

Document downloaded from:

<http://hdl.handle.net/10251/64597>

This paper must be cited as:

Izquierdo Sebastián, J.; Montalvo Arango, I.; Campbell, E.; Pérez García, R. (2015). A hybrid, auto-adaptive, and rule-based multi-agent approach using evolutionary algorithms for improved searching. *Engineering Optimization*. 1-13. doi:10.1080/0305215X.2015.1107434.



The final publication is available at

<http://dx.doi.org/10.1080/0305215X.2015.1107434>

Copyright Taylor & Francis: STM, Behavioural Science and Public Health Titles

Additional Information

A hybrid, auto-adaptive, and rule-based multi-agent approach using evolutionary algorithms for improved searching

Joaquín Izquierdo^a, Idel Montalvo^b, Enrique Campbell^a, Rafael Pérez-García^a

^a*FluIng-IMM, Universitat Politècnica de València, Cno. de Vera s/n, 46022, Valencia, Spain
{jizquier, encamgo1, rperez}@upv.es*

^b*Ingeniousware GmbH, Bahnhofstraße 4a, 76137 Karlsruhe, Germany
imontalvo@ingeniousware.net*

A hybrid, auto-adaptive, and rule-based multi-agent approach using evolutionary algorithms for improved searching

Joaquín Izquierdo^{a1}, Idel Montalvo^b, Enrique Campbell^a, Rafael Pérez-García^a

^a*FluIng-IMM, Universitat Politècnica de València, Cno. de Vera s/n, 46022, Valencia, Spain
{jizquier, encamgo1, rperez}@upv.es*

^b*Ingeniousware GmbH, Bahnhofstraße 4a, 76137 Karlsruhe, Germany
imontalvo@ingeniousware.net*

Selecting the most appropriate heuristic for solving a specific problem is not easy. The reasons are manifold. This article focus on one of these reasons: traditionally, the solution search process has operated in a given manner regardless of the specific problem being solved, and the process has been the same no matter the size, complexity, and domain of the problem. To cope with this situation, search processes should mould the search into areas of the search space that are meaningful for the problem. This article builds on previous work in the development of a multi-agent paradigm through the use of techniques derived from knowledge discovery (data mining techniques) on databases of so-far visited solutions. The aim is to improve search mechanisms, increase computational efficiency, and enrich through rules the formulation of optimization problems – while reducing the search space and catering to realistic problems.

Keywords: evolutionary optimization, data mining, rule extraction, PRIM

1. Introduction

Many optimization problems in engineering have recently transcended the spectrum of what might be called standard optimization (Nicklow et al., 2010; Nguyen, Hartmann, and König, 2012). The well-known reasons are manifold. Among others, problems exhibit non-linearities and lack good conditions of differentiability, are multimodal, and are conditioned by various hard-to-handle constraints. For these problems, global optimization is challenging, and often unattainable. In addition, the numerical methods used to evaluate optimization criteria and constraints are often severely affected by the noise and uncertainty inherent in real-world problems. Consequently, new methods are needed which allow virtually any optimization criterion and constraint and proper handling of the singularities (non-linearity, coexistence of mixed variables, etc.) associated with objective and constraint functions. Such problems have to be addressed through suitable optimization techniques in combination with simulation and data analysis processes.

Evolutionary algorithms (EAs) are stochastic optimization techniques that avoid various mathematical complications (Reed et al., 2013). EAs handle a number of variable populations of solutions for the problem in hand, in pursuance of the best individual(s) who represent the best solution(s) to the problem. The flexibility introduced by evolutionary algorithms has enabled the use of virtually any objective function for evaluating solutions, even when these evaluations require running complex mathematical and/or procedural simulations of the systems under analysis. The literature is very extensive in examples, in particular, in urban hydraulics (Liong and Atiqzama, 2004; Geem, 2006; Izquierdo et al., 2008b; Jin et al., 2008; Montalvo et al., 2010; Shean and McBean, 2010; Bei and Dandy, 2012; Berardi, Laucelli, and Giustolisi, 2012; Wu and Behandish, 2012; Montalvo et al., 2014; Marchi et al., 2014). See also (Zheng, Simpson, and Zecchin, 2013a) for a review.

Nevertheless, each algorithm has its own drawbacks and is better adapted to certain problems than to others. The heuristics behind a certain evolutionary algorithm endow its elements with specific capabilities for efficiently solving some kinds of problems, while being inefficient with problems of a different nature. For example, particle swarm optimization (PSO) experienced

¹ Corresponding author. Email: jizquier@upv.es. Tel.: +34 628028804

major difficulties in the design of water distribution systems (WDS) – the case study of this article – when the candidate diameters for the pipes were listed in an unordered way (Montalvo, 2011). In these cases, the main reason is clearly the lack of a logical order in the options, which produces overly abrupt roaming through the search space.

Clear improvements must be made for EAs to be able to cater to realistic problems. This article considers a multi-agent approach that wraps three elements: hybridization; self-adaptive parameters; and incorporation of knowledge into the search process. In this article, focus is mainly on this last element and its integration in a multi-agent structure in which the other elements are already integrated.

Firstly, it seems reasonable to think that there is no solution to the problem of obtaining a strategic combination of methods that enables any problem to be robustly and efficiently resolved in terms of global optimality. One of the approaches found in the literature in this direction is based in hyper-heuristics (Burke et al. 2009, 2010), where domain-specific low-level heuristic elements, such as mutation operators are created automatically (McClymont et al., 2013, 2014) using such techniques as genetic programming (Koza, 1992). Another approach consists in the joint (hybrid) use of different optimizers, such as PSO, ACO, GA, etc., in the same platform through swarms that share all the resources in a multi-agent structure. There are numerous examples in the literature of multiple optimizers being used together (Onwubolu and Babu, 2004; Javadi, Farmani, and Tan, 2005; Hwang and He, 2006; Vrugt and Robinson, 2007; Shang, Zhao, and Shen, 2009). The proposal in this article considers hybridization of heuristics of different nature within a joint multi-agent based platform (Montalvo et al., 2014).

Also, mechanisms to determine the most appropriate methods for suitable parameter handling (tending toward self-adaptation) (Maier et al., 2014) are also needed. Various self-tuning algorithms have been developed (Vrugt and Robinson, 2007; Vrugt, Robinson, and Hyman, 2009; Hadka and Reed, 2012; McClymont et al., 2013; Johns et al., 2014). Adaptive and self-adaptive parameters are also considered in the proposal considered here (Montalvo et al., 2014).

