

Reviewer Comments:

Reviewer 1

- 5 A. page 1, line 16 - "A widely applied approach is that of Genetic Algorithms..." the authors should cite the Maier et al (2014) study here (references to it do appear later in the paper). This Maier paper is a thorough review of the state of the field, so it can be used to contextualize the work. In fact, I would say that modern terminology calls these approaches "evolutionary algorithms", of which the generic Genetic Algorithm is simply one approach.
- B. page 1, line 18 - If the authors are referring to "tricks" they should cite a few examples. Otherwise, the authors are giving the false impression that they are the first to explore these genetic algorithm parameters, which I do not believe they are.
- 10 C. page 2 - In this list, I think it is important to mention the fact that some of these evolutionary algorithm techniques are not genetic algorithms at all, but rather other algorithms such as differential evolution. That could perhaps be a sixth item in the list.
- D. section "Problem specific variators" - I was somewhat confused by this section, since in it the authors have discussed some of the new operators they are looking at, but it seemed to overlap with the subsequent "Tests" section. I think the organization of these methods sections could use some fixing.
- 15 E. A citation for EPANET should be given, and if the model is freely available, that should be noted.
- F. page 3, line 12 - The authors have consistently forgotten the formatting of 10^5 ; this should be fixed throughout.
- G. A brief mention of how the optimization was carried out would be appropriate, especially in the population size. The population size has been shown to be a very important parameter for assessing GA success.
- 20 H. The "Tests" section should also briefly comment on the performance criterion used (namely, the best optimal solution found so far as a function of time during the run)
- I. The authors should give some general guidance as to how to assess the "significance" of the objective function value. For example, other than the drastic reduction in objective of the first three tests as compared to the last three, is there a difference? In other words, it appears as though the first three tests are the worst, but can we tell the difference with the other 9 tests?
- 25

Reviewer 2

- J. Authors propose the use of two heuristic variations for mutation operator: flatiron and proximity. A doubt exists about their authorship. Do the authors propose these operators? Are they adapted from previous works? In the first case, the operators should be described with more detail. On the contrary, if the operators were taken from other authors, their work should be referenced.
- 30

- K. Proximity operator use values of a number of neighbours. The concept of neighbour in water system is a topological one. Topological operations might be computational heavy and time to reach a solution might increase. Some comment about this fact could be said.
- L. Authors made a set of tests with “[: :] 1x10⁵ function evaluation [: :]”. Later, in Figure 4 up to 1000 generations are represented. It can be deduced that the population used for the algorithm is 100 individuals. However, since Genetic Algorithm is strongly based on population, the population must be explicitly mentioned in the paper.
- M. Apart from the previous comments, other little changes could help improving the paper:
- In line 1-2, page 2, the year of publication for El-Mihoub et al. (2006) might be added.
 - In line 13, page 2, there is a mistake: “[: :] to construct an model [: :]” should be changed by “[: :] to construct a model [: :]”
 - All references to the number of function evaluations (1x10⁵) or fitness values (8.8x10⁵) should be written with superscript or using the “⁵” symbol. This affects to lines 12, 22, 24 and 29 in page 3; and lines 9 and 12 in page 4.
 - In line 6, page 3 there is a word missing: “[: :] 584 nodes, 491 ? and 1 reservoir [: :]”. I guess the word links or pipes is missing. Anyway, please check this number since a network with 584 nodes and (only) 491 links cannot be run with EPANET. If this check is performed, Figure 4 caption should also be corrected.
 - In line 2, page 4 the test 2 is referred to as RM-NPC2. However, in Figure 2 is referred to as RMNPC2 and in Figure 4 is referred to simply as test 2. Please use the same reference throughout the paper.
 - Caption for Figure 4 refers part (a) as results corresponding to test 2 and part (b) as results for test 12. However, the part (a) has a value for the objective function between 5.2e5 and 7.3e5 (coherent with the results of test 12, 5.63e5). It happens the same with part (b), values close to 9e5, more likely to correspond to test 2, 8.8e5. Probably it is just a question of flipping the caption for parts (a) and (b).
 - (This is only my personal point of view. Please forget if you disagree). In scientific texts, it is recommended to use a non-personal language. For that reason I encourage the authors to change the personal style in line 22, page 1 (“[: :] In this paper, we [: :]”) and line 13, page 4 (“[: :] consulting projects, we will [: :]”).

