Adaptive Augmented Evolutionary Intelligence for the Design of Water Distribution Networks

Matthew B. Johns
University of Exeter
Exeter, EX4 4QF
United Kingdom
M.B.Johns@exeter.ac.uk

Herman A. Mahmoud
University of Exeter
Exeter, EX4 4QF
United Kingdom
H.Mahmoud@exeter.ac.uk

Edward C. Keedwell
University of Exeter
Exeter, EX4 4QF
United Kingdom
E.C.Keedwell@exeter.ac.uk

Dragan A. Savic
University of Exeter/KWR
Exeter/Nieuwegein
United Kingdom/Netherlands
D.Savic@exeter.ac.uk

ABSTRACT

The application of Evolutionary Algorithms (EAs) to real-world problems comes with inherent challenges, primarily the difficulty in defining the large number of considerations needed when designing complex systems such as Water Distribution Networks (WDN). One solution is to use an Interactive Evolutionary Algorithm (IEA), which integrates a human expert into the optimisation process and helps guide it to solutions more suited to real-world application. The involvement of an expert provides the algorithm with valuable domain knowledge; however, it is an intensive task requiring extensive interaction, leading to user fatigue and reduced effectiveness. To address this, the authors have developed methods for capturing human expertise from user interactions utilising machine learning to produce Human-Derived Heuristics (HDH) which are integrated into an EA's mutation operator. This work focuses on the development of an adaptive method for applying multiple HDHs throughout an EA's search. The new adaptive approach is shown to outperform both singular HDH approaches and traditional EAs on a range of large scale WDN design problems. This work paves the way for the development of a new type of IEA that has the capability of learning from human experts whilst minimising user fatigue.

CCS CONCEPTS

• **Theory of computation → Design and analysis of algorithm**

GECCO'20, July 8-12, 2020, Cancun, Mexico

DOI: 10.1145/nnnnnnn.nnnnnnn

KEYWORDS

Evolutionary Algorithm; Machine Learning; Human-computer Interaction; Knowledge Guided Search; Water Distribution Network Design; Real-world Application

1 INTRODUCTION

Evolutionary algorithms (EAs) have been used for the optimisation of theoretical and real-world inspired problems in many fields of study. One such area is water systems engineering, more specifically the design and operation of Water Distribution Networks (WDNs). Although solutions produced by EAs can provide mathematically optimal designs, they are often not entirely suitable for real-world application. This is due in part to the multitude of considerations an engineer needs to take into account when designing such complex systems which are difficult to express mathematically in objective functions and constraints. As a result, engineers utilising evolutionary optimisation techniques will often have to manually adjust promising solutions to meet real-world application requirements even following extensive objective function formulation and refinement. This need to involve an expert in the real-world optimisation of complex systems is a recognised issue [1] and thus researchers in the field of water systems engineering have been looking for approaches to address this.

The work presented in this paper proposes a refined approach for the capture and application of engineering expertise to EAs for the optimal design of WDNs. This is achieved through the use of advanced interactive visualisation, machine learning and algorithm mechanics to produce EAs capable of learning from water systems engineers to improve algorithm performance and real-world solution applicability. It is envisaged that the developments presented here will provide a robust foundation for the development of a novel type of Interactive Evolutionary Algorithm

(IEA) which will have the ability to learn and apply complex engineering concepts to real-world WDN design problems.

1.1 Knowledge Guided Search

Evolutionary algorithms have been applied to a substantial range of optimisation tasks and have proven to be versatile approaches capable of locating high quality solutions for extremely complex problems. Their ability to effectively navigate large search spaces whilst avoiding local optima means they can be viewed as truly global search techniques. This ability can be partly attributed to their independence over the problem being optimised, however in the case of large real-world problems where evaluations are expensive and performance is essential, EAs can struggle to obtain quality results in a timely manner.

When applied to the problem of WDN design, an EA is dependent on variation operators such as crossover and mutation to alter the design of the network. These operators are however ignorant of the effect changes made to the chromosome have on the resultant solution. For example, from the EA's standpoint, a change to the diameter of a pipe has no impact on hydraulic performance of connected elements until the solution is evaluated, however an engineer would know the hydraulic head at adjacent nodes and velocity of water in the pipe would be affected. The hydraulic performance of the candidate solution is only known following decoding and evaluation and although this abstraction allows EAs to be applied to a wide variety of problems without excessive alteration, there is clear scope for the integration of problem specific knowledge provided it is implemented in a manner not to the detriment of global search and performance capability.

The integration of problem specific knowledge into EAs within the field of water systems has started to gain traction over the past decade. A popular approach has been to seed the initial population of an EA with solutions generated by hydraulically based engineering inspired heuristics [2–5]. More recently these engineering inspired heuristics have been integrated into the variation operators of EAs with the view of not only improving overall algorithm performance but also solution feasibility [6, 7]. However, these approaches require the need to develop these heuristics manually, usually from hydraulic principles and the observation of engineering practices.