However, as noted by various authors, the significant step will not only derive from combining various heuristics, using hyper-heuristics, or better adjusting parameters to the problem being solved, but also from more directly influencing the way the search is performed. To develop algorithms that adapt their behaviour to the problems being solved, so that the algorithms have more chances to succeed, may be of great interest. A way to achieve this is by combining evolutionary optimization with the introduction of (among other knowledge carriers) rules based on the domain of the problem being solved.

Examples of uses of heuristic knowledge in combination with the search algorithm may be found in the recent literature. For example, in the optimal design of water distribution networks (WDNs), Zheng, Simpson, and Zecchin (2013b) use shortest path methods to reduce the search space by eliminating areas of the search space obtained through the use of classical optimization techniques. Even though initialization with good solutions does not change the optimization algorithm, computational efficiency is often improved by this type of preconditioning. Domain specific rules may be used for this task before starting an EA run. See (Kang and Lansey, 2011; Keedwell and Khu, 2005; Bi, Dandy, and Maier, 2014), among others.

In (Montalvo, 2011) and (Montalvo et al., 2014) problem domain knowledge is used to produce rules that are incorporated into the solution search process. In this approach, during the search, the ranges of certain decision variables are dynamically modified (reduced, if possible) depending on the values already given to other decision variables. Rules developed in this research were exclusively derived from specific knowledge of the problem in hand, specifically, water distribution system design. For example, for the sizing of pipes in a WDN, it is a rule of thumb to reduce or maintain the diameter of the pipes as the system progresses from upstream to downstream. In some sense, using domain rules can be considered a method to better define, in a non-standard way, the objectives and constraints of a given problem.

These ideas work quite well, although there are a couple of drawbacks. Firstly, rules are “hard-coded”: changing the existing source code or adding new code to support the algorithms is

necessary for rule enforcement. Secondly, it is hard to discover new rules to help further improve the search process. Developing rule-based agents within EAs to improve their performance requires the active participation of specialists from the problem domain. It is difficult to develop good rules without a good understanding of the problems in the context of their domain. But even for people with a deep understanding of the problem domain it is hard to define rules that can be generalized and applied to work in combination with evolutionary techniques. It is much easier to analyse how to improve the search in a specific problem instance than to define a generalized way to do it. Even if a generalized way is found, it will need adjusting to be expressed in a programming language and effectively used.

This article concentrates on applying data mining (DM) techniques to the set of solutions evaluated after several generations of a single run of an EA in order to extract rules initially intended to be used by future evolutionary generations so that the search is improved. The objective is to create, together with appropriate computational techniques and simulation, better conditions for optimization. Other attempts in this same direction (Michalski, 2000) have been performed previously. For example, in (Pelikan et al., 2002) use is made of probability distributions of promising solutions to guide the exploration; in (Jourdan et al., 2005) rule induction is used to produce rules to create new individuals. This article uses PRIM (Friedman and Fisher, 1999), and the rules obtained are automatically incorporated in the multi-agent platform.

The global objective of this article is to innovate in solution search quality in complex real optimization problems using a (non-standard) computational methodology that is based on the multi-agent system paradigm. This methodology integrates: various heuristics; auto-adjusting parameters; and knowledge carriers based on problem-based knowledge and on the exploration of the analysed solutions. This methodology will also enable the control and interaction of one or more users as high level agents in the system so that the optimization process is interactive, and can, thus, be guided more efficiently. A computational platform, ASO (for agent swarm optimization), already built by the authors (Montalvo et al., 2014) that uses these ideas has been suitably modified and now enlarged (with the incorporation of knowledge obtained during the optimization process) to offer the possibility of exploring emergent high-level behaviour in rule-based optimization processes based on low-level interaction.

The road map of the article is the following. In the following section, after shortly introducing the multi-agent approach, which includes all the elements of the present proposal, and the use of problem-dependent rules, it is reasoned how other kinds of knowledge extracted from the evolution history may also be useful, and focus on rule extraction mechanisms. Section 3 presents a case study. Finally, Section 4 provides conclusions.

2. Methodology

The enclosing structure of the proposal in this article is an agent-based (Sycara, 1998; Wooldridge, 2009) approach. Multi-agent systems (MASs) have provided an appropriate theoretical and computational basis for solving various problems in different fields.

In addition to ASO (Montalvo et al., 2014), several articles have been published that address agent-based optimization for complex problems. Examples include (Weichhart et al., 2004) who develop a multi-agent optimization system for scheduling problems; while (Persson et al., 2005) use MASs combined with optimization techniques in dynamic distributions of resource allocation; and (Xie and Liu, 2009a,b) use MASs in combinatorial optimization.

ASO is a software platform over which the approach described in this article is built. The application of ASO to the most popular benchmarking problems in the WDN-design literature has produced the best known solutions for these problems (Montalvo, 2011). The authors have also used this package for addressing other real-world complex problems such as the design of waste water systems (Izquierdo et al., 2007); the calibration of a WDN (Izquierdo et al., 2008b); the optimal design of a biomass supply chain at regional level (Izquierdo et al., 2008a); and the

clustering of a water company database to classify pipes with the aim of rehabilitation (Díaz et al., 2008).

ASO hybridizes various algorithms in runtime on a single platform. Also, by introducing several parameters into the inner mechanisms of its various EAs, ASO uses adaptive and auto-adaptive parameters, thus leaving the manual effort that is usually made to initialize new optimization runs in the hands of the agents. Finally, agents are endowed with specific, problem-dependent rules purposely designed to heuristically deal with the solution process. These rules attempt to mimic the judgment of a human expert when approaching a solution to a problem. The next subsection focuses on the main contribution of this article, namely, using knowledge discovery to enrich the search.

2.1 Knowledge discovery enriched search

Evolutionary algorithms have not generally taken advantage of this use of rules, and so have been bound to analyse larger search spaces than necessary. The inclusion of rules may reduce search space by several orders of magnitude. This can also be seen as a way of improving the problem definition by including information that either cannot be easily expressed or depends on the current solution.

As a consequence of using this feature, solutions are efficiently obtained and are closer to reality. Efficiency derives from the fact that simply checking a number of usually simple rules avoids many expensive calculations or simulations (hydraulic simulations in the case study presented in this article). Finally, the fact that the rules have strong problem-dependent meaning definitely brings the solution nearer to reality (Montalvo, 2011).

This idea may be approached by combining the way evolutionary algorithms work with the introduction of knowledge extracted from a suitable database of solutions visited during previous steps of the optimization process. This idea is further developed next.

