solve the following minimization problem.

$$\begin{array}{lll} \text{minimize} & \sum\limits_{x_i \in X} (f_d(x_i) + f_p(x_i, V_{x_i})) & \\ \text{subject to} & \sum\limits_{x_i \in X} g_c(x_i, V_{x_i}) \geq u \cdot P & \\ & V_{x_i} \cap V_{x_j} = \emptyset & \text{for } i \neq j \text{ and } i, j \in 1, ..., n \\ & X \subseteq T & \\ & \mathcal{V} \subseteq T & \end{array}$$

5 Data preparation

For each Woreda two files are prepared with QGIS and the PyGis plug-in, which allows us to write python code and execute it in QGIS.

A grid of 500 meters by 500 meters is prepared, where each point denotes a potential location for a well. Each point has a unique identification number and each point is sampled for the suitability and elevation at that point. All the points that lie within 500 meters of an existing well are removed, since a new well close to an existing well might interfere with the productivity of the existing well.

Next to that a 5 kilometer grid is prepared, where each point denotes the middle of a 5 by 5 km square. In each square the total population is counted with QGIS and assigned to its central point. These points are the virtual villages, which also all get a unique identification number. The villages that have population less than 10 are removed for computability reasons. This is a reasonable restriction since area's with a population of less than 10 people per 25 km^2 may be regarded as empty.

In Figure 8 we can see the Woreda Bilate Zuria with both grids. The potential well locations are the black dots and the blue dots are the villages. We can see that the holes between the black dots are where the existing wells are and we may also note that here all villages have population more than 10 since none got removed.

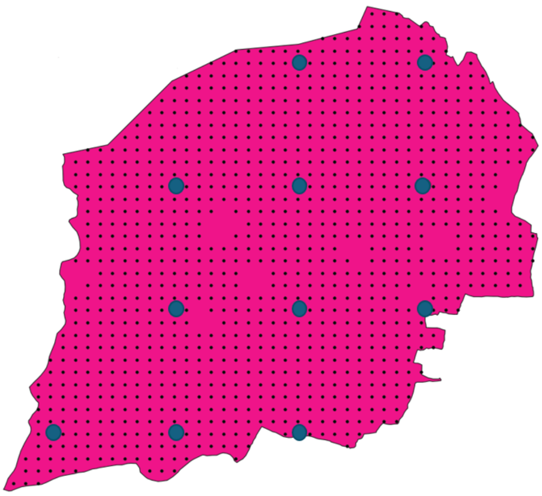


FIGURE 8: Bilate Zuria with grids: potential well locations (black) and villages (blue)

We now create a gigantic distance matrix with the distances from each potential well location to each village and combine all the information and export to two Excel files.

The 'grid' file contains in each row a potential well location (its ID), the distances to all villages, the suitability and the elevation.

The 'village' file contains in each row a village (its ID), the population and the elevation.

In Figure 9 and Figure 10 we see (part) of these Excel exports. Notice that the village ID's in the 'Village'-Excel are the same as the headers in the 'Grid'-Excel to show the distance from the potential well locations to the villages.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	fid	ID	6696	6784	6785	6786	6873	6874	6875	6876	6963	6964	6965	suitability1	elevation1
2	1	343138	2647	13787	9706	6954	20673	16777	13623	11822	23850	20564	18084	0.43100	1392
3	2	343700	1567	12670	8640	6230	19565	15713	12690	11189	22787	19579	17247	0.47200	1400
4	3	343701	1947	13110	8999	6336	19979	16070	12937	11248	23142	19866	17429	0.39700	1406
5	4	343702	2371	13554	9371	6479	20396	16434	13199	11330	23503	20162	17624	0.42800	1401
6	5	344258	2020	9854	6423	5952	16863	13360	11074	10808	20391	17604	15939	0.50500	1426
7	6	344259	1553	10269	6678	5811	17254	13668	11224	10730	20715	17839	16043	0.48800	1421
8	7	344260	1115	10690	6959	5709	17651	13988	11393	10676	21047	18085	16162	0.54200	1417
9	8	344261	759	11118	7264	5650	18053	14318	11582	10645	21385	18342	16296	0.48500	1417
10	9	344262	639	11552	7589	5635	18459	14657	11789	10637	21729	18608	16443	0.51300	1414
11	10	344263	860	11992	7933	5665	18870	15006	12013	10652	22079	18884	16605	0.48600	1415
12	11	344264	1254	12435	8292	5738	19285	15363	12253	10691	22435	19168	16780	0.45000	1417
13	12	344265	1705	12883	8666	5853	19704	15728	12509	10753	22796	19462	16967	0.37200	1421
14	13	344266	2177	13335	9052	6007	20127	16100	12780	10838	23163	19764	17168	0.38500	1418
15	14	344820	2922	8749	5541	5906	15774	12366	10348	10548	19371	16713	15280	0.45300	1433
16	15	344821	2422	9159	5748	5676	16160	12660	10460	10421	19686	16931	15357	0.41200	1437
17	16	344822	1923	9577	5989	5481	16552	12965	10595	10316	20010	17161	15449	0.48400	1442
18	17	344823	1425	10003	6261	5327	16951	13283	10751	10235	20341	17403	15556	0.45200	1430
19	18	344824	928	10435	6560	5216	17354	13612	10928	10178	20678	17655	15679	0.42400	1432
20	19	344825	439	10874	6883	5152	17762	13950	11124	10145	21022	17917	15816	0.46400	1436
21	20	344826	156	11317	7225	5135	18175	14299	11340	10137	21372	18190	15969	0.39300	1438
22	21	344827	597	11765	7586	5168	18593	14656	11572	10153	21728	18472	16135	0.40200	1438
23	22	344828	1091	12217	7961	5248	19014	15021	11822	10194	22089	18763	16315	0.43800	1433

FIGURE 9: 'Grid'-Excel Bilate Zuria

	A	B	C	D
1	fid	id	population	elevation
2	6696	6696	2417.8	1440
3	6784	6784	3041.5	1592
4	6785	6785	2430.7	1636
5	6786	6786	1950.5	1671
6	6873	6873	11218.0	1746
7	6874	6874	7222.1	1761
8	6875	6875	4820.7	1783
9	6876	6876	10384.1	1808
10	6963	6963	15683.4	1905
11	6964	6964	26983.7	1913
12	6965	6965	32208.9	1870

FIGURE 10: 'Village'-Excel Bilate Zuria

To be able to apply this data to find a solution to the minimization problem as described in Section 4, we need values for the variables in the cost functions and the Hydraulic Power function. We use the following values:

$r_p = \frac{1}{5000}$ (no unit)

$r_e = \frac{1}{\frac{500000}{0.32}}$ (no unit)

$Q = 0.03 \frac{m^3}{h}$

$\eta = 0.70$ (no unit)

$hours = 10 \cdot 365 \cdot 24h$

These values have been recommended by Acacia Water or obtained through empirical observations. We use a time period of 10 years to calculate the hydraulic power over because we may assume the well will be used for at least ten years.

This gives us

$$HP = \begin{cases} \frac{1000 \cdot 9.81 \cdot h \cdot 0.03 \cdot 10 \cdot 365 \cdot 24}{0.70 \cdot 1000} \approx 36.8 \cdot 10^3 \cdot h & h \geq 0 \\ 0 & h < 0 \end{cases}$$

The value of r_e has been chosen such that when multiplied with the hydraulic power it gave approximately the same order of magnitude as r_p multiplied with some sample distances. Here, the 0.32 comes from the price of one kWh of energy, at the time of writing the model. For r_p the value $\frac{1}{5000}$ has been chosen since we estimated that for the cost of one well we can buy 5000 meters of pipeline.

6 One Well

Recall that to find the optimal well location for one well, we want to solve the following program:

$$\begin{array}{ll} \text{minimize} & f_d(x) + f_p(x, V) \\ \text{subject to} & g_c(x, V) \geq u \cdot P \\ & x \in T \\ & V \subseteq T \end{array}$$

To solve this we use a brute-force and greedy approach. We try all possible options and then greedily take the cheapest options until we have reached our goal.

For each potential well location we calculate the cost to connect to each village. Recall that this is the weighted sum of the pipeline costs and energy costs to account for Hydraulic Power. Then for each village we calculate the cost per capita, so the connection costs divided by the population. We then order the villages from lowest to hightest cost per capita. Starting at the village with the lowest cost per capita keep connecting villages until the population of the connected villages has exceeded the required amount of (total population) * (percentage to connect). We have thus calculated the near-minimal cost to connect (total population) * (percentage to connect) people to a potential well location. Since we greedily take the cheapest connections we are not guaranteed the absolute minimal cost. Add to this cost the drilling cost. Remember that this has been defined as $\frac{1}{\text{Suitability}} \cdot C$.

Repeat this for each potential well location and obtain a configuration and cost for each of them.

The chosen well location is then the one with the lowest cost: the sum of the connection costs of the chosen well location and villages and the digging cost of the well.

Let us consider the following example in Figure 11 and suppose we want to connect 70% of the population:

FIGURE 11: Example map with villages (blue dots) and potential well locations (black crosses)

For simplicity we assume that there is no elevation difference in this map. In Table 2 we can see the distances and population and we see that the suitability of potential well location 1 is 0.4 and that it is 0.3 for potential well location 2.

	Village 1	Village 2	Village 3	Village 4
Potential well 1	1000 m	1000 m	2000 m	2000 m
Potential well 2	2000 m	2000 m	1000 m	1000 m
Population	400	350	50	250

TABLE 2: Distances between potential locations and villages

We now calculate the connection cost per capita per village.

Recall that the connection cost is defined as follows.

$f_p(x,V) = r_p \cdot C \cdot \sum_{i\in V} d(x,i) + r_e \cdot C \cdot \sum_{i\in V} HP(x,i)$. Since we have no elevation difference in this map this simplifies to $f_p(x,V) = r_p \cdot C \cdot \sum_{i\in V} d(x,i)$.