1.2 Interactive Evolution

Interactive Evolution [8] aims to combine human knowledge and judgement through interaction with an EA during the search process. The most common interaction is often to assess solution fitness; however, the user can also interact with other algorithm operators such as variation and selection. When applying an EA to a problem, especially those of a real-world nature, it is often non-explicit conditions which are hard to characterize. In the field of product design for example, a designer relies on human intuition to make subjective decisions as to the aesthetic qualities of a design such as a piece of furniture [9]. In cases such as these, the fitness criterion cannot be explicitly formulated, often requiring a side-by-side comparison to evaluate a solution effectively. In addition, user interaction can be used to guide the search of the algorithm to increase convergence and help avoid the trappings of local optima.

The complexity of WDN problems mean they are difficult to solve, not only from a mathematical standpoint but also from sociological, political and other subjective perspectives. In the field of water systems, most of the research is focused on simulation model development and application refinement of optimisation algorithms. The problem comes when an EA returns a mathematically optimal solution, but the solution becomes infeasible when considering the subjective preferences of the engineer [10]. Researchers have recently developed methods using interactive evolution to calibrate models by allowing the user to incorporate unmodeled objectives into the search process [11]. Interactive evolution is becoming an increasingly popular area of research, especially where problems require the subjective responses of human users to guide the search of EAs [8]. Singh et al [11] utilised human input and the Non-dominated Sorting Genetic Algorithm II (NSGAII) [12] to identify solutions for groundwater problems which were both mathematically optimal and feasible. This was achieved through the use of human responses which were combined with other criteria to compute solution fitness. Although the interactive aspect of the technique was simple (solution ranking) the interactive EA produced high-quality solutions, outperforming the standard version of NSGAII.

These developments in the field of interactive evolution, specifically in water systems engineering, suggest interactive evolutionary algorithms have the capability to produce solutions to complex real-world problems.

1.3 Machine Learning

The aim of this work is to capture and embed human expertise into an EA. This has been achieved in previous research through the expression of 'rules of thumb' which are integrated into an EA through heuristics. However, this is often a difficult task as most decisions made by an engineer are founded on intuition and 'feel' rather than a specific set of rules. Therefore, in this research we propose the utilisation machine learning techniques as a tool to learn behaviour from engineering interactions. Machine learning algorithms use computational techniques to learn directly from data to produce models without relying on explicit instruction sets or pre-defined equations. These methods, in particular decision trees [13] and neural networks [14] have been used to address problems in WDN including pipe deterioration modelling, model calibration, leakage detection and demand forecasting. Random forests, utilised in this work, are ensemble learning approaches for classification and regression problems and are constructed from multiple decision trees. Whilst decision trees are simple, computationally efficient and interpretable they are prone to overfitting.

1.4 Multi-objective Water Distribution Network Design

The optimal design of WDNs has been initially formulated as a single-objective problem through minimizing cost (denoted by NC in Eq. 1), with pipe diameters given as decision variables whilst the layout of pipes, connectivity and nodal demands are considered as input data. However, a number of authors claimed that the optimal design of a WDN is a multi-objective problem since it involves trade-off between conflicting objectives such as cost and network reliability. The reliability is the ability of providing adequate supply to consumers under both standard and unexpected operation conditions. Todini [15] resilience index (RI) is widely used as a surrogate measure of reliability and its defined as the capability of the network to provide more head than required at each demand node in order to have a sufficient excess to be dissipated internally in case of failure present such as pipe burst. Prasad and Park [16] combined both resilience index and diameter uniformity coefficient (Eq. 2) to provide a better representation of the reliability of loops in the network. Thus, the design optimization problem in this work is expressed as dual-objective function through minimizing network cost and maximizing network resilience index and can be written as follows:

$$Minimize \qquad NC = \sum_{i=1}^{p} u_i(d_i) \cdot l_i \tag{1}$$

$$Maximize \qquad RI = \frac{\sum_{j=1}^{n} c_j \cdot q_j \left(h_j^{avl} - h_j^{req}\right)}{\left(\sum_{k=1}^{r} q_k \cdot h_k + \sum_{n=1}^{pp} \frac{P_n}{\gamma}\right) - \sum_{j=1}^{n} q_j \cdot h_j^{avl}} \tag{2}$$

$$c_j = \frac{\sum_{d=1}^{pn} d_d}{pn \cdot max\{d_d\}} \tag{3}$$

Where: $u_i(d_i)$ = unit cost of pipe of a given diameter; l_i = length of pipe *i; p* and *n* = number of pipes and demand nodes in a given network; q_j, h_j^{avl} and h_j^{req} = available demand, available pressure head and required pressure head at node *j*; q_k and h_k = supply and elevation head at reservoir *r*; *r* and *pp* = number of reservoirs and pumps in a given network;P_p = power of pump *n;* γ = specific weight of water;c_j = uniformity at the node *j*; *pn* = number of pipes connected to node j; and d_d = diameter of pipe *d* connected to node *j*.

