Drinking Water
Engineering and Science
Discussions

Interactive comment on "Hybridisation of brownboost classifier and glowworm swarm based optimal sensor placement for water leakage detection" *by* Rejeesh Rayaroth and Sivaradje Gopalakrishnan

Anonymous Referee #2

General comments:

The manuscript presents a method to identify optimal pressure measurement positions in a water distribution system for leak detection by utilising classification and meta-heuristic optimisation algorithms. However, the proposed method is questionable and the scientific style how this method is presented in this paper is insufficient. Besides the manuscript being poorly readable, the article lacks necessary technical details crucial for understanding the methodology as well as references to related key literature. Furthermore, some parts in the methods section are missing references at all, resulting

in plagiarism (intended or not), especially in the parts describing brownboost classifiers and glowworm swarm optimisation—the key methods of this paper. The paper contains 8 full pages (pages 6 to 13) copying ideas from other authors without citing them adequately. Even without these concerns, the novelty and usefulness of the proposed methods is arguable. In my opinion, the only novelty lies in the authors applying very specific algorithms (brownboost classifier, glowworm swarm) on a problem that has been already addressed in former literature with various other classification and optimisation algorithms. The results—recommending to put 50 to 500 sensors in a network of 17 km (up to a sensor every 34 (!) meter)—are without doubt unfeasible for applying the methodology on real-world systems. In conclusion, the reviewer could not recommend the manuscript for publication in DWES.

Nevertheless, the authors are invited to consider the following more specific comments:

Specific comments:

Introduction and Related works:

-The introduction is actually missing an introduction: Why is finding leaks important? How do water utilities find leaks (see for example Puust et al. 2010)? What are classical methods, what are recent methods? What are benefits of model-based approaches? After a few incoherent introductory sentences, the introduction lists scientific literature at random. For the reader it is impossible to find any reasoning behind the order of the references listed in the paper.

-Page 1 Line 28: The sentence "leaks in WDN are detected through machine learning techniques" is not true, these methods are just one way to detect leaks, in fact, used only in scientific literature and not in practice at all.

-In general, the listed literature is a mix of different methods for leak detection, which are even focussing on different physical effects caused by a leak (e.g. flow, noise, pressure without even mentioning it here) as well as a wild mix between leak localisation,

leak detection and optimal sensor placement methods. Therefore, they can't be compared and listed in the way they are presented in introduction and the related works section.

-The literature review is insufficient: Only very recent literature is listed—the oldest publication is from 2015. Older key literature as well as novel key literature for this topic is missing. In my opinion, the state of the art is not well described. I recommend the authors to review the literature in the papers that the mention and identify the key papers for this topic.

-The knowledge gap is not well defined. There are only sentences dropped in the introduction like "however, XXX was not improved", "but YYY was not carried out" or "ZZZ was not reduced" without stating if that was even the purpose or aim of the listed paper and without giving any further explanations.

-Page 3 Line 8: Throughout the paper it is mentioned that the method allows to detect leaks with higher accuracy, but it is not clear what is meant. Higher than what?

-Page 3 Line 20: with "neat" diagram sounds strange.

-Page 3 Line 23: Related works: Is the work listed in the introduction not also related work? What is the difference of this section to the previous one?

-Page 3 Line 13: The paper states the difference between normal data and abnormal data, but it is not clear throughout the paper what this terms actually means and how data is identified as abnormal.

-Page 3 Line 34: "Optimisation was not carried out in effective manner" is missing an explanation why it is not effective. The same can be found on Page 4 Line 16

-Page 4 Line 17: What is the difference in this context between optimal sensor placement and sampling design?

Methodology

-In general, the methodology section is not described in an understandable way and not outlined clearly, hence, it does not allow a reproduction by fellow scientists.

-Page 4 Line 25 to 29 would better fit in the introduction section.

-It is not clear if the paper deals with leak detection or leak localisation, since the literature review deals one time with detection and then switches to localisation and vice versa. For example, sentences in the methodology section like on Page 4 Line 29: "in order to detect the leak location" are confusing.

-Page 5 Line 2: What is a water distributer system?

-Page 5 Line 1-5: Why do the authors introduce a graph description of a water distribution system if it is not used later on? Additionally, a reference is missing to previous literature of how to describe a water distribution system as a mathematical graph. For sure, this is not the idea of the authors.

-Page 5 Line 5: Reference to EPANET is missing (Rossman 2000)

-Page 5 Line 8: How are leaks simulated in this paper with EPANET? Why did the authors use extended-period simulations, it seems there is not need for this.

-An important parameter is the leak's size, because this parameter has an effect on the size of the pressure drop and hence the detectability, but the leak size is not mentioned throughout the paper at all. In fact, while reading the paper, it is not clear if there are any leak simulations performed at all. If that is the case, the whole method proposed by the authors is very questionable, because the definition of normal and abnormal pressures does not make sense at all. Can the authors please clarify this point, because it is crucial for the whole publication?

-Figure 1: Besides the bad resolution and that some of the text in the figure is cut away, the figure is not very informative. What does "abnormal pressure data nodes are distributed" even mean?

-The enhanced brownboost classifier method is missing a crucial citation to the original paper by (Freund 2000), who invented this classification method. This is a clear case of plagiarism.This situation is further aggravated by the fact that this is one of the two key methods in this paper.