For potential well 1 and village 1, with distance 1000 m and population 400, we have the following connection cost per capita: $\frac{1000}{5000\cdot 400} = 0.00050$. Recall that we defined $r_p = \frac{1}{5000}$ as our ratio between drilling and pipeline costs. In Table 3 we can see the costs per capita for each potential well location and village.

	Village 1	Village 2	Village 3	Village 4
Potential well 1	0.00050	0.00057	0.00800	0.00160
Potential well 2	0.00100	0.00114	0.00400	0.00080

TABLE 3: Connection cost per capita for each potential well location and village

We now pick the cheapest connections until we have reached the desired population. The total population of this map is 1050 and so 70% is 735.

For potential well location 1 the cheapest connection is Village 1, with population 400. Then Village 2, which gives us a total of 750, which is bigger than 735 so we stop.

The total costs for potential location 1 including drilling cost are now $\frac{1}{0.4}C + \frac{1}{5000}(1000 + 1000)C = 2.9C$

For potential well location 2 the cheapest connection is Village 4, with population 250. Then Village 1, which gives us a total of 650. Then Village 2, which gives us a total of 1000, which is bigger than 735 so we stop. We may note here that we overshoot by almost 300, which is not optimal.

The total costs for potential location 2 including drilling cost are now $\frac{1}{0.3}C + \frac{1}{5000}(2000 + 2000 + 1000)C = 4.3C$

Clearly, it is the best option to choose potential well location 1 and connect it to Village 1 and 2.

7 Multiple wells

Recall that to find the optimal well locations, we want to solve the following program:

$$
\begin{array}{lll}
\text{minimize} & \sum\limits_{x_i \in X} (f_d(x_i) + f_p(x_i, V_{x_i})) & \\
\text{subject to} & \sum\limits_{x_i \in X} g_c(x_i, V_{x_i}) \geq u \cdot P & \\
& V_{x_i} \cap V_{x_j} = \emptyset & \text{for } i \neq j \text{ and } i, j \in 1, ..., n \\
& X \subseteq T & \\
& \mathcal{V} \subseteq T &
\end{array}
$$

To solve this we may rewrite the general description of a Uncapacitated Facility Location Problem, as described in Section 2, to fit our situation. The facilities F are the potential well locations, the cities C are the virtual villages. The opening cost f_i of well i is $\frac{1}{\text{Suitability(i)}}$ and the connection cost c_{ij} between well i and village j is the weighted sum of pipeline cost and energy cost for hydraulic power, $c_{ij} = r_p \cdot C \cdot d(i,j) + r_e \cdot C \cdot HP(i,j)$.

The objective is to open (drill) a subset of the wells F and connect each village to an open well so that the total cost is minimized.

We may solve this with an algorithm adapted from [10] with approximation factor 1.61. For this algorithm it is required that our connection costs satisfy the triangle inequality (such that it is a metric space).

Proposition 7.1. *The connection costs c_{ij} satisfy the triangle inequality.*

Proof. Suppose to the contrary that there exist a well i, village j and another village k such that $c_{ij} > c_{ik} + c_{kj}$. Then,

$$r_p \cdot C \cdot d(i,j) + r_e \cdot C \cdot HP(i,j) > r_p \cdot C \cdot d(i,k) + r_e \cdot C \cdot HP(i,k) + r_p \cdot C \cdot d(k,j) + r_e \cdot C \cdot HP(k,j)$$

which we may rewrite as

$$r_p \cdot d(i,j) + r_e \cdot HP(i,j) > r_p \cdot (d(i,k) + d(k,j)) + r_e \cdot (HP(i,k) + HP(k,j)) \tag{1}$$

However, the distances $d(i,j)$ are taken *as the crow flies* and hence are the shortest way to go from i to j. So $d(i,j)$ satisfies the triangle inequality. Likewise, the hydraulic power $HP(i,j)$ depends only on the height difference h between i and j, which satisfies the triangle inequality too.

Hence we have that $d(i,j) \leq d(i,k) + d(k,j)$ and $HP(i,j) \leq HP(i,k) + HP(k,j)$, which we may combine to obtain

$$r_p \cdot d(i,j) + r_e \cdot HP(i,j) \leq r_p \cdot (d(i,k) + d(k,j)) + r_e \cdot (HP(i,k) + HP(k,j))$$

for $r_p, r_e > 0$, which contradicts Equation 1 and so we have shown that the connection costs c_{ij} satisfy the triangle inequality.

□

The algorithm works as follows:

1. We introduce a notion of time. The algorithm starts at time 0. At this time, all villages are unconnected, all wells are unopened, and the budget of every village j, denoted

by B_j , is initialized to 0. At every moment, each village j offers some money from its budget to each unopened well i. The amount of this offer is computed as follows:

If j is unconnected, the offer is equal to $max(B_j - c_{ij}; 0)$ (i.e., if the budget of j is more than the cost it has to pay to get connected to i, it offers to pay this extra budget to i);

If j is already connected to some other well i' , then its offer to well i is equal to $max(c_{i'j} - c_{ij}; 0)$ (i.e., the amount that j offers to pay to i is equal to the amount j would save by switching its well from i' to i).

2. While there is an unconnected village, increase the time, and simultaneously, increase the budget of each unconnected village at the same rate (i.e., every unconnected village j has $B_j = t$ at time t), until one of the following events occur. If multiple events occur at the same time, process them in an arbitrary order.

(a) For some unopened well i, the total offer that it receives from villages is greater than or equal to the cost of opening i. In this case, we open well i, and for every village j (connected or unconnected) which has a non-zero offer to i, we connect j to i.

(b) For some unconnected village j, and some well i that is already open, the budget of j is greater than or equal to the connection cost between j and i. In this case, we connect village j to well i.

In the original algorithm as described by [10] connections were made when the offer was exactly equal to the opening costs or when the budget was exactly equal to the connection cost. For our real life problem we found this infeasible and so we adjusted it to allow for greater than or equal. The verification of this adjusted algorithm can be found in the Appendix.

Because of this we lose some accuracy. In the proof for the approximation factor of 1.61 it is assumed that the offers and opening cost, and budget and connection cost, are always equal when connecting. To make sure the 'greater than' case is not much bigger than the 'equal' case, we take t to increase with only 0.01 at each iteration. We may therefore assume the total offer can never overshoot the opening cost with more than 0.01 times the number of nonzero offers. For the biggest Woreda (Maji) the biggest number of nonzero offers was 6 at a time. For the smallest Woreda (Bilate Zuria) this was five. Hence it was at most 0.06 off, and we should increase our approximation factor by 0.06.

Note that by taking such small time steps the computations for all ten Woreda's took about five hours. Suggestions for a faster, but slightly less accurate implementation of the algorithm can be found in the Appendix.

We now use this algorithm on the data and parameters as prepared in Section 5 to obtain our final results.

Like the one well algorithm, we also want to be able to connect to 50% or 70% of the population for the multiple well algorithm, such that we can compare one well versus multiple wells. We find the connections to 50% and 70% of the population by applying a post-process to the multiple well configuration that connects to all villages that we obtained using the algorithm above.

We take the configuration obtained with the multiple well algorithm above and we let each chosen well and the cities connected to it be called a star. To find a configuration that connects 50% or 70% of the population we slightly adjust our process of Section 6.

We calculate for each star the cost per capita and order them from low to high. We now keep picking stars until we have exceeded the desired population connected. Now we go back to the last star that we added to the configuration, and calculate for each of the villages in this star the cost per capita. Once again order from low to high and add these single villages to the already gathered stars until the desired population is reached. We will call this algorithm the Star Picking algorithm.

Let us consider an example result of the multiple well algorithm, where all villages have been connected, in Figure 12. Here we can see three wells and the villages they are connected to, and the population of each village. For simplicity we take all connection costs to be 10 C and all well drilling costs to be 5 C. We will use the Star Picking Algorithm to find a configuration that connects 70% of the population to a well. The total population of this area is 1000 and so we want to connect 700 people.

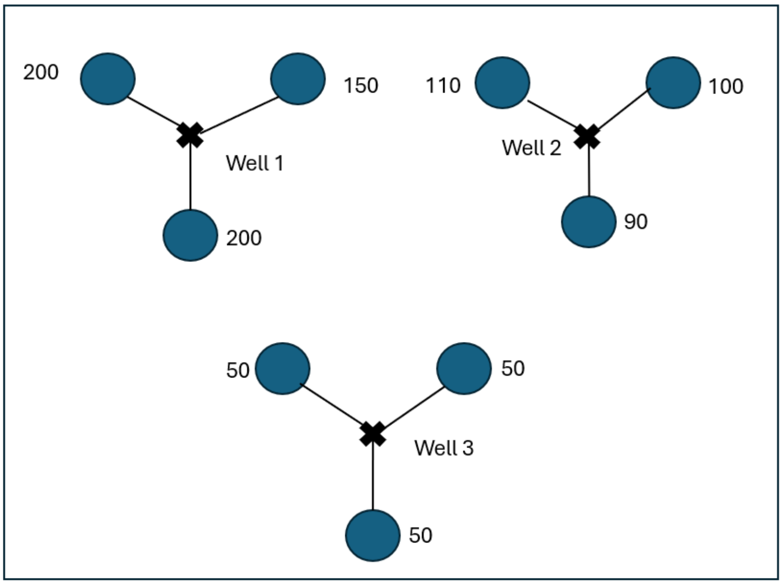


FIGURE 12: Example map with villages (blue dots) and wells (black crosses)

We start by calculating the cost per capita of each star. We name the stars after their well number.