A generated solution to the above problem is represented by a vector of integer numbers, in which each element is the value of a pipe diameter in that solution. The value of the vector ranges from one up to the number of commercially available diameters. The quality of the generated solution is assessed based on the objective functions.

The optimization problem is subject to the hydraulic constraints. These involve satisfying continuity at each demand node, conserving energy in loops and ensuring that available pressure head at each node is always equal to or above the required pressure head. The constraints require solving conservation of mass and energy equations to determine the nodal pressure heads, flows in pipes for a given network, and are automatically satisfied by using the well-known EPANET2.0 [17] hydraulic simulator.

2 EXPERIMENTAL SETUP

The experimentation presented in this paper comprises of 3 core components: interaction capture, HDH machine learning and integration of HDHs into EAs. As stated previously, the aim of this paper is to develop refined techniques for the automatic capture and use of engineering knowledge within EAs with the view to developing a new type of IEA for the optimisation of WDN design problems.

The approach presented here involves several WDN design problems of varying size and complexity. Three small scale problems are used to train the HDH models and two large, real-world design problems are used to evaluate the performance of the algorithms on test.

2.1 Engineering Interaction Capture

To facilitate the capture of engineering knowledge a framework was developed to allow engineers to interact with WDN optimisation problems. The HOWS framework [18, 19] consists of two core components, a server and an interactive visualisation client (IVis). The system is designed so that computationally expensive operations such as network configuration, automatic optimisation, hydraulic simulation and objective function evaluation are handled by the server. The information generated by the server is sent to IVis which is designed to visualize the data and facilitate easy interaction with the problem at hand. IVis utilises advanced three-dimensional rendering techniques to present the user with an intuitive representation of a WDN. Several visualization techniques are employed to provide topological, hydraulic and optimisation information to assist the user to make efficient decisions. Figure 1 shows the Blacksburg network [20] being interacted with in IVis.

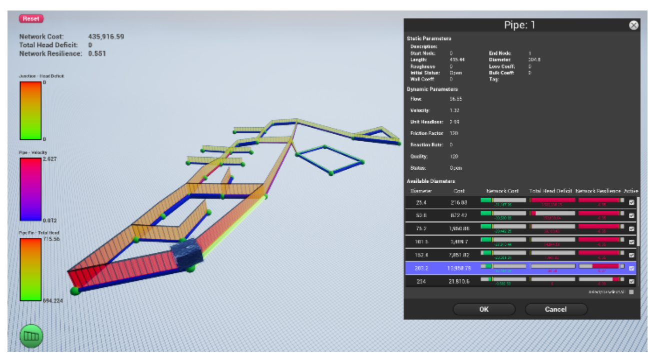

Figure 1: HOWS Framework Interactive Visualization (IVis) Client

As stated previously, WDNs are complex systems consisting of many interconnected components, the most prevalent of which are pipes and junctions. Within IVis these components are represented using simple 3D geometry, cylinders for pipes and spheres for junctions. The primary decision variable in most WDN design problems is the diameter of pipes, this is conveyed to the user by scaling the diameter of the representative cylinder so that it is directly proportional to the diameter of the pipe. This enables the engineer to efficiently identify diameters without the need to inspect individual pipes. The topology of a network is defined by the position of the junctions (spheres) in the 3D domain which intuitively conveys metrics such as distance and elevation change between pipes. IVis uses colour to communicate a variety of hydraulic and optimisation information to the user. By default, the hydraulic head violation at each junction is displayed to the engineer using a linear colour scale where red indicates violation (not enough hydraulic head) and green indicates sufficient head, this allows the user to quickly identify problem areas within the network.

Additional hydraulic data is displayed to the engineer through 'pipe fins', these are aligned along the length of each pipe and can be used to convey up to three separate parameters simultaneously. By default, hydraulic head of connecting junctions is represented by fin height, with this value being reinforced using a linear colour gradient where red is high, and green is low. In addition, each fin has evenly spaced vertical lines running its length, these are moved in the direction of flow at a speed relative to the velocity of water in that pipe. The inclusion of these 'high bandwidth' visualisation techniques provide the engineer with all the relevant information needed to make most decisions without the need to constantly change visualisation settings. It is hoped this will aid the engineer make informed decisions in a timelier manner, increasing interaction capture volume.