During the execution of EAs, typically the number of solutions evaluated represents a small percentage of the total solution space corresponding to the problem being solved. Nevertheless, the number of solutions evaluated is still considerable, and most evolutionary techniques use just a small proportion at a time. Many of the solutions evaluated during the search process are “forgotten” after one generation, and the combined experience of several generations is often poorly exploited.

DM techniques can enable a deeper insight into the many “good” solutions that have been glimpsed and then rapidly disregarded because they were dominated by better solutions during an ephemeral moment in the evolution process. Based on a database obtained by suitably recording those disregarded solutions, DM techniques can help better understand and describe how a system could react or behave after the introduction of changes.

The main proposal of this article is to integrate DM techniques within the evolutionary work as a step for dynamically generating knowledge that can be used to improve the efficiency of solution search processes.

The description of the process, summarized in Figure 1, is the following.

When initializing an evolutionary process only random solutions are generally available. At the beginning the algorithm uses its specific search mechanisms, perhaps aided by some clear problem-dependent rules, until some iteration point. At the same time, analysed solutions are stored in a suitable database. The knowledge extraction algorithm is then launched, thus generating a number of facts and/or rules. This new knowledge is automatically encoded. During a new batch of iterations these rules are applied, while a new database is generated. After this iteration, a new knowledge extraction process is launched that may take advantage of the old rules. The new knowledge is then enforced and the process is repeated until convergence. This will eventually accelerate convergence to the desired solutions.

Before descending to specifications, a number of factors influencing this process must be pinpointed. Section 3 will provide specific examples when presenting the case studies.

Firstly, if there are well-defined problem-dependent rules that can be clearly encoded, they should be implemented from the outset, thus forcing the process to produce solutions closer to reality. Advantage should also be taken of mechanisms as in (Bi, Dandy, and Mayer, 2014) that include problem domain knowledge during initialization.

Secondly, iteration points where the evolutionary process can be stopped may be devised in several ways, for example, after a fixed number of iterations or when a significant improvement has been achieved and stabilized.

(FIGURE 1 NEAR HERE)

Figure 1. Putting knowledge extraction and EA to work together

Thirdly, for a typical database, it may happen that some of the variables are irrelevant for discovering new rules since they exhibit an almost constant value in a very high percentage of registers. At a given stage of the evolution, these “constant” values may correspond either to optimal (target) values or to variables that have not been completely explored. A decision must be made according to the evolution stage. In early evolution stages these values may be disregarded since they very likely correspond to poorly explored solutions, thus corresponding to local minima from where the EA should escape by using its stochastic abilities. On the other hand, in advanced evolution stages, these values should be directly transformed into hard rules for those variables, since the solutions may correspond to so-far good solutions. In any case, these variables may be (temporally) eliminated from the database for the current situation.

Various data mining techniques may be used to extract knowledge from the database of the so-far analysed solutions. In general, any of these databases is composed of a collection X of n -dimensional vectors (corresponding to the decision variables) and a set of target vectors Y (corresponding to the objectives; in the case where there is only one objective, objects in Y are scalars). In this article, the PRIM rule extraction algorithm (Friedman and Fisher, 1999) is used.

2.2 Rules to tailor the search process

Data mining techniques have been successfully applied to areas that handle large volumes of data as tools to scan the available information and thus track down understandable and useful knowledge (Friedman and Fisher, 1999), (Hastie, Tibshirani, and Friedman, 2001), (Bouguessa et al., 2009), (Hsu and Chen, 2009). IF-THEN rules are a common way of passing knowledge, since they are both easy to understand and implement in software programs. Moreover, adaptive techniques enable new rules to be added and existing rules to be removed or modified. For this reason, rules are used by most existing techniques to produce knowledge (Gonçalves et al., 2006), (Kamwa, Samantaray, and Joos, 2009), (Hasperué, Lanzarini, and de Giusti, 2012).

Rules based on the domain of the problem can be enforced and amalgamated with evolutionary techniques. The use of rule-based agents is, for example, one of the principles in ASO, which includes rules for taking the EA in use closer to the problem being solved, as stated above.

A process that endows an evolutionary optimization process with the ability to dynamically generate rules aimed at being used to improve the efficiency of the solution search process is here explored.

In many applications, interesting associations among variables often occur at a relatively high concept level. For example, design patterns in a database made out of solutions for a water distribution network obtained during an EA run may not show any substantial regularity at the primitive data level, such as the individual solution fitness level, but may show some interesting

regularities at several high concept level(s), including, for example, relationships among certain variables, such as pipe diameters.

Given a database of solutions obtained during a run of an evolutionary optimization process, it would be interesting to discover associations among decision variables showing that the presence of some values for certain variables in a solution implies the presence of other specific values for other variables in the same solution. Some rule induction algorithms can automatically select attributes and ranges of attributes that occur in their premises.

Focus is here on the so-called subgroup discovery (Klosgen, 1996; Wrobel, 1997; Lavrac et al., 2004), which aims at finding patterns in the data corresponding to subgroups with interesting properties. This contrasts with the development of global models (such as classification trees or logistic regression models) that aim at a good global performance.

The patient rule induction method (PRIM) suggested by Friedman and Fisher (1999) is referred to as a “bump-hunting” algorithm. Bump-hunting algorithms are used to find regions in the input variable space that are associated with a relatively high (or low) mean value for the outcome(s). A region is described by conjunctive conditions using the input variables and is associated with the mean value of the output in that region.

PRIM can be used for questions of data analysis in which the analyst is interested in finding combinations of values for the input variables that result in similar characteristic values for the outcome variables. Specifically, a set of subspaces of the input variable space is sought within which the values of the output variables are significantly different from the average value over the entire domain.

This results in a very concise representation, since usually only a limited set of dimensions of the input space variable is restricted. That is, a subspace is characterized by upper and/or lower limits in just a few of the input dimensions.

PRIM is easy to interpret, does not need variable transformations and imputation for missing values, and no prior assumptions on data are necessary.

Several successful applications of the PRIM have been presented in various areas such as: geology (Friedman and Fisher, 1999); marketing (Friedman and Fisher, 1999); finance (Becker and Fahrmeir, 2001); medicine (Kehl and Ulm, 2005); bioinformatics (Cole, Galic, and Zack, 2003; Liu et al., 2004); process optimization (Chong and Jun, 2005; Chong and Jun, 2008); and human water use (Kwakkel and Timmermans, 2012).

3 Case study

In this paragraph the above ideas are applied to the design of a WDN. It studies the San José de las Lajas network, a small city in Cuba, to test PRIM. Also, the results obtained are discussed.