Author responses:

We thank the reviewers for their thorough suggestions. The responses to the individual point follow below.

To reviewer 1:

- 5 A. We have added the reference to Maier to this point in the text and modified the text to include the broader term of evolutionary algorithms.
- B. We have added the reference to El-Mihoub's review of HGA variants to this point of the text. We have also rephrased some parts of the introduction to avoid giving the impression of claiming to be the first to investigate this.
- C. We have added this insight that the hybridization changes the GA's place in the EA taxonomy to the end of the list.
- 10 D. We have restructured the methods section accordingly.
- E. We have added the reference. The network model is property of a Dutch utility and is not freely available.
- F. We have reformatted to 10^5 accordingly.
- G. We have added the various settings of the algorithm to the relevant part of the methods section.
- H. We now explicitly note the performance criterion in the same section.
- 15 I. We have added some thoughts about this to the discussion.

To reviewer 2:

- J. Now describe both variators in some more detail, and have rephrased to stress that only one of them is custom made.
- 20 K. We have added a comment on the retrieval of neighbours. (This was added to the description of the flatiron mutator, which uses neighbours, whereas the proximity mutator does not).
- L. We have added this information to the methods section.
- M. We agree with the reviewer on all items mentioned, and have modified the text accordingly.

Revised Manuscript with track changes:

Technical note: Problem specific variators in a genetic algorithm for the optimization of drinking water networks

5 Karel van Laarhoven¹, Ina Vertommen¹, Peter van Thienen¹

¹KWR Watercycle Research Institute, Nieuwegein, 3430 BB, the Netherlands

Correspondence to: Karel van Laarhoven (karel.van.laarhoven@kwrwater.nl)

Abstract. Genetic algorithms can be a powerful tool for the automated design of optimal drinking water distribution networks. Fast convergence of such algorithms is a crucial factor for successful practical implementation at the drinking water utility level. In this technical note, we therefore investigate the performance of a suite of genetic variators that was tailored to the optimisation of a least-cost network design. Different combinations of the variators are tested in terms of convergence rate and the robustness of the results during optimisation of the real world drinking water distribution network of Sittard, the Netherlands. The variator configurations that reproducibly reach the furthest convergence after 10^5 function evaluations are reported. In the future these may aid in dealing with the computational challenges of optimizing real world networks.

15 1 Introduction

Optimization techniques have been applied to the design (or more specifically, the dimensioning) of water networks for decades (see Bieupoude et al., 2012 and De Corte and Sørensen, 2013, for overviews). A widely applied approach is that of Genetic Algorithms (GA) (Holland, 1975; Goldberg, 1989) and other members of the overarching family of evolutionary algorithms (EA) (Maier et al., 2014). Though the classic genetic algorithm is very powerful, a number of approaches have been developed to improve convergence in water network design problems. More specifically, the various mechanisms of the genetic algorithm are commonly expanded, replaced or combined with heuristic tricks or complete heuristic algorithms to improve performance (as reviewed by e.g. El-Mihoub et al., 2006). The algorithms which include these are commonly referred to as Hybrid Genetic Algorithms (HGA) or Memetic Algorithms (MA). An overview of these approaches is summarized below. Following one of the approaches, a selection of custom heuristic variators has been implemented in Gondwana, a generic optimization tool for drinking water networks (Van Thienen and Vertommen, 2015). In this paper, ~~we describe these~~ these variators are described and it is demonstrated how they contribute to significantly faster convergence in a sample optimization problem case study.

2 Hybrid Genetic Algorithms

A cornerstone of the HGA approach (Krasnogor and Smith, 2005; El-Mihoub et al., 2006) is the observation that classic GA are especially well suited for quickly locating global optima in the solution space, but subsequently have difficulty converging to the optimum locally within a reasonable number of iterations. To mitigate this, GA are augmented with Local Search (LS) methods. These are algorithms that iteratively modify a given solution towards a predefined optimization criterion. LS methods find local optima relatively quickly, but are generally unable to escape this local optimum in favor of a possibly different global optimum. The resulting HGA therefore profits from the strengths of both techniques and yields better solutions. El-Mihoub et al. (2006) identify the following general ways in which GA capabilities can be expanded through hybridization:

1. A GA solution can be improved by running it through a LS method. This can be done to improve the final solution of the GA. Alternatively, the LS algorithm can be applied to refine intermediate solutions to promote the representation of different promising areas of the solution space within the population.
2. The number of iterations that are needed to achieve convergence can be reduced by replacing classic genetic operators with different ones to guide the search through the solution space.
3. Alternatively, system specific knowledge can be used to modify the genetic operators in such a way that they only result in viable solutions. This does not guide the search, but prevents time loss due to the evaluation of many illegal solutions, which may arise from random variations in heavily restricted GA problems.
4. The population size needed to achieve convergence can be reduced by dynamically controlling candidate selection with a LS method.
5. System specific knowledge can be used to construct a model to quickly approximate the results of fitness functions that are expensive to calculate, speeding up the evaluation of the GA objectives.

It is worth noting that the possible resulting HGA form a broad class of algorithms, and that individual HGA might come to fall under categories different from GA within the taxonomy of EA (Calegari et al., 1999).

3 Case study

3.1 Problem specific variators

Within the field of water network design optimization, algorithms that guide the GA to reduce the size of the search space is a specific challenge in current research (Maier et al., 2014). Table 1 lists a collection of genetic operators that was composed to tune a GA to the optimization of a least cost design (Alperovitz and Shamir, 1977; Savic and Walters, 1997). This type of problem varies pipe diameters throughout the network in search of the minimum network costs while achieving a minimum pressure at each node. In addition to several classic GA variators (Holland, 1995; Goldberg 1989), two heuristic variators are used that were constructed with the goal of a least cost design in mind. In terms of the classification of hybrid metaheuristics by (Talbi, 2002), the resulting HGA is a low-level teamwork hybrid.

Formatted: Heading 2

The heuristic flatiron mutator was custom made to enhance convergence according to approach 2 in the list above. It guides the search past a type of artefact that commonly occurs in intermediate solutions for the least cost design problem. This artefact occurs when classic mutation causes a larger diameter pipe to be surrounded by smaller diameter pipes, which is hydraulically insensible. These artefacts can take a long time to disappear through random mutation only. The flatiron mutator speeds up convergence by detecting and 'smoothing out' these artefacts every iteration as follows:

1. for the mutating pipe, obtain the neighbour IDs from a lookup table with neighbouring pipes per pipe (it is worth noting that this lookup table is created at the start of the optimization, thereby limiting its impact on computation);
2. if the pipe connects to exactly 1 or 2 neighbouring pipes, compare the diameter of the mutating pipe to those of its neighbours;
3. if the mutating diameter is larger than the diameter of all neighbours, reduce it to the largest diameter among neighbours;

The heuristic list proximity mutator enhances convergence according to approach 3 in the list above. It is equivalent to the classic 'creep mutator' (Sivanandam & Deepa, 2007): it functions as the regular random mutation of a single pipe diameter, except that the possible outcomes of the mutation are limited to values close to the value prior to mutation. - It limits the possible outcomes of a mutation to diameters close to the original value. This mutator is typically used because large deviations from the original diameter are likely to cause hydraulically inviable solutions.

In order to evaluate the influence of the developed problem specific variators, a series of tests was performed on a case study.

3.2 Case study network

In order to evaluate the influence of the developed problem specific variators, a series of tests was performed on a case study.

The case study consists on the design of part of the existing drinking water distribution network of the Dutch village Sittard. The network has a total length of 10.8 km and has 1000 connections, including connections to a school, a residential building with 32 apartments and a care farm for mental patients. The network is fed by a single reservoir and has a mean total demand of 15 m³/h. The network is represented by an EPANET (Rossman, 2000) model consisting of 584-583 nodes/junctions, 491-491 pipes, 140 valves and 1 reservoir. The network is displayed in Fig. 1.

3.3 Optimization

For the design of the network, the minimization of the product between pipe diameter and pipe length (surrogate for costs) was considered as the objective, constrained by a minimum pressure at each node equal to 34 meters. The decision variables were the pipe diameters, that could be chosen from the set of available diameters summarized in Table 2. A population of 100 individuals with an elitism rate of 15% was used for each optimization. Selection between candidates was achieved through tournament selection with a tournament size of 2 and with the objective function - the product of pipe length and diameter - as the performance criterion.