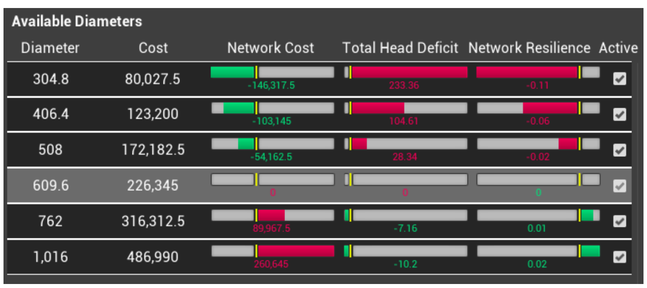

Figure 2: HOWS IVis Look-ahead Dialog

Due to the multi-objective nature of the problem it is important for the engineer to consider all objectives and constraints during the manual optimisation process and not favour any single objective. This was achieved by defining an 'equilibrium' point from the reference Pareto front for each training problem. The equilibrium point is defined as the point on the Pareto front that has the smallest Euclidean distance from the ideal point (minimum feasible cost and maximum resilience). From this the user is given a target value for each primary objective which is within 10% of the equilibrium point.

For this set of experiments the engineer is presented with a randomly generated solution to the problem and tasked with manually optimising the network so that both objective values meet or surpass the target values provided whilst ensuring solution feasibility (no constraint violations). These values are presented to the engineer in the top left of the screen and updated automatically when changes to the network are made. The engineer interacts with the network by clicking on its components. Clicking on a junction will provide the engineer with relevant hydraulic head information highlighting any constraint violation. Selecting a pipe will bring up a dialog detailing the pipe's hydraulic status and if a decision variable, will present the engineer with a selection of available diameters. A core feature of IVis is the 'look-ahead' system which is designed to give the engineer detailed information on what would happen if the diameter of a selected pipe is changed. Figure 2 shows a detailed view of the look-ahead dialog in IVis. Upon selecting a pipe in the network, IVis sends a request to the server for the available diameters for that pipe. Upon receipt of the request, the server calculates the objective and constraint values, in this case, network cost, total head deficit, and network resilience for each potential diameter change. This feature aids the user make more educated and precise decisions with a view of increasing interaction quality.

To gather interaction data used to generate the HDHs, three engineers were asked to manually solve three small WDN design problems, namely Two Loop [21], Hanoi [22], and Blacksburg [20]. Two-loop consists of a single water source and 8 decision pipes with 14 available pipe diameters to choose from. Hanoi is a representation of a single water source network consisting of three loops, based upon the trunk main layout for the city of Hanoi, Vietnam. The network consists of 34 decision pipes and 6 available pipe diameters. The final network, Blacksburg, is a representation of a single source network consisting of multiple loops and branches. It consists of 35 pipes 23 of which are decision variables and 14 available pipe diameters.

Each time an engineer changes the diameter of a pipe the server logs the state of the network, these interaction logs can then be used to train machine learning models to predict what an engineer would do given a selected pipe and network state.

2.2 Learning Human Derived Heuristics

The developments presented here build upon earlier work [23, 24] that used machine learning techniques to develop models that would predict, given a randomly selected pipe, what diameter an engineer choose given the hydraulic state of the network. This early work paved the way to producing a generalisable approach where models can be trained on interaction data obtained on small WDN problems and then applied to large scale problems. This is important as it is almost impossible for a user to make a large WDN optimal through manual intervention, a process that will likely lead to user fatigue [25].

The previous work utilised a decision tree-based learning approach which was selected due to the ability to visualise and interpret the models generated, however this approach is prone to overfitting and can impact model accuracy. To address this, the work presented here utilises random forest [26] machine learning methods to generate the Human Derived Heuristics as they are less prone to overfitting whilst reducing error due to variance and bias.

The random forest method requires a fixed input schema. In this approach the following seven normalized features, local to the selected pipe, are considered: the current diameter, velocity, upstream head deficit, downstream head deficit, pipe influence, flow and length. These parameters were selected as they are often important considerations for an engineer when selecting a new diameter.

2.3 Integrating Heuristics into EAs

The trained regressor and classifier models are integrated into an EA through the mutation operator. The HDH based mutation operator has been developed to replace the standard mutation procedure of an EA. Three HDH based mutation operators are presented in this work: HDH – Regressor (HDH-R), HDH – Classifier (HDH-C) and HDH – Adaptive (HDH-A). All HDH based operators function in a similar manner, firstly by decoding the chromosome and randomly selecting a pipe (decision variable) in the network. The selected pipe's current diameter, velocity, upstream head deficit, downstream head deficit, pipe influence, flow and length are applied to the trained model. For the HDH-R the model will predict a new diameter for the pipe, this value is compared to the list of available diameters and the closest is applied to the selected pipe. For the HDH-C the classifier model predicts whether a pipe should be increased or decreased, this prediction is used alongside the list of available diameters to either increment or decrement the pipe's diameter.