This system is a medium size real-world network with a fixed layout (see Fig. 3 for the layout) fed by a tank, and consisting of 273 pipes and 183 consumption nodes with distinct consumptions amounting to 176 l/s. A population of about 58,000 inhabitants is served by this network.

The design considers minimizing the investment cost while maintaining a pressure of 30 meters of water column at every consumption node. This pressure should be maintained even under failure conditions in any single pipe.

The set of available diameters consists of six different values ranging from 12 to 40 inches. Considering that the network has 273 pipes, the solution space of the problem will contain $6^{273}=2.72 \times 10^{212}$ possibilities.

As said, the objective for this case study tried not only to minimize the initial investment cost of the network, but also to maximize its tolerance against pipe failures, see (Martínez-Rodríguez et al., 2011). To this purpose, satisfaction of the stated minimum pressure in the network was

imposed not only for normal conditions, but also for failure conditions. This evaluation of the tolerance implied running the analysis of the network as many times as existing pipes plus one (normal condition) every time the objective functions were evaluated. That is to say, network analyses were performed without considering any pipe failure and then considering one pipe failure at a time. Simultaneous failures in pipes were not considered.

Four swarms with four different types of agents were introduced for optimizing the design of the water network taken as the case study: standard PSO particles; PSO-based agents that were dynamically changing the range of the decision variables based on the results of the PRIM algorithm; PSO-based agents that used rules introduced *a priori* for deciding diameter values trying to maintain or reduce them from upstream to downstream; and a fourth type of agent with similar characteristics to the third type, but also considering knowledge derived from PRIM. These different agents shared the principles of PSO but had their own rules for deciding the ranges where solutions would be searched for each decision variable.

A total of 25 executions were run under the same conditions. Each of these executions tried to improve the best solutions obtained until reaching 800 iterations without any improvement in the best solution found (termination condition). Iterations are understood as changes of position of all the agents participating in the search.

All the types of agents were trying to find a dominant solution and the most successful were able to reproduce (clone) themselves during the search. See clonation details in (Montalvo et al., 2014).

In the search for the Pareto front several agents are able to find a dominant solution. Nevertheless, there are agents acting as leaders of the remaining agents during the search (Montalvo et al., 2014). These leaders also belong to the Pareto front but are closer to the utopia point, where the best of all the objectives are found. After 25 runs, results showed that in 19 cases the agent leading the swarm was directly influenced by the performance of PRIM. Therefore, these agents dynamically moved the range of the decision variables following the results of the PRIM algorithm.

(FIGURE 2 NEAR HERE)

Figure 2. Pareto front obtained without (swarm 1) and with (swarm 2) PRIM

Figure 2 represent the difference between the swarms using PRIM and the swarms that did not include PRIM in their calculations. For the sake of clarity and since the main focus is on the differences between using PRIM or not, the four classes of swarms have been grouped into two, namely, those taking advantage of PRIM and those not considering the additional information provided by PRIM. In this figure, Swarm 1 consists of both particles from the standard PSO algorithm and the PSO algorithm using the rule ‘lower diameters downstream’ introduced *a priori*. Swarm 2 consists of particles from the standard PSO using, additionally, PRIM, and particles of PSO using the rule ‘lower diameters downstream’ introduced *a priori* and combined with PRIM.

The use of PRIM gives a competitive advantage to some particles. In effect, one has to note that particles using PRIM are the first to reach positions close to the utopia point and this means they become search leaders. Their leadership implies that they influence those particles that are not using PRIM, guiding them to improve their positions in the solution space. This indirect benefit enables particles not using PRIM to reach the Pareto front. Nevertheless, leadership positions and the closest positions have been mainly dominated by particles using PRIM and this reveals their competitive advantage. These particles show a better ability to dominate other particles when competing to be part of the Pareto front. It can also be seen that these particles are much closer to the utopia point as a group, because it is there where the swarm leaders are and, as consequence, where the “desirable” solutions can be found. Reducing the space to get closer to

those “desirable” solutions forces particles using PRIM to explore the zone near the utopia point much more than the rest of the particles that are not following the PRIM results.

These results show the convenience of using the knowledge obtained through data mining for guiding the search. Even when the partial results of the PRIM algorithm at early stages do not show a clear “understanding” of which rules to follow in order to improve the search, the repeated use and improvement of the results obtained with PRIM can successfully help agents in their search. It is worth noting here that agents just based on *a priori* rules did not perform better than the agents that tried to “learn” during the search process. Nevertheless, their participation and also the participation of the classic PSO agents were crucial for identifying niches of good solutions during the “learning” process. Agents using “knowledge” generated during several iterations were in a better situation to choose the best way towards good positions in the solution space.

The larger presence of agents using PRIM in the Pareto front is a clear sign that:

- This type of agent is able to find good positions in the objective space faster than the others. Agents taking more time to find Pareto positions cannot replace those already positioned and, as a consequence, will not be shown as part of the Pareto front. It is difficult to measure exactly how much faster agents using PRIM reach the Pareto front; it is even harder to extrapolate the measurement to different execution conditions. The intention of the tests executed in this research was not to focus on obtaining an exact performance measurement but to check which agents are able to arrive first in Pareto front positions.
- This type of agent is able to maintain dominant solutions. Only the dominance of others opens the possibility for integration into the Pareto front. In the long term, the other agents have fewer chances to succeed in dominating the solutions found by the agents supported by PRIM; that is why the presence in the Pareto front of agents not using PRIM (swarm 1 in Figure 2) is less notable – especially near the utopia point.
- This type of agent has a better chance of success in a wider scope of problems. The value brought by PRIM into the search process is not the result of rules created *a priori* by experts, but the result of the analysis of many solutions of a specific domain. This implies an adaptation of the relation between variable ranges to the problem itself, based on the analysis of many solutions and its corresponding projection in objective space.
- This type of agent is compatible with the *a priori* introduction of rules based on the criteria of experts in the problem domain. These rules, expressed also as relations among variable ranges, have a clear influence on the solutions obtained.

In the particular problem studied here, this last point may be emphasized. The resulting solutions show a relatively good organization of the diameters in the network. This enhances the influence of the rules introduced *a priori* devoted to maintaining or reducing diameters from upstream to downstream. Figure 3 clearly shows this effect. In effect, pipe diameters closer to the tank are bigger (represented by thicker lines) than the diameters found when moving away from the tank in any direction. The size of nodes in Figure 3 is also an indicator of pressure.