Star	Costs (C)	Capita	Cost per capita (C)
1	$3 \cdot 10 + 5 = 35$	$200 + 200 + 150 = 550$	0.064
2	$3 \cdot 10 + 5 = 35$	$110 + 100 + 90 = 300$	0.117
3	$3 \cdot 10 + 5 = 35$	$3 \cdot 50 = 150$	0.233

TABLE 4: Costs per capita for each star

From Table 4 we can see Star 1 has the lowest cost per capita so we pick that star to be part of our final configuration. We have now connected 550 people. We proceed with the next cheapest star, which is Star 2. We have now connected 850 people, meaning we overshot the 700 people we wanted to connect. Hence we will look at the separate connections within Star 2, which can be seen in Table 5.

Village 1 has the cheapest cost per capita and so we add it to our configuration, meaning we have now connected $550 + 110 = 660$ people. If we now add the next cheapest option Village 2 we have connected 760 people, which is more than 700 so we stop.

This gives us the final configuration in Figure 13, with a total cost of $5 \cdot 10 + 2 \cdot 5 = 40C$

Village	Costs (C)	Capita	Cost per capita (C)
1	10	110	0.0909
2	10	100	0.1000
3	10	90	0.1111

TABLE 5: Costs per capita for each connection

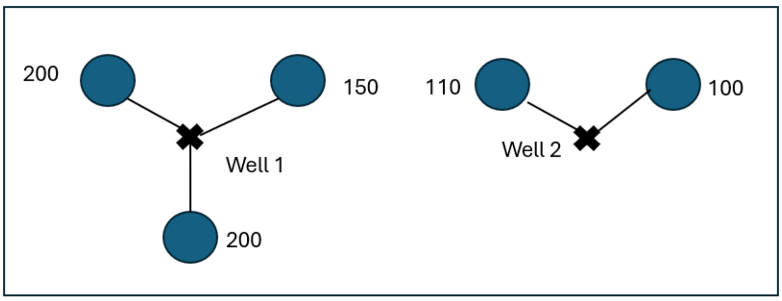


FIGURE 13: Results of the Star Picking algorithm to connect 70% of the population

8 Results

All obtained maps with well and village locations and tables with cost per Woreda per algorithm, as well as the number of wells and kilometers of pipeline can be found in the Appendix, Section 11.4 and 11.5.

Comparing the maps of the 'One well'-algorithm we can see that the well locations to connect 50%, 70% and 100% of the population are quite similar. For connecting 50% of the population we can see that there are very few villages. Apparently 50% of the population lives in about 4 villages for most Woreda's. For the connecting 70% of the population it may be noted that the chosen villages are clustered around the well, as would intuitively be expected to result in a configuration with lowest costs. For connecting 100% of the population the wells appear to be in the middle of the villages, which is according to intuition too, as for the one well case the most costs will come from pipeline and it is hence useful to have a minimal amount of that. We may conclude that the 'One well' algorithm performs according to intuition.

Comparing the maps of the 'Multiple well'-algorithm we can see that for connecting 100% of the population the map has its wells nicely distributed over the areas. For connecting 50% and 70% of the population the wells form clusters and far less villages are connected. This makes sense since the stars with cheapest cost per capita are chosen, making small villages unattractive. For all percentages we see that quite a lot of wells are chosen compared to the amount of villages. It appears to be that about 3 to 4 villages share one well.

In Figure 14, 15 and 16 we can see how the different algorithms compare in costs. The Woredas have been ordered by area, with the Woreda with the smallest area on the left and the biggest area on the right. We see that for Woreda's with large area placing one well to connect 100% of the population is very expensive. This makes sense as there are a lot of pipelines to be laid in a large area. We consequently also see that placing multiple wells to connect 100% of the population is much cheaper, especially for large Woreda's.

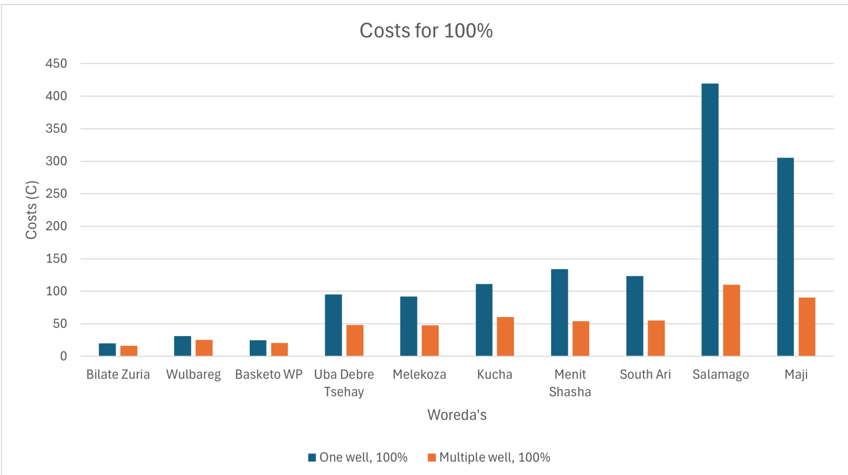


Figure 14: Comparison of algorithms connecting 100% of the population by their cost.

For connecting 70% of the population it is usually cheaper to place one well than to place multiple wells: only for some big Woreda's does the multiple well algorithm work better. This can be explained by the fact that the wells have been chosen from the 100% configuration and have thus been optimized for that case and not the 70% case. Some suggestions and idea's tried to solve this are discussed in the Discussion.

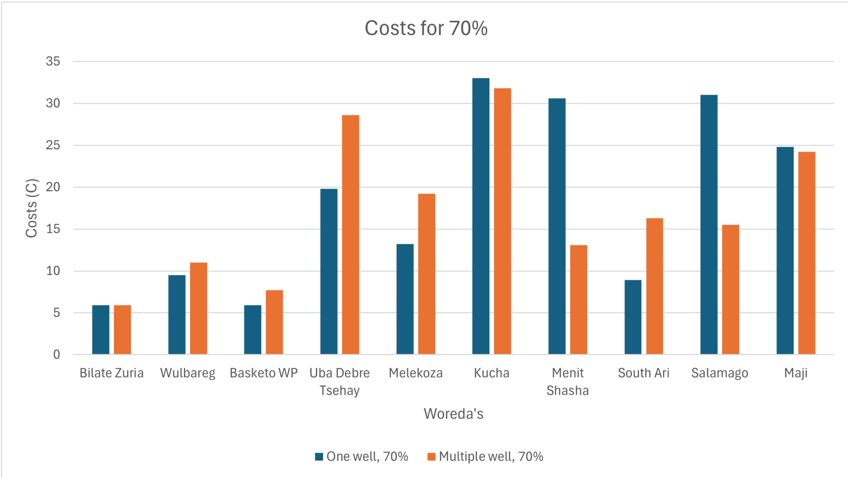


FIGURE 15: Comparison of algorithms connecting 70% of the population by their cost.

For connecting 50% of the population it is usually cheaper to place one well, except for some of the biggest Woreda's. We may explain this by the same reasons as for connecting 70% of the population.

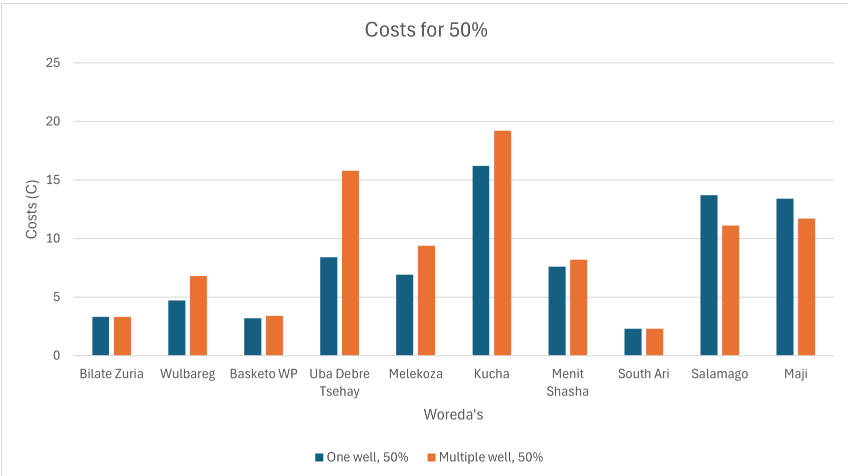


FIGURE 16: Comparison of algorithms connecting 50% of the population by their cost.

As expected, when connecting 100% of the population the total distance from well to village is always much shorter when placing multiple wells compared to placing only one well. This total distance can roughly be taken as the amount of pipeline needed, although it does not take elevation into account.

In Figure 17 we can see that this is not the case for the 50% and 70% algorithms. We see that for Uba Debre Tsehay and South Ari the one well algorithm uses less pipeline than the multiple well algorithm, and for Bilate Zuria they are equal. Notable is also that for South Ari for 50% the kilometers of pipeline is very small. This is explained easily with the fact that only one well is placed which is connected to only one village, the same village for both algorithms too.

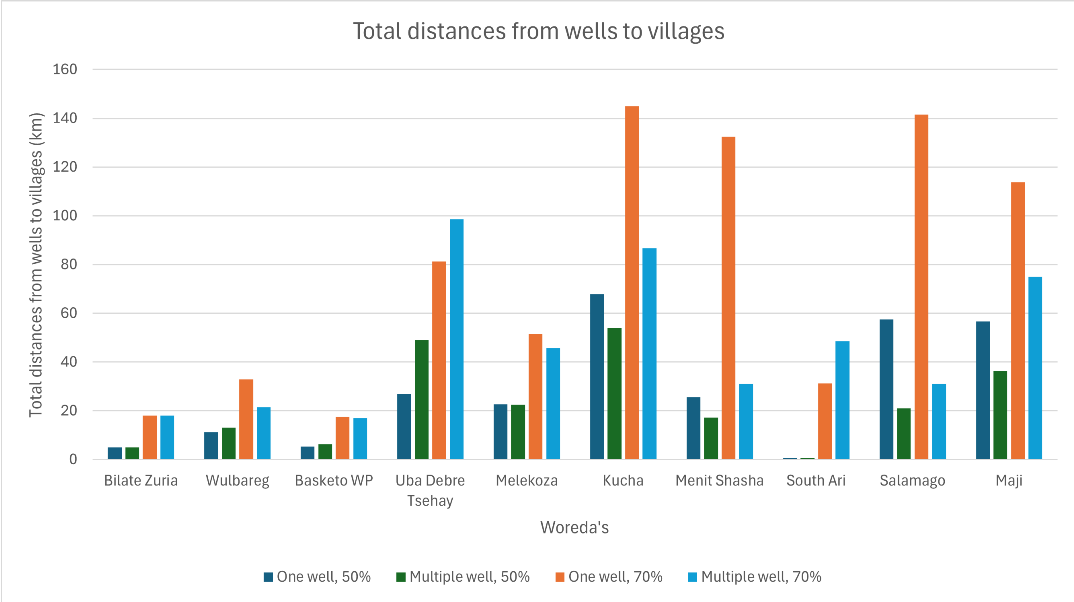


FIGURE 17: Comparison of total distances from wells to villages

In Figure 18 we can see in brown the Kebeles (villages) that have been marked by Acacia as the areas where a further detail study for placing a well will be done, meaning a well will be placed within each brown area. When we compare this to the wells placed by the one well algorithm to connect to 100% of the population, we see that for most Woreda's the chosen location is relatively close to the chosen Kebele.