Previous work [23, 24] suggests that the regressor based HDH approach performs well during the initial stages of a search whilst the classifier based HDH approach is less susceptible to premature convergence, performing well in the later stages of the search.

This led to the development of an adaptive approach which aims to combine both HDH mutation methods (HDH-R & HDH-C), applying HDH-R at the start of the search and HDH-C towards the end. This is achieved using a Hypervolume Gradient Monitor (HGM), a process which directly controls the application probability of both HDH methods, this is calculated as follows:

$$\mathbb{P}(m_R) = \frac{g_c}{g_i} \tag{4}$$

Where g_i is the initial gradient of the hypervolume curve, g_c is the current gradient of the hypervolume curve and $P(m_R)$ is the probability of utilising the HDH-R guided mutation. The gradient of the hypervolume curve is calculated at the end of each generation, using a comparison between the current hypervolume and that of five generations previous. If HDH-R is not utilised, then HDH-C is used instead. This process is designed to apply the regressor based HDH aggressively at the beginning of the search and smoothly transition to the classifier HDH as the search progresses and the rate of convergence slows.

An important factor to consider when implementing such an operator is computational efficiency. In the case of this WDN design problem, the most computationally intensive task is the hydraulic simulation, conducted during solution evaluation. Therefore, to preserve computational efficiency, it is crucial not to incur any additional hydraulic evaluations as this would drastically increase runtime, especially for large real-world networks. Due to the dependency the HDH models have on a solution's hydraulic information (pressure, flow and velocity), mutation cannot follow crossover, without the need to evaluate the resultant solution. Therefore, in the HDH variant EAs the mutation operator precedes the crossover operator in order to preserve the hydraulic information from the parent solution.

It was found in previous experimentation [23] that when completely replacing an EA's standard mutation with HDH mutation, although early performance was boosted, resulted in premature convergence. Thus, a straightforward method for combining standard and HDH mutation was devised where each time the HDH mutation operator was invoked, there was a probability of the standard mutation being used instead. Through experimentation it was found that a probability of P(m)=0.5 resulted in the best all round performance, therefore in the following experimentations, the application probability of all HDH based approaches will be set at P(m)=0.5. The EAs used to assess the performance of the new HDH based approaches are NSGAII and the Strength Pareto EA 2 (SPEA2) algorithm [27], both were selected due to their ability to perform well on this WDN design problem [28, 29].

Two large networks from the literature [30, 31], previously unseen by machine learning model, were used for testing the performance of the universal HDHs implemented in the EA in the proposed method. The universal HDHs are created from the cumulative interaction from all users using the generalization method presented by [24]. The first test network is Modena network which includes 317 pipes (all decision variables), 268 demand nodes, 4 reservoirs with fixed head within 72.0 m to 74.5 m and 13 available pipe diameters. The total number of combinations to cover the full solution search space is 13^{317}. The second network is Balerma Irrigation Network (BIN). The BIN consists of 454 relatively short-length pipes, 443 demand nodes, and the water supplied to the network through four reservoirs. There are ten available diameters, ranging from 113.0 mm to 581.8 mm, for each decision pipe, resulting in a search space equal to 10^{454} combinations. The required pressure head for all demand nodes in both test networks is maintained at 20m.

3 RESULTS AND DISCUSSION

For each problem presented in this section the parameters of the base algorithm (SPEA2 and NSGAII) remain constant. A population size of 100, single-point crossover and a mutation probability of 1/n where n is the number of decision variables are used. Regarding the HDH variants, the probability that the HDH method is applied during mutation is fixed at P=0.5. Each algorithm is run 30 times for a total of 500,000 fitness evaluations. For each test problem the hypervolume [32] and Inverted Generational Distance (IGD) [33] performance metrics are used to evaluate the algorithms on test. Hypervolume is calculated for each problem using the theoretical best (utopia) and worst (nadir) points in the solution space. The IGD is calculated using a reference Pareto front obtained from Wang et al. [29].

The first set of results presented are from the Modena problem. Figure 3 shows the average hypervolume for each algorithm over the search. The first observation is that all HDH based algorithms outperform their respective 'standard' algorithm (NSGAII and SPEA2) over the entirety of the allotted 500,000 fitness evaluations. The highest performing HDH variant during the initial stages of the search is HDH-R, boosting convergence rate for both standard algorithms. The next best performing algorithm is the adaptive HDH variant (HDH-A) followed by the classifier-based algorithm (HDH-C). This behaviour is somewhat expected as the regressor based heuristic is able to make larger changes in terms of pipe diameter compared with the classifier approach which is restricted to making incremental diameter changes.