Maintaining or reducing diameters from upstream to downstream not only helps to improve the solutions obtained in an engineering sense but also reduces the solution space of the problem. Reducing the solution space of a problem increases the chances of finding good solutions and is an advantage when the evaluation of the objective function requires significant computational effort. The tolerance to failure, for example implied running the analysis of the network not only for normal condition, but also for the cases defined by every pipe of the network when out of service. This kind of evaluation, requiring more time and resources than a single evaluation, encourages the idea of taking and exploiting as much as possible all the results obtained from the objective function(s) in order to learn from them and then decide the next zone to explore. Simultaneous failure in more than one pipe or any other device was not considered in this

example. Including simultaneous failure and other objectives will further encourage the idea of reducing the solution space by identifying those zones where the combination of variable values does not make sense for the problem being solved.

(FIGURE 3 NEAR HERE)

Figure 3. Design of the network with problem-based rules and the influence of PRIM

4. Conclusions

This article has presented an approach based on distributed optimization to solve realistic optimization problems. The proposed approach is a structure that integrates several metaheuristics with different characteristics cooperating together. The parameters used by the various metaheuristics are adjusted adaptively and self-adaptively, using the mechanisms built into each optimization metaheuristic. The agents rely on a knowledge-based system that encapsulates both problem-dependent rules, and other relevant rules obtained through the integration of elements of artificial intelligence technologies such as DM. Specifically, rule extraction through PRIM has been considered.

To manage the complexity associated with the solution of realistic problems in engineering in general, and water resources in particular, used has been made of a multi-agent system, ASO, on which a distributed simulation and experimentation platform has been developed. This platform provides the appropriate general-purpose infrastructure for numerical optimization using various categories of heuristics.

The results for the case study presented a multi-objective optimization problem that includes nonlinear properties and mixed continuous-integer variables, and demonstrates that agent-based optimization is a promising and effective approach for solving sophisticated optimization problems in an adaptive and distributed manner. At least, the agent based approach is as good as the isolated execution of the same optimization methods used. However, the interaction provided by the exchange of information and the adaptive mechanisms may help the development of emerging and synergistic effects – and lead to more effective solutions for a wide variety of realistic optimization problems, and improved effectiveness in terms of convergence speed and reduction of the search space.

Several advantages of the proposed optimization approach can, in addition, be mentioned. Firstly, the search results are reused and selectively stored in a dynamic database from where likelihood areas and/or rules are extracted with the aim of improving the search efficiency; secondly, the defined environment is scalable with respect to its expandability by further optimization methods or technologies encapsulated in agents; thirdly, modelling complexity for implementing hybrid optimization approaches is relatively low, due to the fact that it is not the global behaviour of the environment that has to be defined explicitly, but only the local behaviour of the agents; as a result, and fourthly, the approach is directly applicable to other water resources and engineering problems. Last but not least, the interaction of one or more users with the platform provides improved usability of the computational framework. In contrast, the main drawback of this approach is the increasing complexity of design and implementation of software for agent-based optimization.

Considering the existing technology from the points of view of software and hardware, the capabilities for solving complex optimization problems have significantly increased when compared with the situation ten years ago. In the water field, as in other engineering problems, the biggest challenge today remains less in the capacity for solving a problem and much more in the ability to identify the proper problem statement corresponding to the real needs of the situation to be solved. For several years, problem statements have been constrained by the techniques available to solve them. Today, techniques such as the one presented here, open the door to a world of many possibilities. In the case of water distribution system design, for

example, good engineering solutions will not come from a super optimization algorithm but from a proper problem statement and an algorithm able to reproduce engineering thinking.

5. References

- Becker, U., and L. Fahrmeir. 2001. "Bump hunting for risk: a new data mining tool and its applications." *Computational Statistics* 16: 373-386.
- Bei, W., and G. C. Dandy. 2012. "Retraining of metamodels for the optimization of water distribution systems." In *The 14th Water Distribution Systems Analysis Conference*, ASCE, Adelaide, South Australia., pp. 36-47.
- Berardi, L., D. Laucelli, and O. Giustolisi. 2012. "A decision support tool for operational optimization in WDNEXL." In *The 14th Water Distribution Systems Analysis Conference*, ASCE, Adelaide, South Australia, pp. 48-65.
- Bi, W., G. C. Dandy, and H. R. Maier. 2014. "Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge." *Environmental Modeling and Software*, <http://dx.doi.org/10.1016/j.envsoft.2014.09.010>.
- Bouguessa, M., and S. Wang. 2009. "Mining Projected Clusters in High-Dimensional Spaces." *IEEE Transaction on Knowledge Data Engineering* 21 (4): 507-522.
- Burke, E. K., M. Hyde, G. Kendall, G. Ochoa, E. Ozcan, and J. Woodward. 2009. "A classification of hyper-heuristics approaches." In: *Handbook of Metaheuristics* (M. Gendreau & J.-Y. Potvin, eds.). International Series in Operations Research & Management Science: Springer, Berlin.
- Burke, E. K., M. Hyde, G. Kendall, G. Ochoa, E. Ozcan, and R. Qu. 2010. "Hyper-heuristics: A Survey of the State of the Art." *Computer Science Tech. Rep.*, University of Nottingham, Nottingham.
- Chong, I. G., and C. H. Jun. 2005. "Performance of some variable selection methods when multicollinearity is present." *Chemometrics and Intelligent Laboratory Systems* 78: 103-112.
- Chong I. G., and C. H. Jun. 2008. "Flexible patient rule induction method for optimizing process variables in discrete types." *Expert Systems with Applications* 34 (4): 3014-20.
- Cole, S. W., Z. Galic, and J. A. Zack. 2003. "Controlling False-negative Errors in Microarray Differential Expression Analysis: a PRIM Approach." *Bioinformatics* 19: 1808-1816.
- Díaz, J. L., M. Herrera, J. Izquierdo, I. Montalvo, and R. Pérez-García. 2008. "A Particle Swarm Optimization derivative applied to cluster analysis." In *Proceedings of 4th Biennial Meeting, iEMSs 2008: International Congress on Environmental Modelling and Software*, Barcelona, Spain.
- Friedman, J. H., and N. I. Fisher. 1999. "Bump hunting in high-dimensional data." *Statistics and Computing* 9: 123-143.
- Geem, Z. W. 2006. "Optimal cost design of water distribution networks using harmony search." *Engineering Optimization* 38 (3): 259-280.
- Gonçalves, L. B., M. M. B. R. Vellasco, M. A. C. Pacheco, and F. J. de Souza. 2006. "Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel Neuro-Fuzzy Model for Pattern Classification and Rule Extraction in Databases." *IEEE Transactions Systems on Man and Cybernetics.—Part C: Applications and Reviews* 36 (2): 236-248.
- Hadka, D., and P. Reed. 2012. "Borg: An auto - adaptive many - objective evolutionary computing framework." *Evolutionary Computation* 2012, 1-30.
- Hasperué, W., L. Lanzarini, and A. de Giusti. 2012. "Rule Extraction on Numeric Datasets Using Hyper-rectangles." *Journal of Computing and Information Science* 5 (4): 116-131.