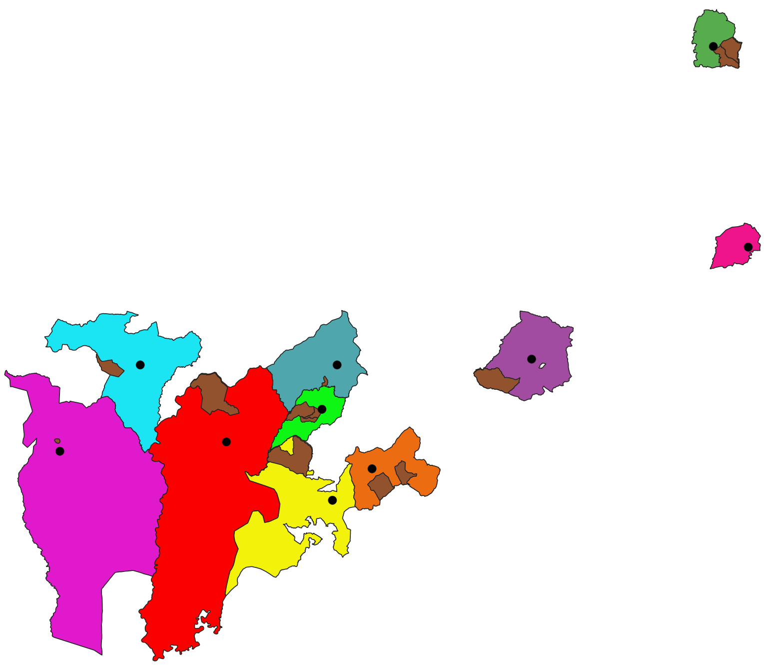


FIGURE 18: Chosen Kebeles (brown) and One well connected to 100% of the population (black)

9 Discussion

As we have seen in the results the multiple well algorithm to connect 50% and 70% percent of the population, which we called the 'Star Picking Algorithm', does not perform as wished. To connect 70% of the population, in five out of ten cases the one well algorithm performs better, and they perform the same for one case. To connect 50% of the population the one well algorithm performs better in 6 out of 10 cases, and equal to the Star Picking algorithm in 2 cases. Of course it could be that placing one well is truly the best solution, but then the Star Picking algorithm should be able to pick that up too, but for connecting 70% of the population only for the Bilate Zuria Woreda does the Star Picking algorithm pick one well. When connecting 50% of the population, 3 Woreda's get assigned one well.

In this Discussion we will showcase why this might be the case, what has been tried to improve this and give suggestions for further research.

Firstly, it makes sense that the Star Picking algorithm is not perfect. We already have to account for an error by choosing to use an approximation algorithm to connect 100% of the population, as the problem is NP hard. We then proceed to, loosely speaking, pick the cheapest 50% or 70% from this 100%, from which we can clearly say that it is not optimal. But, for some Woreda's it ends up with the exact same cost as the one well algorithm which shows us the algorithm is not just nonsense.

Multiple attempts at other algorithms have been made and tested on the case where we connect 70% of the population.

To account for the fact that the best 70% are only chosen in the end an Integrated Approach has been tried. This meant that we made being connected to 70% a stopping condition for the algorithm. Apart from terminating faster this did not help. We have to make sure the 'best' connections are chosen first, or at least before termination.

To do so we first changed the algorithm such that the budget of each city was not increased equally, but based on their population. Each round the budget of each city was increased with 0.0001·population. In this way, the bigger cities should be able to connect faster and thus before termination. The results of this were worse than the Star Picking algorithm and so the idea was dismissed.

Secondly, it was tried to artificially make connections with big cost per capita more expensive. At the beginning of the calculations for each Woreda three potential well locations were chosen and their connection cost per capita to all villages was calculated and sorted from low to high. Then the 70th Quantile (and later 90th) was picked as the bar value for being a large cost per capita for that Woreda. Then we check each connection cost calculated while performing the algorithm. If it was larger than the bar value it was artificially raised to 100000 to make this connection unattractive to connect for the algorithm. The results of this were much worse than the Star Picking algorithm and so the idea was dismissed.

It has also been considered to design an entirely new algorithm. For example based on the idea in [17]. Here they create a 'Universal Well', with no drilling costs, and the connection costs are the lost profits from the cities that are connected to this Universal Well. In our case however, there are no profits and so we cannot determine what the loss of a village-well connection would cost. Next to that it is not possible to ensure a predetermined percentage of connection with this method.

Another idea was to transform the Uncapacitated Facility Location Problem into a

Partial Set Covering Problem as in [8]. This does have the percentage as a stopping condition but requires a predetermined distance between well and city, such that the well can 'cover' the city. This would mean we would not be minimizing over the biggest costs, the pipeline and hydraulic cost, but only over the drilling costs.

Further research could look into finding an algorithm for a Partial Facility Location Problem, where only a given percentage of the cities needs to be connected.

Besides the Star Picking algorithms there are also some discussion points for the other algorithms.

For the multiple well algorithm we have rewritten the problem as an Uncapacitated Facility Location Problem. But in reality, the problem is not uncapacitated but capacitated. Depending on the location there is certainly a maximal amount of water the well can deliver per day. However as this widely varies per location and to keep the model somewhat simple this has been dismissed. Once it is know how the suitability relates to the capacity/yield of a well a Capacitated Facility Location Problem could be applied and solved.

In general it should be noted that all the costs have been expressed in C, the well drilling costs, and C was assumed to not depend on the location. In reality however, the well drilling costs do vary, depending on the required depth which depends on the ground layers and where the water can be found.

Furthermore the elevation difference has been simplified a lot. In this model the total difference has been taken, the orange arrow in Figure 19. But as we can see from the same picture the pipeline actually has to be quite a bit longer than that. Next to that the formula for Hydraulic Power does not take the friction of the pumped water with the pipe into account.

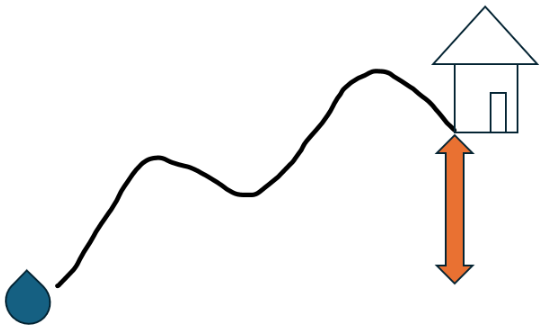


Figure 19: Elevation difference between well and village

In general it should be mentioned that by using approximation algorithms and approximating the real life variables used in the model, for example r_e and r_p, the results of the algorithms are not exact and should be taken as approximate. Further research should look into what values relate to reality the best such that a more accurate approximation of the costs and locations can be made.

10 Conclusion

We have seen how with the currently available data it is not possible to relate suitability and probability of a successful well. It is recommended to extend the data set with failed wells and obtain more data from the current data sparse regions. Then it might be possible to find a relation between suitability and probability, which can be used to improve the algorithms to find well locations.

We then proceeded with the assumption that the suitability and probability are equal. With this we found costs and configurations for placing one well and multiple wells. We saw that when the well(s) should be connected to 50% or 70% of the population the cheapest option is usually to place only one well. We saw that this means that the algorithm for picking multiple wells does not perform perfectly and saw what was tried to improve this. Suggestions for further algorithms consists of the Universal Well idea and changing the UFLP to a Partial Set Cover problem. In practice, this under-performance of the multiple well algorithm is not a problem, as both the multiple well and one well algorithm can be run to determine the location and the cheapest option can be chosen. This is possible because both algorithms have a small run time of a couple of seconds per Woreda.

For connecting to 100% of the population we saw that it was cheaper to place multiple wells, especially for the Woreda's with large area's. It should be noted that calculating the multiple well algorithm to connect all villages took about 5 hours to compute, but this can be reduced to 1 hour or 30 minutes when one can accept slightly higher costs. We also saw that the area's chosen by Acacia, in which a further detail study will be done and eventually a well will be placed, are relatively close to the well locations chosen by the one well algorithm that connects 100% of the population for most of the Woreda's.

For placing multiple wells the problem has been regarded as an Uncapacitated Facility Location Problem. This is an NP hard problem and hence we should make use of an approximation algorithm. We chose an algorithm with approximation factor 1.61 as inspired by [10]. To this approximation factor we have to add 0.06 to compensate for the correction we had to make to the algorithm to make it feasible for our real life application. That is, in [10] the offer and cost need to be exactly equal to be accepted, and in our case the budget is also allowed to be bigger.

All in all we have created two algorithms that can deliver a feasible well configuration per Woreda. Depending on which percentage of the population is desired to be connected it is better to choose the multiple well algorithm or the one well algorithm.