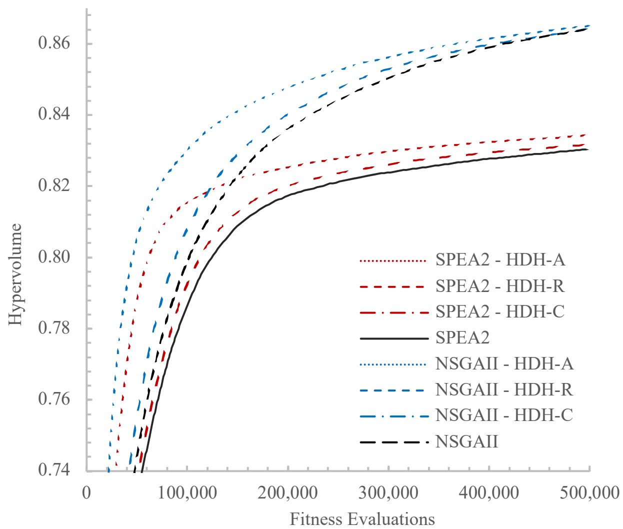

Figure 3: Mean Hypervolume for the Modena Problem – SPEA2 – HDH and NSGAII – HDH Variants

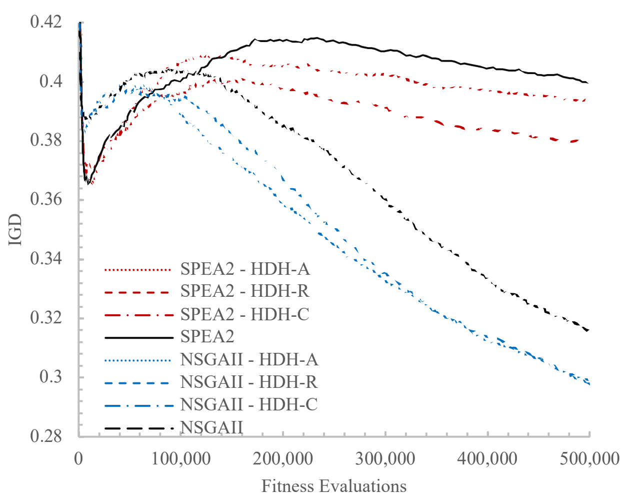

Figure 4: Mean Inverted Generational Distance for the Modena Problem – SPEA2-HDH and NSGAII-HDH Variants

Figure 4 shows the average IGD for the algorithms on test for the Modena problem. Interestingly the algorithms exhibit different order of performance regarding IGD than the Hypervolume results and vary between base algorithms. Regarding the SPEA2 based algorithms, SPEA2 - HDH-C achieves a better IGD throughout nearly the entirety of the search followed by SPEA2 – HDH-A and finally HDH-C. However, NSGAII – HDH-A achieves the lowest IGD out of all the NSGAII results followed by the other two HDH approaches.

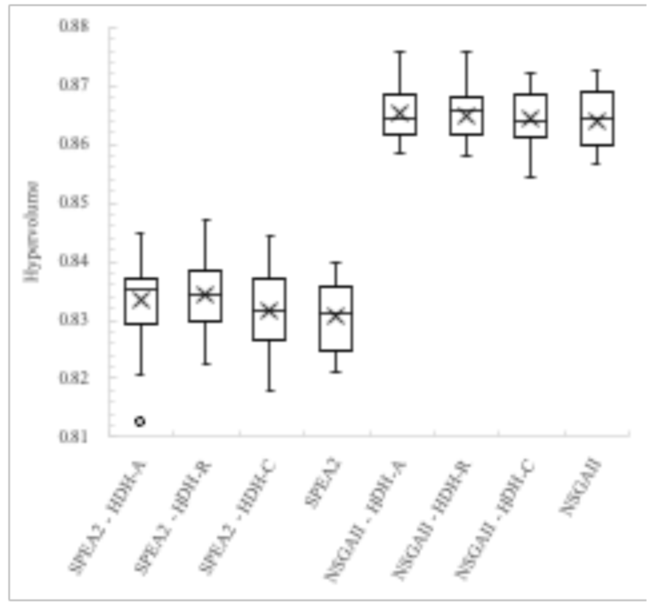

Figure 5: Hypervolume Results for the Modena Problem - SPEA2-HDH and NSGAII-HDH Variants

Figure 5 shows the final hypervolume results (30 runs) for each of the algorithms on test following the allotted 500,000 fitness evaluations. All HDH achieve a higher average hypervolume than their standard counterparts. However, utilising statistical testing (Mann-Whitney U [34]) no statistically significant difference between the NSGAII - HDH variants and NSGA – II. SPEA2 – HDH-R and SPEA2 - HDH-A where found to produce statistically different distributions from SPEA2.

The IGD results for the final generation for the Modena problem are shown in Figure 6. With regard to the SPEA2 based algorithms, both HDH-A and HDH-R achieve statistically better results than the other two algorithms. NSGAII – HDH-A is the best performing algorithm in terms of IGD, outperforming the other HDH variants and standard algorithm.