- Hastie, T., R. Tibshirani, R., and J. Friedman. 2001. *The Elements of Statistical Learning*. Springer-Verlag, New York.
- Hsu, C. M., and M. S. Chen. 2009. "On the Design and Applicability of Distance Functions in High-Dimensional Data Space." *IEEE Transactions on Knowledge Data Engineering* 21 (4): 523-536.
- Hwang, S. F., and R. S. He. 2006. "A hybrid real-parameter genetic algorithm for function optimization." *Advanced Engineering Informatics* 20 (1): 7-21.
- Izquierdo, J., I. Montalvo, R. Pérez, and V. S. Fuertes. 2007. "Design optimization of wastewater collection networks by PSO." *Computers and Mathematics with Applications* 56 (3): 777-784.
- Izquierdo, J., R. Minciardi, I. Montalvo, M. Robba, M., and M. Tavera. 2008a. "Particle Swarm Optimization for the biomass supply chain strategic planning." In *Proceedings of 4th Biennial Meeting, iEMSs 2008: International Congress on Environmental Modelling and Software*, 1272-1280, Barcelona, Spain.
- Izquierdo, J., I. Montalvo, R. Pérez, and M. Tavera. 2008b. "Optimization in water systems: a PSO approach." *Business and Industry Symposium (BIS)*, Ottawa, Canada.
- Javadi, A. A., R. Farmani, and T. P. Tan. 2005. "A hybrid intelligent genetic algorithm." *Advanced Engineering Informatics* 19 (4): 255-262.
- Jin, X., J. Zhang, J. L. Gao, and W. Y. Wu. 2008. "Multi-objective optimization of water supply network rehabilitation with non-dominated sorting Genetic Algorithm-II." *Journal of Zhejiang University SCIENCE A* 9 (3): 391-400.
- Johns, M. B., E. Keedwell, and D. Savic. 2014. "Adaptive locally constrained genetic algorithm for least-cost water distribution network design." *Journal of Hydroinformatics*, 16(2): 288-301.
- Jourdan, L., D. Corne, D. Savic, and G. Walters. 2005. "Preliminary investigation of the 'learnable evolution model' for faster/better multiobjective water systems design." In *Evolutionary Multi-Criterion Optimization* (pp. 841-855). Springer Berlin Heidelberg.
- Kamwa, I., S. R. Samantaray, and G. Joos. 2009. "Development of Rule-Based Classifiers for Rapid Stability Assessment of Wide-Area Post-Disturbance Records." *IEEE Transactions on Power Systems* 24 (1): 258-270.
- Kang, D., and K. Lansey. 2011. "Revisiting optimal water-distribution system design: issues and a heuristic hierarchical approach." *Journal of Water Resources Planning and Management ASCE* 138 (3): 208-217.
- Keedwell, E., and S. T. Khu. 2005. "A hybrid genetic algorithm for the design of water distribution networks." *Engineering Applications of Artificial Intelligence* 18 (4): 461-472.
- Keedwell, E., and S. Khu. 2006. "Novel cellular automata approach to optimal water distribution network design." *Journal of Computational and Civil Engineering* 20 (1): 49-56.
- Kehl, V., and K. Ulm. 2005. "Responder Identification in Clinical Trials with Censored Data." *Computational Statistics & Data Analysis* 50 (5): 1338-1355.
- Klosgen, W. 1996. "Explora: A multipattern and multistrategy discovery assistant." In Fayyad UM, Piatetsky-Shapiro, Smyth P, Uthurusamy R, editors. *Advances in Knowledge Discovery and Data Mining*. Cambridge: AAAI Press.
- Koza, J. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.
- Kwakkel, J. H., and J. S. Timmermans. 2012. "Safe Operating Spaces for Human Water Use: Applying Exploratory Modeling and Patient Rule Induction to ANEMI." *2012 International*

Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany.

- Lavrac, N., B. Kavsek, P. A. Flach, and L. Todorovski. 2004. "Subgroup discovery with CN2-SD." *Journal of Machine Learning Research* 5: 153-88.
- Liong, S. Y., and M. Atiquzzama. 2004. "Optimal design of water distribution network using shuffled complex evolution." *Journal of Institution Engineers Singapore* 44 (1): 93-107.
- Liu, X., Y. Minin, Y. Huang, D. B. Seligson, and S. Horvath. 2004. "Statistical Methods for Analyzing Tissue Microarray Data." *Journal of Biopharmaceutical Statistics* 14: 671-685.
- Maier, H. R., Z. Kapelan, J. Kasprzyk, J. Kollat, L. S. Matott, M. C. Cunha, G. C. Dandy, M. S. Gibbs, E. Keedwell, A. Marchi, A., Ostfeld, D. Savic, D. P. Solomatine, J. A. Vrugt, A. C. Zecchin, B. S. Minsker, E. J. Barbour, G. Kuczera, F. Pasha, A. Castelletti, M. Giuliani, and P. M. Reed. 2014. "Evolutionary Algorithms and Other Metaheuristics in Water Resources: Current Status, Research Challenges and Future Directions." *Environmental Modeling and Software*, <http://dx.doi.org/10.1016/j.envsoft.2014.09.013>.
- Marchi, A., G. C. Dandy, A. Wilkins, A., and H. Rohrlach. 2014. "Methodology for comparing evolutionary algorithms for optimization of water distribution systems." *Journal of Water Resources, Planning and Management ASCE* 140 (1): 22-31.
- Martínez-Rodríguez, J. B., I. Montalvo, J. Izquierdo and R. Pérez-García. 2011. "Reliability and Tolerance Comparison in Water Supply Networks." *Water Resources Management* 25: 1437-1448.
- McClymont, K., E. C. Keedwell, D. Savic and M. Randall-Smith. 2013. "A General Multi-objective Hyper-Heuristic for Water Distribution Network Design with Discolouration Risk." *Journal of Hydroinformatics* 15 (3): 700-716.
- McClymont, K., E. C. Keedwell, D. Savic and M. Randall-Smith. 2014. "Automated construction of evolutionary algorithm operators for the bi-objective water distribution network design problem using a genetic programming based hyper-heuristic approach." *Journal of Hydroinformatics*, 16(2): 302-318.
- Mezura-Montes, E., and A. G. Palomeque-Ortiz. 2009. "Self-adaptive and deterministic parameter control in differential evolution for constrained optimization." In Efrén Mezura-Montes (Ed.), *Constraint-Handling in Evolutionary Optimization*, Springer-Verlag, Studies in Computational Intelligence, vol.198, 95-120.
- Michalski, R.S. 2000. "Learnable evolution model: Evolutionary processes guided by machine learning." *Machine Learning*, 38(1-2): 9-40.
- Montalvo, I. 2011. *Diseño óptimo de sistemas de distribución de agua mediante Agent Swarm Optimization*. PhD doctoral dissertation. Universitat Politècnica de València, Valencia, Spain.
- Montalvo, I., J. Izquierdo, M. Herrera, and R. Pérez-García. 2014. "Water supply system computer-aided design by Agent Swarm Optimization." *Computer-Aided Civil and Infrastructure Engineering* 29 (6): 433-448.
- Montalvo, I., J. Izquierdo, S. Schwarze, and R. Pérez-García. 2010. "Multi-objective particle swarm optimization applied to water distribution systems design: An approach with human interaction." *Mathematical and Computer Modelling* 52: 1219-1227.
- Nguyen, V. V., D. Hartmann, and M. König. 2012. "A distributed agent-based approach for simulation-based optimization." *Advanced Engineering Informatics* 26: 814-832.
- Nicklow, J. W., P. M. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld, A. Singh, and E. Zechman. 2010. "State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management." *Journal of Water Resources Planning and Management ASCE* 136 (4): 412-432.