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0,63 cm + Indent at: 1,27 cm

Formatted: Heading 2 Char

Formatted: Heading 2

Formatted: Superscript

Formatted: Heading 2

3.4 Tests

A total of 16 tests were performed, wherein different values-rates for the specific variators were considered in order to assess their influence on the network design results. For each test, a total of 1×10^5 function evaluations and 10 runs were performed, i.e., each test was repeated 10 times in order to assess the mean, standard deviation, best and worst results obtained for each test. Table 3 provides an overview of the different tests, including the considered variator values and obtained results. Tests 1 and 2 consist of different value-rate combinations of the classical-naïve random mutation (RM) and one-point crossover (NPC). The heuristic proximity mutation was added in tests 3 to 8. Different values-rates for the heuristic flatiron mutation were considered in tests 9 to 14. Tests 15 and 16 further explore the influence of the one-point crossover on the performance of the algorithm.

4 Results

From the obtained results (Table 3) it is clear that the consideration of the heuristic flatiron mutation (FM) and proximity mutation (LPM) significantly improved the obtained results for the optimization problem. These results are graphically reported in Fig. 2.

Considering only the naïve random mutation and one-point crossover, the best results after 1×10^5 function evaluations were achieved with a mutation rate equal to 0.05 and a crossover rate equal to 0.95. In this case the average objective function value was 8.8×10^5 . Adding a proximity mutation does not improve the results, but considering only a proximity mutation and no random mutation, has a significant influence on the outcomes. Considering-With only the proximity mutation, the best results after 1×10^5 function evaluations were achieved for a proximity mutation rate of 0.05. With this value, the best results were achieved for both the mean, best and worst values for the objective function. Fig. 3a illustrates the influence of this variator on the computed objective function values.

Adding a flatiron mutation further improved the obtained results. The best results after 1×10^5 function evaluations, on average, were obtained for a combination of a crossover rate of 0.95, with a proximity mutation rate equal to 0.05 and a flatiron mutation rate equal to 0.8. The best result within one test only was obtained for a slightly lower flatiron mutation rate, equal to 0.7. Fig. 3b illustrates the influence of this flatiron mutation rate on the obtained objective function values.

The effect of the problem specific variators can also clearly be seen on the shape of the convergence curves. Fig. 4 illustrates the mean, mean \pm standard deviation, and mean $\pm 2 \times$ standard deviation convergence curves for tests number 2 (RM-NPC2) and 12 (FM4). The proximity and flatiron mutations lead to smoother curves and a faster convergence. The standard deviation between results of the different runs is also much lower, which means that the results are more stable.

Formatted: Heading 2

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

5 Discussion and conclusions

The results presented in this paper clearly illustrate the value of applying heuristic, non-classical variators in drinking water distribution system design optimizations using genetic algorithms. While the difference between the test with random mutation and the other tests is especially noticeable in Fig. 2, it is worth noting that the smaller differences between the other individual tests indicate a significant difference in convergence as well. In Fig 4a, for instance, it can be seen that, in FM4, the average objective function value of 6×10^5 was reached in around 700 generations, about 1.4 times faster than in LPM2, LPM5 and LMP6.

In ~~our~~ the tests, the combination of a low rate for the proximity mutation with a high rate of the flatiron mutation leads to the best results after 1×10^5 function evaluations (test numbers 13:FM5 and 14:FM6), i.e. the fastest convergence. All tested combinations which include either the flatiron or the proximity mutation exhibit a similar or worse performance. Albeit slower, particularly stable results were obtained with the proximity mutation (rate=0.1) and no flatiron mutation. These runs show the smallest standard deviation in the results after 1×10^5 function evaluations.

In ~~our~~ future research and consulting projects with Gondwana, we will continue to use this combination of variators will be used in order to deal with the computational challenges of larger real world networks.

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

15 Acknowledgments

The authors wish to thank Henk Vogelaar from Waterleiding Maatschappij Limburg (WML) for providing the Sittard network model used in the calculations.

References

Bieupoude, P., Azoumah, Y., and Neveu, P.: Optimization of drinking water distribution networks: Computer-based methods and constructal design. *Comp. Env. Urb. Sys.* 36, 434-444, 2012.

Calegari, P., Coray, G., Hertz, A., Kobler, D., and Kuonen, P.: A taxonomy of evolutionary algorithms in combinatorial optimization. *J. Heur.* 5, 145-158, 1999.

De Corte, A., and Sorensen, K.: Optimisation of gravity-fed water distribution network design: a critical review. *Europ. J. of Oper. Res.* 228, 1–10, 2013.