The average hypervolume over the allotted fitness evaluation for the BIN problem are displayed in Figure 7. For this more complex problem the HDH based approaches have an increased impact on the performance of the standard algorithms. The initial performance of both HDH-R algorithms initial is promising; however, the adaptive methods SPEA2 – HDH-A and NSGAII-HDH-A overtake at 150,000 and 220,000 evaluations respectively. The classifier based HDH algorithms also go on to outperform HDH-R in both cases, however, does not catch the adaptive algorithms.

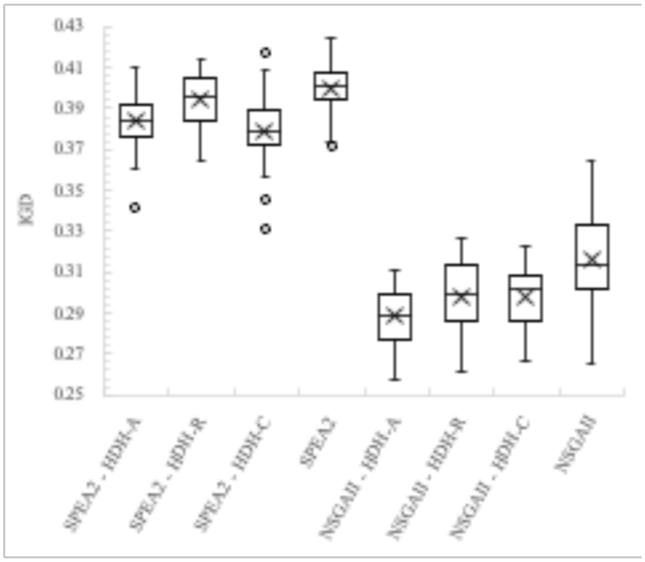

Figure 6: Inverted Generational Distance Results for the Modena Problem - SPEA2-HDH and NSGAII-HDH Variants

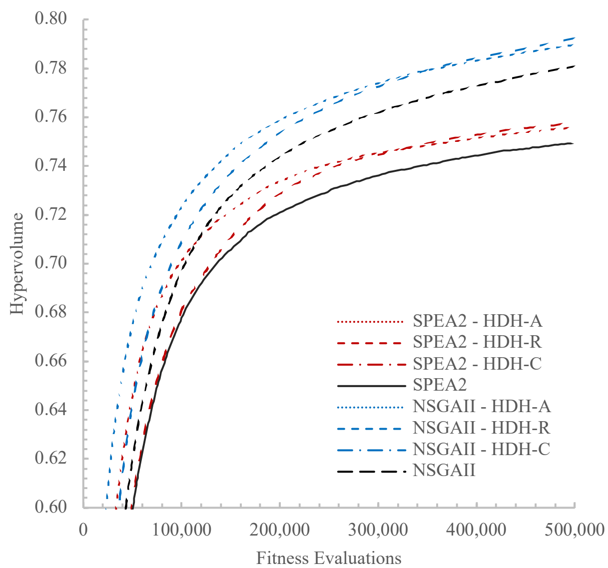

Figure 7: Mean Hypervolume for the Balerma Irrigation Network Problem – SPEA2 – HDH and NSGAII – HDH Variants

Figure 8 displays the average IGD results for the BIN problem. Initially the HDH algorithms reduce their population's IGD at a faster rate than the standard algorithms. Interestingly SPEA2 achieves a better result than its HDH counterparts from approximately 100,000 fitness evaluations onwards. In the case of NSGAII there is less distinction between the algorithms with NSGAII – HDH-R performing the worst during the majority of the search.

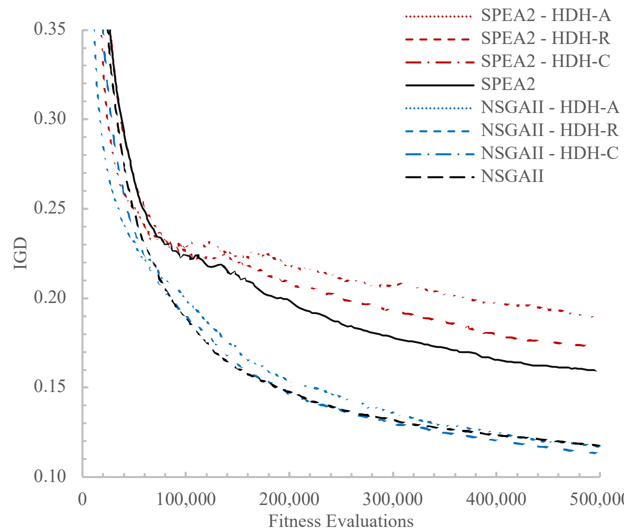

Figure 8: Mean Inverted Generational Distance for the Balerma Irrigation Network Problem – SPEA2-HDH and NSGAII-HDH Variants