- Onwubolu, G. C., and B. V. Babu. 2004. *New Optimization Techniques in Engineering*, Springer-Verlag.
- Pelikan, M., D. E. Goldberg, and F. G. Lobo. 2002. "A survey of optimization by building and using probabilistic models." *Computational optimization and applications* 21(1): 5–20.
- Persson, J. A., P. Davidsson, S. J. Johansson, and F. Wernstedt. 2005. "Combining agent-based approaches and classical optimization techniques." In *Proceedings of the European workshop on Multi-Agent Systems (EUMAS 2005)*, 260–269.
- Reed, P. M., D. Hadka, J. D. Herman, J. R. Kasprzyk, and J. B. Kollat. 2013. "Evolutionary multiobjective optimization in water resources: The past, present, and future." *Advances in Water Resources* 51: 438-456.
- Shang, W., S. Zhao, and Y. Shen. 2009. "A flexible tolerance genetic algorithm for optimal problems with nonlinear equality constraints." *Advanced Engineering Informatics* 23 (3): 253–264.
- Shean, H., and E. McBean. 2010. "Hydraulic calibration for a small water distribution network." In *Proceedings 12th Water Distribution Systems Analysis Symposium*, Tucson, Arizona: K. Lansey, C. Choi, A. Ostfeld, and I. Pepper, 2010.
- Sycara, K. P., 1998. "Multiagent systems." *American Association for Artificial Intelligence* 19 (2): 79-92.
- Vrugt, J. A., and B. A. Robinson. 2007. "Improved evolutionary optimization from genetically adaptive multimethod search." *Proceedings National Academy of Science USA*. 2007 January 16; 104(3): 708–711.
- Vrugt, J. A., B. A. Robinson, and J.M. Hyman. 2009. "Self - adaptive multimethod search for global optimization in real - parameter spaces." *IEEE Transactions on Evolutionary Computation* 13 (2): 243-259.
- Weichhart, G., M. Affenzeller, A. Reitbauer, and S. Wagner. 2004. "Modelling of an agent based schedule optimisation system." In *Proceedings of the IMS International Forum*, Cernobbio, Italy, 79-87.
- Wooldridge, M., 2009. *An Introduction to Multiagent Systems*, John Wiley & Sons.
- Wrobel S., 1997. "An Algorithm for multi-relational discovery of subgroups." *Proceedings of the 1st European Conference on Principles of Data Mining and Knowledge Discovery*. 1997, Trondheim, Norway. Berlin/Heidelberg: Springer.
- Wu, Z. Y., and M. Behandish. 2012. "Real-time pump scheduling using genetic algorithm and artificial neural network based on graphics processing unit." In *Proceedings Water Distribution System Analysis Conference*, Adelaide, Australia, 2012, 1088-1099.
- Xie, X. F., and J. Liu. 2009a. "Graph coloring by multiagent fusion search." *Combinatorial Optimization* 18 (2): 99–123.
- Xie, X. F., and J. Liu 2009b. "Multiagent optimization system for solving the traveling salesman problem (TSP)." *IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics* 39(2): 489–502.
- Zecchin, A. C., A. R. Simpson, H. R. Maier, A. Marchi, A., and J. B. Nixon. 2012. "Improved understanding of the searching behavior of ant colony optimization algorithms applied to the water distribution design problem." *Water Resources Research* 48 (9): W09505.
- Zheng, F., A. R. Simpson, and A. C. Zecchin. 2013a. "A decomposition and multistage optimization approach applied to the optimization of water distribution systems with multiple supply sources." *Water Resources Research* 49 (1): 380-399.

Zheng, F., A.R. Simpson, and A.C. Zecchin. 2013b. "A coupled binary linear programming-differential evolution algorithm approach for water distribution system optimization." *Journal of Water Resources Planning and Management* (in press).