We would like to give the following recommendations:

- Obtain data about failed well drillings such that a relation between suitability and probability might be found.
- Find out what the costs C of a drilling should be based on the suitability, and find out what the costs of a kilometer of pipeline and hydraulic power are. Then an actual approximate cost can be given, rather than one expressed in C.
- When all villages in an area should be connected, use the multiple well algorithm.
- When not all villages in an area should be connected, use both the one well and multiple well algorithm and choose the cheapest option.

References

[1] *Acacia Water*. Accessed 03/12/2024. 2024. URL: https://acaciawater.com/Home/.

[2] Jaroslaw Byrka and Karen Aardal. "An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem". In: *SIAM Journal on Computing* 39.6 (2010), pp. 2212–2231.

[3] Roger C Calow, Alan M MacDonald, Alan L Nicol, and Nick S Robins. "Ground water security and drought in Africa: linking availability, access, and demand". In: *Groundwater* 48.2 (2010), pp. 246–256.

[4] Wikimedia Commons. *File:Ethiopia in Africa (-mini map -rivers).svg*. Accessed 17/12/2024. 2011. URL: https://commons.wikimedia.org/wiki/File:Ethiopia_in_Africa_%28-mini_map_-rivers%29.svg.

[5] Wikimedia Commons. *File:Southern Nations, Nationalities, and People's Region in Ethiopia.svg*. Accessed 17/12/2024. 2011. URL: https://commons.wikimedia.org/wiki/File:Southern_Nations,_Nationalities,_and_People%27s_Region_in_Ethiopia.svg.

[6] Stephen Cook. "The P versus NP problem". In: *Clay Mathematics Institute* 2.6 (2000), p. 3.

[7] Gérard Cornuéjols, George Nemhauser, and Laurence Wolsey. *The uncapicitated facility location problem.* Tech. rep. Cornell University Operations Research and Industrial Engineering, 1983.

[8] Mark S Daskin and Susan Hesse Owen. "Two New Location Covering Problems: The Partial P-Center Problem and the Partial Set Covering Problem". In: *Geographical Analysis* 31.3 (1999), pp. 217–235.

[9] Sudipto Guha and Samir Khuller. "Greedy strikes back: Improved facility location algorithms". In: *Journal of algorithms* 31.1 (1999), pp. 228–248.

[10] Kamal Jain, Mohammad Mahdian, and Amin Saberi. "A new greedy approach for facility location problems". In: *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing.* 2002, pp. 731–740.

[11] Martin Kopecký, Martin Macek, and Jan Wild. "Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition". In: *Science of the Total Environment* 757 (2021), p. 143785.

[12] Compendium voor de Leefomgeving. *Jaarlijkse hoeveelheid neerslag in Nederland, 1910-2022.* Accessed 03/12/2024. 2023. URL: https://www.clo.nl/indicatoren/nl050809-jaarlijkse-hoeveelheid-neerslag-in-nederland-1910-2022.

[13] Shi Li. "A 1.488 Approximation Algorithm for the Uncapacitated Facility Location Problem". In: *Automata, Languages and Programming.* Ed. by Luca Aceto, Monika Henzinger, and Jiří Sgall. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 77–88. ISBN: 978-3-642-22012-8.

[14] Getachew Alem Mera. "Drought and its impacts in Ethiopia". In: *Weather and climate extremes* 22 (2018), pp. 24–35.

[15] David B Shmoys, Éva Tardos, and Karen Aardal. "Approximation algorithms for facility location problems". In: *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing.* 1997, pp. 265–274.

[16] The Engineering Toolbox. *Pump Power Calculator*. Accessed 10/12/2024. 2003. URL: `https://www.engineeringtoolbox.com/pumps-power-d_505.html`.

[17] Francis J Vasko, Dennis D Newhart, Kenneth L Stott Jr, and Floyd E Wolf. "A large-scale application of the partial coverage uncapacitated facility location problem". In: *Journal of the Operational Research Society* 54.1 (2003), pp. 11–20.

11 Appendix

11.1 Python code

All code and relevant README can be found on `https://github.com/annajgesink/Bachelor-Thesis-Water-Wells`

11.2 Verification of multiple well algorithm

In the following section we will verify the multiple well algorithm with 100% of the population connected. As we have slightly adjusted the algorithm by [10], it is necessary to verify it behaves as expected. Recall that we adjusted the algorithm in such a way that we also connect villages to a well when their offer is greater than or equal to the opening cost, and not just when the cost is exactly equal to the offer. Likewise, we also connect a village to a well when the budget is greater than or equal to the connection cost, instead of exactly equal. We will start with an intuitive approach to verify these changes. In Figure 20 we can see that the wells are indeed well spread over the area and connected to a well close to them, as expected.

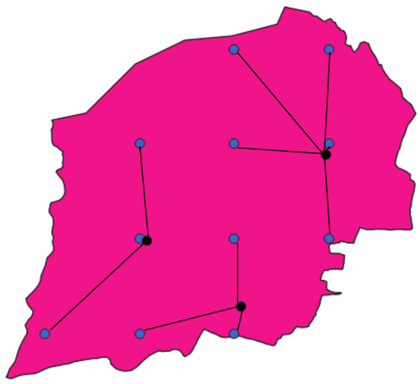


FIGURE 20: Connections between villages and chosen wells, for Bilate Zuria Woreda for 100% of the population connected.

We will now test this algorithm on some small examples to verify its performance. Take the configuration in Figure 21. We may then obtain an Excel sheet with data similar to the data preparation in Section 5, as can be seen in Figure 22. We see the distances from each potential location $10, 11, 12, 13, 14$ to all the villages $100, 200, 300, 400, 500, 600$. We also see that all potential locations have Suitability 0.4 and Elevation 1000 meters. Intuitively, in this configuration the best location to place the well would be location 12, with a cost of $\frac{1}{0.4}C + \frac{1}{5000}(2800 + 2800 + 1000 + 1000 + 2800 + 2800)C + 0C = 5.14C$. When we run the algorithm we obtain chosen location 12 with given costs of 5.14 C.

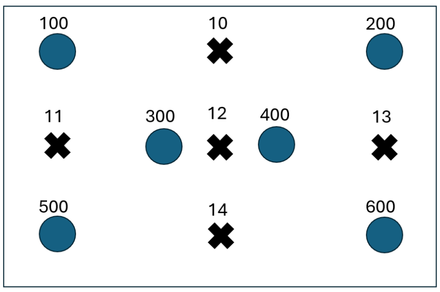


FIGURE 21: Fictional map with villages (blue dots) and potential well locations (crosses)

	A	B	C	D	E	F	G	H	I	J
1	fid	ID	100	200	300	400	500	600	suitability1	elevation1
2	1	10	2000	2000	2200	2200	4500	4500	0.4	1000
3	2	11	2000	4500	1000	3000	2000	4500	0.4	1000
4	3	12	2800	2800	1000	1000	2800	2800	0.4	1000
5	4	13	4500	2000	3000	1000	4500	2000	0.4	1000
6	5	14	4500	4500	2200	2200	2000	2000	0.4	1000

FIGURE 22: Distances for fictional map

Now we change the elevation of 10,11,12 and 13 to 0. This means that the water has to be pumped up from the well to the villages, making these locations more expensive. We expect that 14 is chosen, as this is now the cheapest option. The algorithm also chooses 14, with a cost of 5.98 C.

Changing the elevation back to 1000 for all potential well locations, but now setting the suitability of 10,11,12 and 13 to 0.2, we again expect and obtain location 14 with cost 5.98 C, as locations with lower suitability are more expensive.

Now if we repeat the village configuration in Figure 21 three times with distances of 5 kilometer between the clusters, we obtain Figure 23. We would expect the algorithm to now favour a multiple well configuration, which it does. Of each cluster the middle potential well location is chosen.

If we now set one well in each cluster to a higher suitability this well is chosen. Further changes in the configuration behave as expected too and have been omitted.

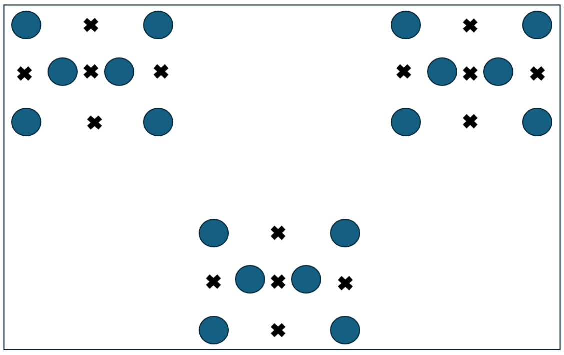


FIGURE 23: Fictional map with villages (blue dots) and potential well locations (crosses)

11.3 Improving the computation time

As we saw in Section 7 computing the multiple well configuration to connect 100% of the population takes about 5 hours. Naturally, we want to decrease this time.

This was done by taking 'smarter' time steps.

Firstly, we initialize all budgets with 1 instead of 0, as it was found that none of the Woreda's make connections before the budget is 1.

Secondly, we increase the budget with 0.1 (instead of 0.01), and then when a connection is about to be made, we go back in time with steps of 0.01 until the difference between the opening cost and offers from the villages (or between the budget and the connection cost) is less than 0.05, which was taken as an acceptable error.

Doing this decreased the computation time from 5 to 1 hour. However, it also very slightly increased the costs.

To reduce the computation time to 30 minutes we take the original algorithm as described in Section 7 and again initialize all budgets with 1. But instead of increasing the budget with 0.01 we increase the budget with 0.1. As expected, this slightly increases the costs.

The results of all three methods to compute the connection costs to connnect 100% of the population with multiple wells can be seen in Table 6.