25 van Thienen, P., and Vertommen, I.: Gondwana: A Generic Optimization Tool for Drinking Water Distribution Systems Design and Operation, *Proc. Eng.*, 119, 1212-1220, 2015.

Holland, J.: *Adaptation in Natural and Artificial Systems*, The University of Michigan, Michigan, USA, 1975.

Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, ISBN:0201157675, 1989.

Krasnogor, N., and Smith, J.: A tutorial for competent memetic algorithms: model, taxonomy and design issues, IEEE Transactions on Evolutionary Computation, 9, 474-488, 2005.

El-Mihoub, T.A., Hopgood, A.A., et al.: Hybrid Genetic Algorithms: A Review. Eng. Let., 13, 124-137, 2006.

Maier, H.R., Kapelan, Z., et al.: Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environ. Mod. Soft., 62, 271-299, 2014.

Alperovitz, E., and Shamir, U.: Design of Optimal Water Distribution Systems. Wat. Resource. Res., 13, 885-900, 1977.

Savic, D.A., and Walters, G.A.: Genetic Algorithms for Least-Cost Design of Water Distribution Networks, J. Wat. Res. Plan. Man., 67-77, 1997.

Talbi, E.G.: A Taxonomy of Hybrid Metaheuristics. J. Heur., 8, 541-564, 2002.

10 Sivanandam, S.N., and Deepa, S.N.: Introduction to Genetic Algorithms, Springer, ISBN: 9783540731894, 2007.

[Rossman, L.A.: EPANET 2 User's Manual. United States Environmental Protection Agency, Cincinnati, U.S.A, 2000.](#)



15 **Figure 1: EPANET model of the drinking water distribution network of Sittard (Netherlands), consisting of 584-583 nodes/junctions, 491 links-pipes, 140 valves and 1 reservoir.**

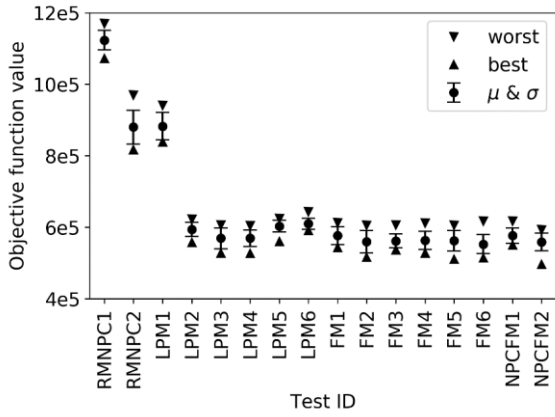
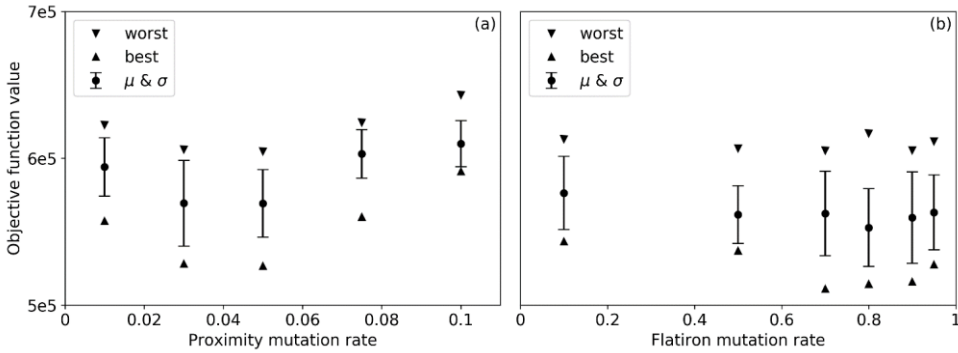


Figure 2. Overview of the obtained results for the different tests.



5 Figure 3. (a) Influence of the proximity mutation rate on the obtained objective function values in tests LPM2, LPM3, LPM4, LPM5 and LPM6. (b) Influence of the flatiron mutation rate on the obtained objective function values in tests FM1, FM2, FM3, FM4, FM5 and FM6.