-It is not clear why the authors use a brownboost classifier at all, since it is invented for noisy environments. The authors are testing their method on simulations, which are not noisy. What is the reason why this classifier was chosen and no other one?

-It is not clear why a k-NN classifier is used before the brownboost classifier. Is it a k-NN classifier or is it just the application of equation 1 on the pressure data?

-It is not clear why the outcome of Equation 1 on Page 7 is binary (0 or 1 as stated on Page 7 Line 19). Looking at the equation, the outcome is supposed to be a floating point number between 0 and 1.

-Page 9 Line 10: How and to what extend does the brownboost classifier improve the classification accuracy and compared to what?

-Similarly to the brownboost classifier method, once again, the glowworm swarm optimisation model is missing a crucial citation to the original work by (Krishnanand and Ghose 2006), the inventors of this algorithm. All the equations in this section can be found in the paper of Krishnanand and Ghose 2006. This is the second clear case of plagiarism which is again aggravated by the fact that this is the second of two key methods in the author's work! Hence, both key methods of this paper are presented in such a way that they were developed by the authors, but in fact, they were not. In total, the paper contains 8 full pages (pages 6 to 13) copying ideas from other authors without citing them adequately.

-Page 11 Equation 11: Parameter gamma is not defined or mentioned in the text. Furthermore, maybe the most important part of a sensor placement algorithm is how to compute the objective function. It is not clear through the whole paper how the

authors actually compute the this function nor what the objective function means in the context of this paper at all.

-Page 12 Equation 13: There is an error in the equation. $L_b(t)$ is in the subscript of the sum.

-Algorithm 2: Since glowworm swarm optimisation is a heuristic method, it cannot be guaranteed that it leads to the optimal solution / optimal node for sensor placement.

Simulation settings:

-In my opinion the machine on which the algorithm has been implemented is not important if the computation time is not discussed in the result section.

-The settings of the constants in the optimisation algorithm (beta, gamma, rho) is not mentioned here, but for optimisation this is of high interest.

-The paper is missing a figure showing the DMA in Barcelona crucial for a further understanding of the results of this paper. Furthermore, it is not clear how the authors get the hydraulic network. Did they get it from researchers in Barcelona? Then it might be also necessary to cite the publication where this network has been introduced for the first time.

-Figure 4: The resolution of the figure is very bad. This has to be improved. Additionally, the figure shows a standard EPANET network (Net 3). Looking at this figure and the fact that the Barcelona network is missing, it is not clear to the reviewer if the authors actually used the Barcelona network for the simulations in this paper, since important materials (Barcelona network model) is not shown.

-Page 14 Line 5-11: It is not clear why the authors have chosen the abbreviations, for example, RM for "node with demand"?

Simulation results:

-In my opinion the convergence and convergence speed of the optimisation method is

of interest, but not mentioned here.

-It is not clear why the authors have chosen the two methods (SVM and Graph-partitioning) for comparison of their method? It seems that these methods are chosen at random from literature? Why haven't the authors chosen other methods that might perform better?

-It is not clear how the authors decide between normal and abnormal pressure data? What does it even mean in this context? Pressure in WDS is also dependent, where in the system it is measured (elevation, roughness values of pieps, valve settings, . . .) so just classifying points according to their pressure won't result in finding leaks automatically. Did the authors generate data by simulating leaks? How many leaks where simulated? What was the leak size?

-In general, it is not clear throughout the paper how the results are generated. The paper shows only sample calculations without detailed explanation. After the sample calculations, tables are listed with numbers and it is not clear, where this numbers come from.

-Using classification time as a measure for the performance of the algorithms is in context of sensor placement very questionable. Furthermore, the reviewer does not see the benefit of a classification time being 36 ms in contrast to 72 ms, since both are very fast. The interesting question would be the convergence time of the optimal sensor placement method, which is not listed in this paper.

-Does the number of pressure data in Figures 5 to 8 correspond to the number of sensors? If that is the case the method would be useless, because deploying 50-500 pressure sensors in a water distribution system of total pipe length of 17 km results in a pressure sensor every 340 to 34 meter. This is a highly unrealistic number of sensors for such a small distribution system. Certainly, no water utility would be able to afford that number of sensors.

-For optimal sensor placement algorithms the most interesting outcome is the location where sensors should be placed. The optimal sensor positions are not shown in this paper.

-A final comment about the use of abbreviations: The authors define abbreviations like EBBC-GWO multiple times in the paper without using it. Basically, in each section the abbreviations are defined again, which is certainly not the purpose of abbreviations at all.

-Finally, there are a lot of repetitions of paragraphs, hence, the paper is not concise.

Freund, Y. (2001). An adaptive version of the boost by majority algorithm. Machine Learning, 43(3), 293–318. http://doi.org/10.1023/A:1010852229904

Krishnanand, K. N., & Ghose, D. (2006). Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications. Multiagent and Grid Systems, 2(3), 209–222. http://doi.org/10.3233/MGS-2006-2301

Rossman, L A. EPANET 2 USERS MANUAL. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-00/057, 2000.

Puust, R., Kapelan, Z., Savic, D. A., & Koppel, T. (2010). A review of methods for leakage management in pipe networks. Urban Water Journal, 7(1), 25–45. http://doi.org/10.1080/15730621003610878