Woreda	Standard algorithm	Smart time steps	Bigger time steps
Basketo WP	20.6	20.6	20.6
Bilate Zuria	16.3	16.9	17.0
Kucha	60.6	61.0	61.3
Maji	90.2	91.3	91.1
Melekoza	47.9	47.5	47.7
Menit Shasha	54.1	55.1	55.1
Salamago	110.2	110.6	111.0
South Ari	54.9	55.4	56.3
Uba Debre Tsehay	48.4	48.7	49.0
Wulbareg	25.4	23.6	23.6

Table 6: Costs (in C) for the three different methods to connect 100% of the population with multiple wells

11.4 Results: maps

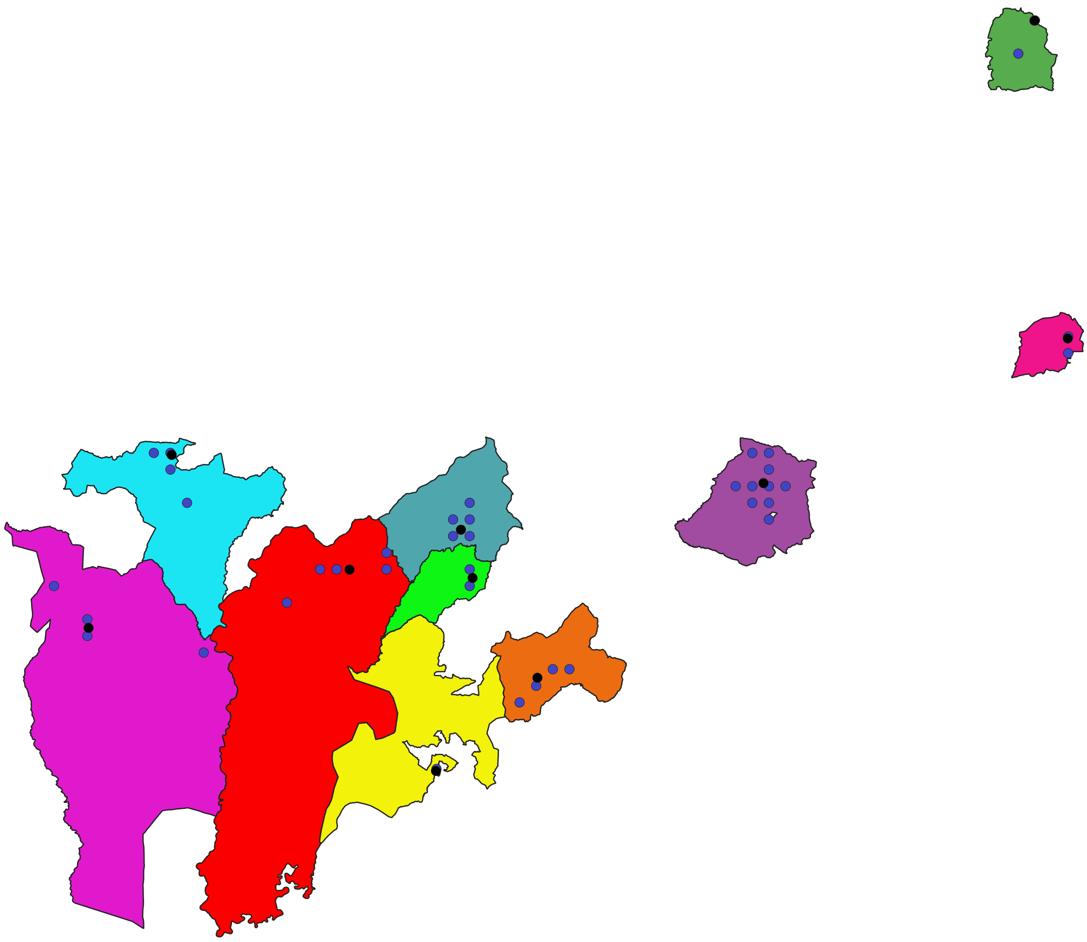
In the Figures below we can see the placement of the wells and the villages they are connected to for the cases of placing one well, with percentage of population to connect 50%, 70% and 100%, and the case for placing multiple wells. For multiple wells too we have the case to connect 50%, 70% and 100% of the population. The wells are shown as black dots and the villages as blue dots.

Recall from the Introduction (Section 1) that the colours on the map correspond to the following Woreda's:

Woreda	Colour
Basketo SP	bright green
Bilate Zuria	pink
Kucha	dark purple
Maji	light purple
Melekoza	dark blue

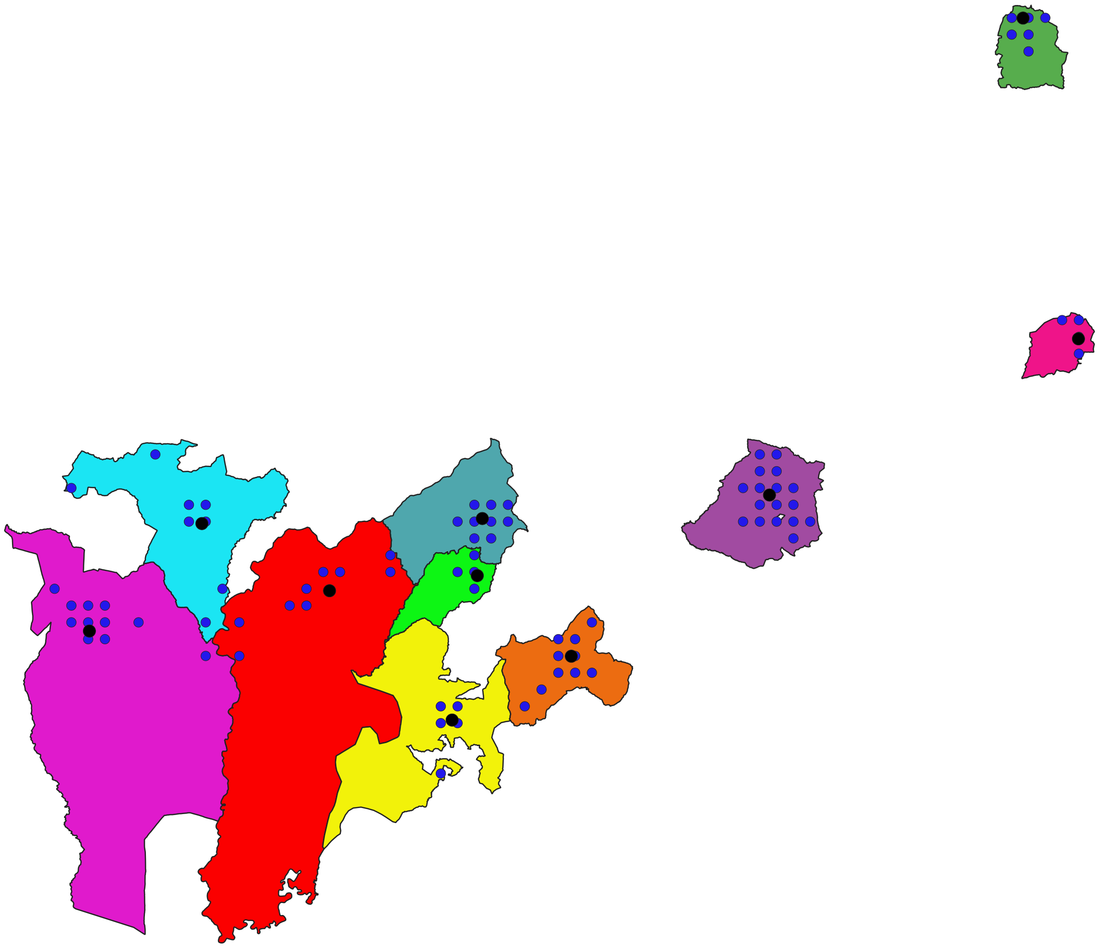
Woreda	Colour
Menit Shasha	light blue
Salamago	red
South Ari	yellow
Uba Debre Tsehay	orange
Wulbareg	dark green

Table 7: Woreda's and their respective colours



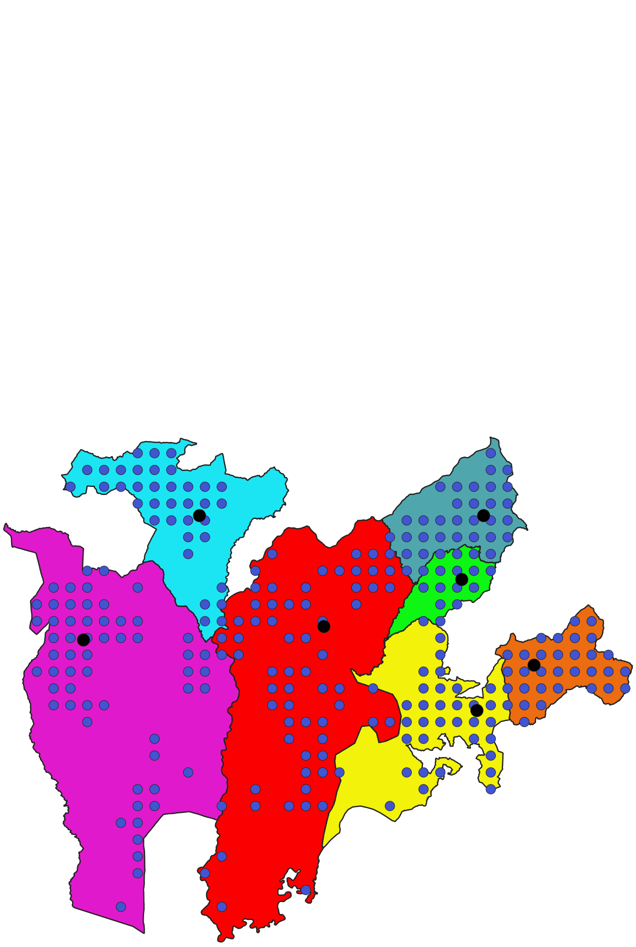
One well connected to 50% of the population

:



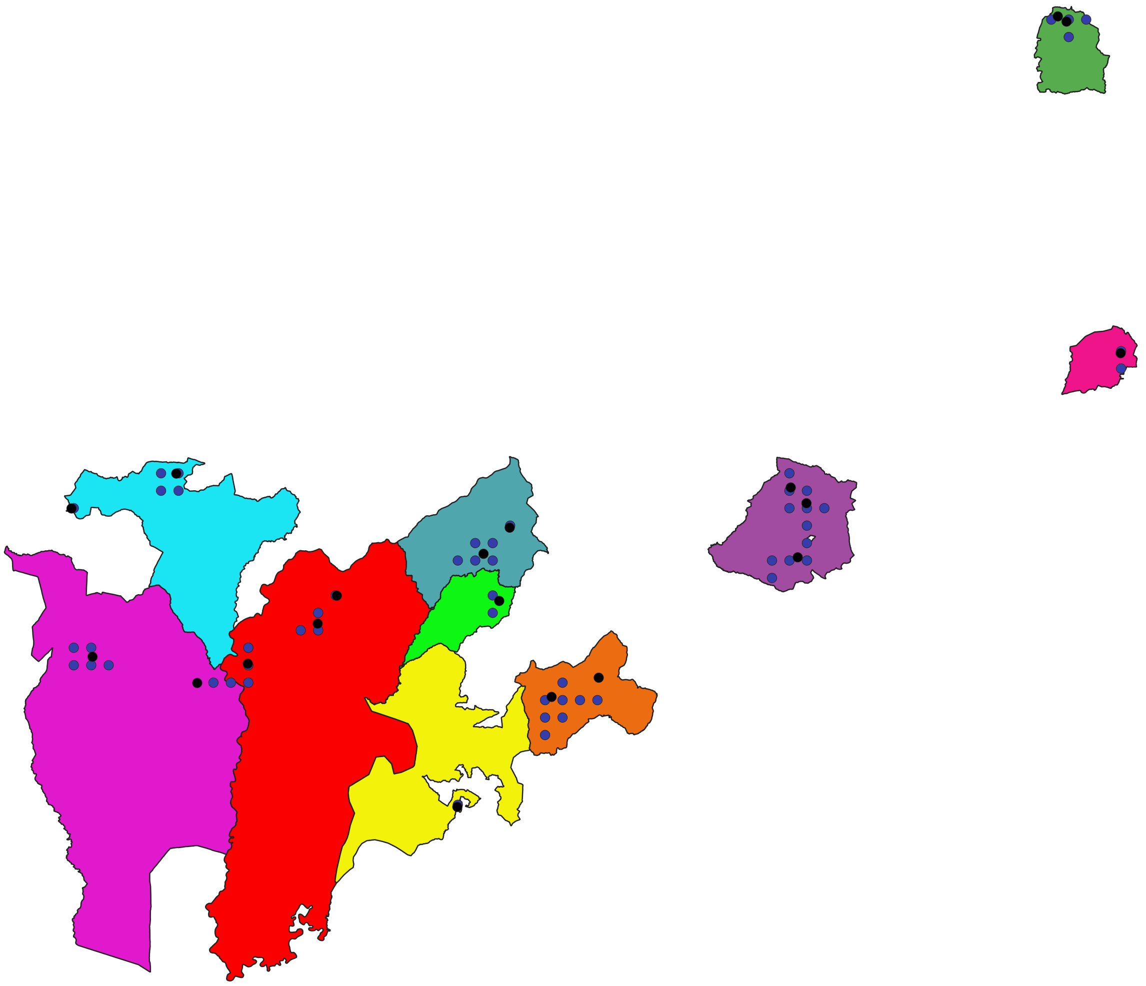
One well connected to 70% of the population

:



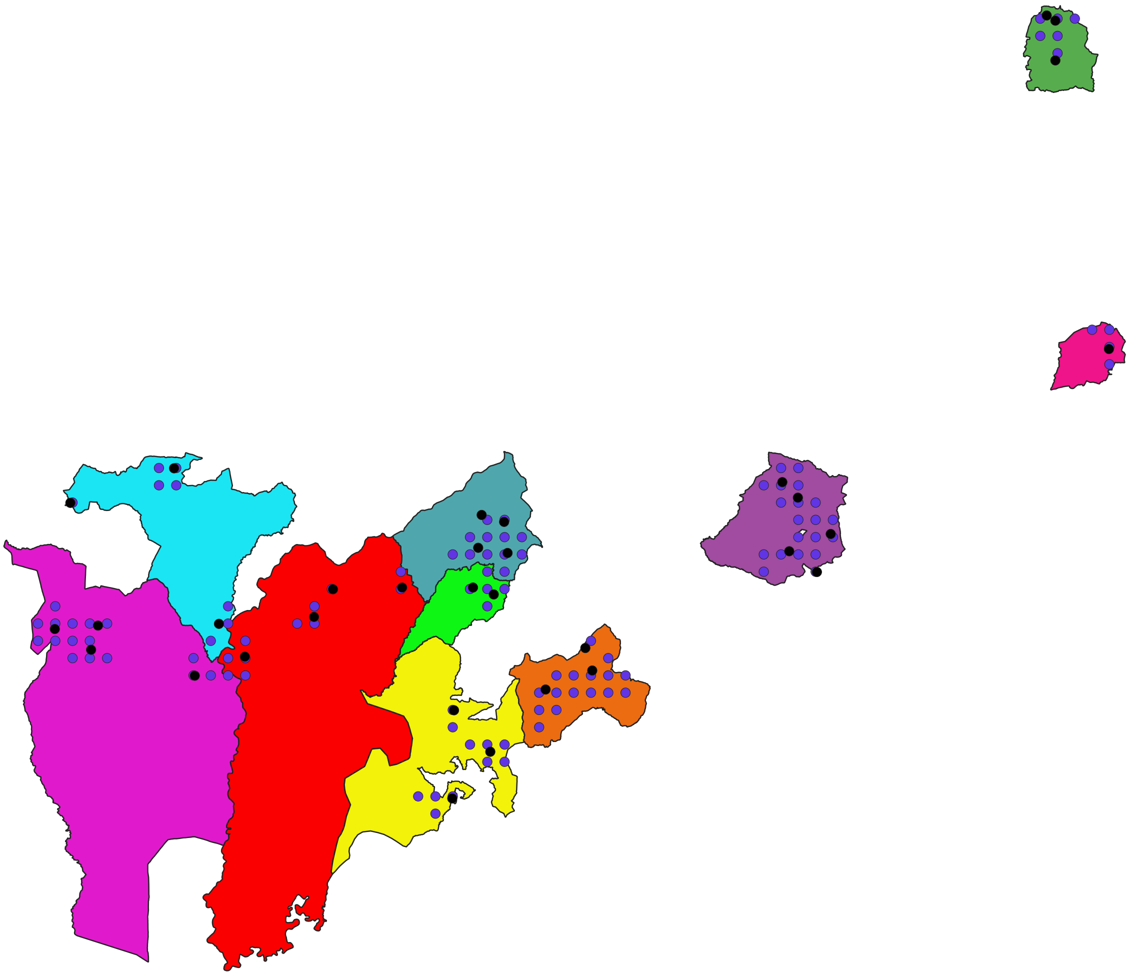
One well connected to 100% of the population

:



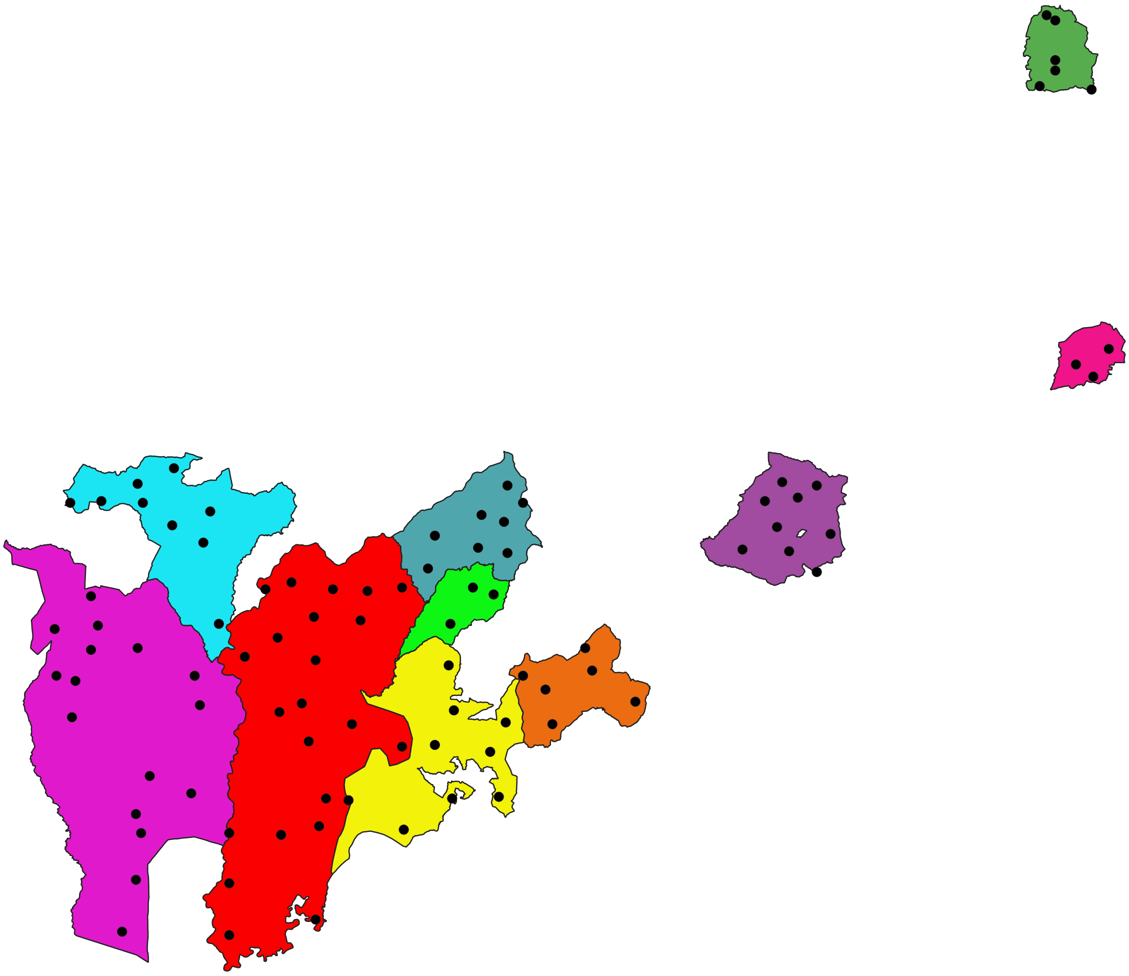
Multiple wells connected to 50% of the population