LIST OF FIGURE CAPTIONS

Figure 1. Putting knowledge extraction and EA to work together

Figure 2. Pareto front obtained without (swarm 1) and with (swarm 2) PRIM

Figure 3. Design of the network with problem-based rules and the influence of PRIM

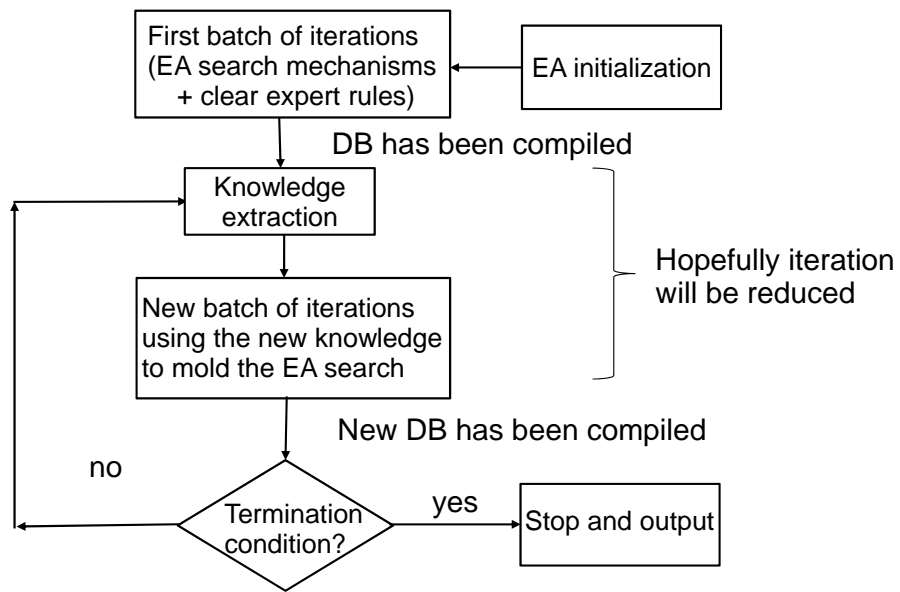


Figure 1. Putting knowledge extraction and EA to work together

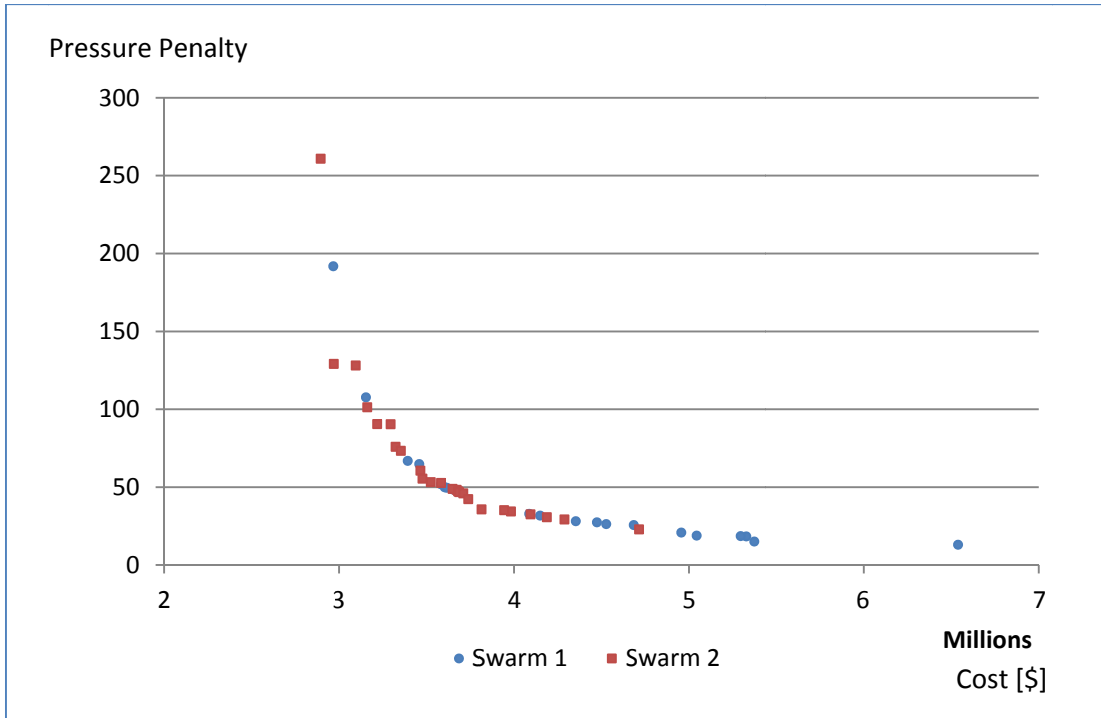


Figure 2. Pareto front obtained without (swarm 1) and with (swarm 2) PRIM

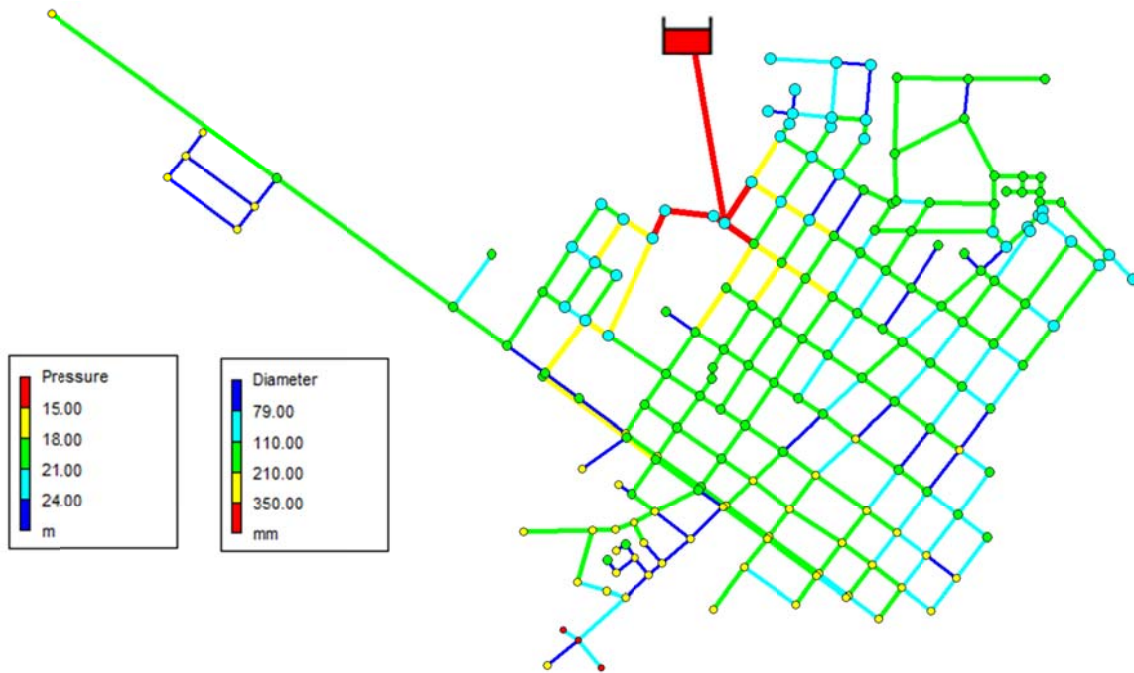


Figure 3. Design of the network with problem-based rules and the influence of PRIM