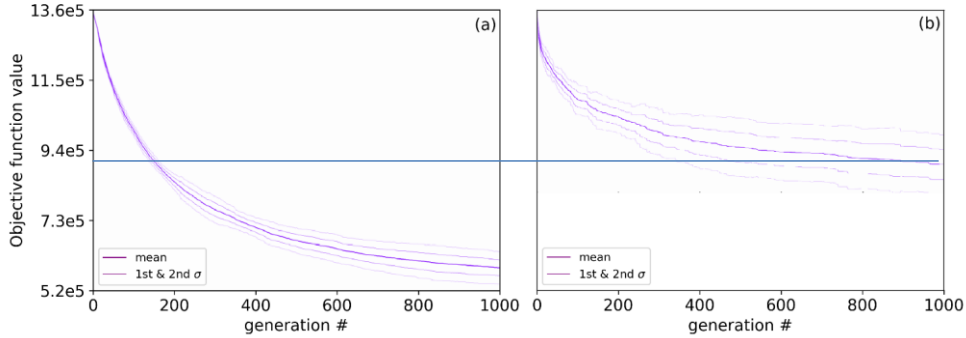


Figure 4. Convergence curves (mean, first and second standard deviations of 10 runs) obtained for tests **2-RMNPC2** and **12-FM4**. (a) **One-point crossover, proximity mutation and flatiron mutation (FM4)**. (ab) **Random mutation and one-point crossover (test 2-RMNPC2)**. (b) **One-point crossover, proximity mutation and flatiron mutation (test 12)**.

5

Table 1. Specific variators. Types include mutators (m) and crossover (c), classic (C) and heuristic (H).

variator	acronym	type	description
random mutation	RM	mC	assign a random value within a prescribed range to a parameter or subdivision index
n point crossover	NPC	cC	mix lists of decision variable (attribute/parameter) values
selection mutation	SeM	mC	random selection from a list of predefined values
flatiron mutation	FM	mH	give an object the same value as (or the minimum or maximum of) its neighbours on both sides, provided each side has only one neighbour
list proximity mutation	LPM	mH	random selection from n nearest neighbour values in an ordered list of allowed values

Table 2. Available pipe diameters (mm).

0	13.2	21.2	36	42.6	58.2	66	72.8
87.3	101.6	130.8	147.6	163.6	190	200	

10 Table 3. Problem specific variator values considered in the different tests and obtained results for 10 runs with 1×10^5 function evaluations each. The best results are indicated in bold.

Test	Crossover rate (N)	Random Mutation (N)	Proximity mutation (H)	Flatiron mutation (H)	Mean	Std.	Best	Worst
1 RM-NPC1	0.9	0.1	0	0	1.12E+06	2.76E+04	1.07E+06	1.17E+06

2	RM-NPC2	0.95	0.05	0	0	8.80E+05	4.75E+04	8.18E+05	9.70E+05
3	LPM1	0.95	0.05	0.01	0	8.83E+05	3.83E+04	8.39E+05	9.41E+05
4	LPM2	0.95	0	0.01	0	5.94E+05	1.99E+04	5.57E+05	6.23E+05
5	LPM3	0.95	0	0.03	0	5.69E+05	2.92E+04	5.28E+05	6.06E+05
6-5	LPM4	0.95	0	0.05	0	5.69E+05	2.30E+04	5.27E+05	6.05E+05
7	LPM5	0.95	0	0.075	0	6.03E+05	1.65E+04	5.60E+05	6.24E+05
8-6	LPM6	0.95	0	0.1	0	6.10E+05	1.57E+04	5.91E+05	6.43E+05
9-7	FM1	0.95	0	0.05	0.1	5.76E+05	2.49E+04	5.43E+05	6.13E+05
10-8	FM2	0.95	0	0.05	0.9	5.60E+05	3.11E+04	5.16E+05	6.05E+05
11-9	FM3	0.95	0	0.05	0.5	5.62E+05	1.97E+04	5.37E+05	6.07E+05
12-10	FM4	0.95	0	0.05	0.95	5.63E+05	2.55E+04	5.28E+05	6.12E+05
13-11	FM5	0.95	0	0.05	0.7	5.62E+05	2.88E+04	5.11E+05	6.05E+05
14-12	FM6	0.95	0	0.05	0.8	5.53E+05	2.64E+04	5.14E+05	6.17E+05
15-13	NPC-FM1	0.9	0	0.05	0.8	5.77E+05	2.16E+04	5.50E+05	6.17E+05
16-14	NPC-FM2	0.8	0	0.05	0.8	5.59E+05	2.46E+04	4.97E+05	5.93E+05