:



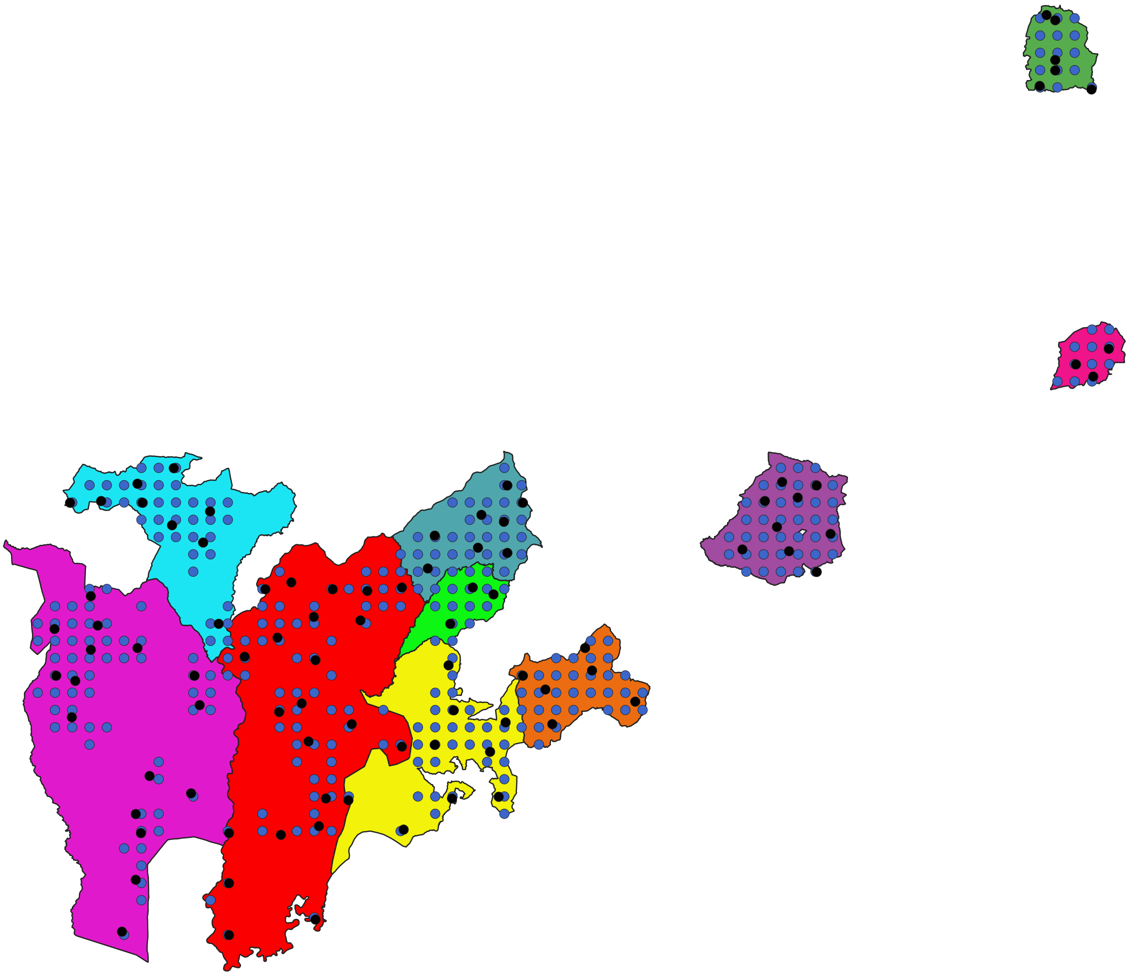
Multiple wells connected to 70% of the population

:



Multiple wells connected to 100% of the population (wells only)

:



Multiple wells connected to 100% of the population

:

11.5 Results: costs

In the following Tables we can see the costs for each Woreda as found by the different algorithms. We also note the number of wells placed and the total kilometers of pipeline.

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	3.2	1	5.3
Bilate Zuria	3.3	1	5.0
Kucha	16.2	1	67.8
Maji	13.4	1	56.7
Melekoza	6.9	1	22.7
Menit Shasha	7.6	1	25.6
Salamago	13.7	1	57.4
South Ari	2.3	1	0.6
Uba Debre Tsehay	8.4	1	26.9
Wulbareg	4.7	1	11.2

Table 8: Woreda's and their respective costs for one well for percentage 50%

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	5.9	1	17.5
Bilate Zuria	5.9	1	18.0
Kucha	33.0	1	144.9
Maji	24.8	1	113.7
Melekoza	13.2	1	51.5
Menit Shasha	30.6	1	132.4
Salamago	31.0	1	141.5
South Ari	8.9	1	31.3
Uba Debre Tsehay	19.8	1	81.2
Wulbareg	9.5	1	32.8

Table 9: Woreda's and their respective costs for one well for percentage 70%

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	24.7	1	106.1
Bilate Zuria	20.23	1	87.8
Kucha	111.3	1	463.3
Maji	305.4	1	1512.0
Melekoza	92.0	1	405.4
Menit Shasha	134.0	1	643.1
Salamago	419.4	1	1945.3
South Ari	123.7	1	601.9
Uba Debre Tsehay	95.4	1	405.0
Wulbareg	31.3	1	120.6

Table 10: Woreda's and their respective costs for one well for percentage 100%

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	3.4	1	6.3
Bilate Zuria	3.3	1	5.0
Kucha	19.2	3	54.0
Maji	11.7	2	36.3
Melekoza	9.4	2	22.4
Menit Shasha	8.2	2	17.2
Salamago	11.1	3	21.0
South Ari	2.3	1	0.6
Uba Debre Tsehay	15.8	2	49.1
Wulbareg	6.8	2	13.1

Table 11: Woreda's and their respective costs for multiple well algorithm for 50%

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	7.7	2	17.0
Bilate Zuria	5.9	1	18.0
Kucha	31.8	5	86.7
Maji	24.2	4	74.9
Melekoza	19.2	4	45.8
Menit Shasha	13.1	3	31.0
Salamago	15.5	4	31.1
South Ari	16.3	3	48.5
Uba Debre Tsehay	28.6	3	98.6
Wulbareg	11	3	21.5

Table 12: Woreda's and their respective costs for multiple well algorithm for 70%

Woreda	Cost in C	Nr of wells	Total distance in km
Basketo WP	20.6	3	68.3
Bilate Zuria	16.3	3	46.3
Kucha	60.6	9	176.0
Maji	90.2	16	262.6
Melekoza	47.9	8	127.4
Menit Shasha	54.1	9	151.4
Salamago	110.2	22	287.2
South Ari	54.9	9	160.9
Uba Debre Tsehay	48.4	6	166.6
Wulbareg	25.4	6	58.2

Table 13: Woreda's and their respective costs for multiple well algorithm for 100%

For easier comparison between different algoritms we now supply one table with all the results combined. In the top block we can see the costs in C, in the middle block the number of wells and in the bottom block the total distance from wells to villages in kilometer.

	One well, 50%	One well, 70%	One well, 100%	Multiple well, 50%	Multiple well, 70%	Multiple well, 100%
Basketo WP	3.2	5.9	24.7	3.4	7.7	20.6
Bilate Zuria	3.3	5.9	20.2	3.3	5.9	16.3
Kucha	16.2	33	111.3	19.2	31.8	60.6
Maji	13.4	24.8	305.4	11.7	24.2	90.2
Melekoza	6.9	13.2	92	9.4	19.2	47.9
Menit Shasha	7.6	30.6	134	8.2	13.1	54.1
Salamago	13.7	31	419.4	11.1	15.5	110.2
South Ari	2.3	8.9	123.7	2.3	16.3	54.9
Uba Debre Tsehay	8.4	19.8	95.4	15.8	28.6	48.4
Wulbareg	4.7	9.5	31.3	6.8	11	25.4

Basketo WP	1	1	1	1	2	3
Bilate Zuria	1	1	1	1	1	3
Kucha	1	1	1	3	5	9
Maji	1	1	1	2	4	16
Melekoza	1	1	1	2	4	8
Menit Shasha	1	1	1	2	3	9
Salamago	1	1	1	3	4	22
South Ari	1	1	1	1	3	9
Uba Debre Tsehay	1	1	1	2	3	6
Wulbareg	1	1	1	2	3	6

Basketo WP	5.3	17.5	106.1	6.3	17	68.3
Bilate Zuria	5	18	87.8	5	18	46.3
Kucha	67.8	144.9	463.3	54	86.7	176
Maji	56.7	113.7	1512	36.3	74.9	262.6
Melekoza	22.7	51.5	405.4	22.4	45.8	127.4
Menit Shasha	25.6	132.4	643.1	17.2	31	151.4
Salamago	57.4	141.5	1945.3	21	31.1	287.2
South Ari	0.6	31.3	601.9	0.6	48.5	160.9
Uba Debre Tsehay	26.9	81.2	405	49.1	98.6	166.6
Wulbareg	11.2	32.8	120.6	13.1	21.5	58.2

TABLE 14: Costs, number of wells and total km of pipeline for different Woreda's for different algorithms