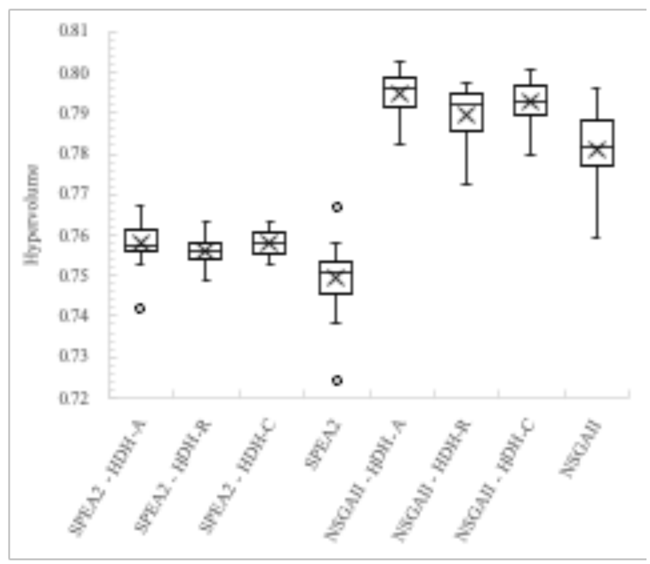

Figure 9: Hypervolume Results for the Balerma Irrigation Network Problem - SPEA2-HDH and NSGAII-HDH Variants

The final generation hypervolume results for the BIN problem are reported in figure 9. The HDH based algorithms exhibit the same performance characteristics for both base algorithms with the adaptive method achieving the greatest hypervolume, followed by HDH-C and finally HDH-R.

Figure 10 presents the IGD results for the BIN problem. Interestingly for the SPEA2 based algorithms the HDH variants perform significantly worse than the standard algorithm, with the classifier based approach performing the best out of the other HDH methods. This is not the case for the NSGAII algorithms, where both the adaptive and classifier approaches find significantly lower IGD values compared with the HDH-R and the standard algorithm.

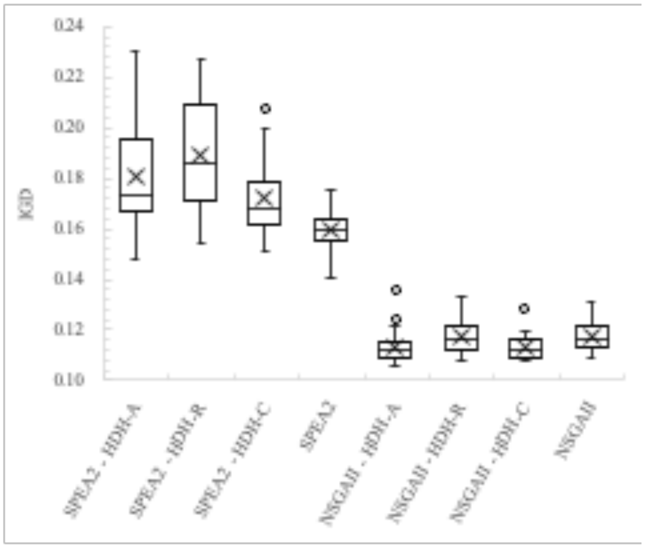

Figure 10: Inverse Generational Distance Results for the Balerma Irrigation Network Problem - SPEA2-HDH and NSGAII-HDH Variants

Figure 11 presents the resultant non-dominated results from all runs for both SPEA2 and SPEA2 – HDH-A for the BIN problem. In addition, the reference Pareto front is displayed along with the equilibrium point, which was used as a target for the engineering interaction capture procedure. This figure goes some way to explain the Hypervolume and IGD results of both algorithms. It is clear why SPEA achieves a lower IGD as it is able to find lower cost solutions in the lower resilience range, whereas HDH-A is drawn towards the equilibrium point due to the influence of the HDHs. The adaptive HDH influences the search of the algorithm and 'pulls' the Pareto front towards the equilibrium point. This is a clear example that the preferences of a human expert have indeed been imparted to the algorithm. This suggests that any subjective preferences when designing a WDN could be captured and conveyed to the algorithm's search, generating solutions more in line with the desires of the engineer.

From these results it can be observed that the inclusion of the HDH methods, especially the adaptive technique, exhibit greater performance benefits on the larger of the two test problems (BIN). This could be due in-part to the networks used to train the HDH models. The nature of these training models often requires the engineer to find the feasible (in terms of hydraulic head) solution space prior to fully addressing the objective functions. However, the Modena problem is easily made feasible, due to its looped nature and expansive selection of available diameters. In fact, all algorithms found at least one feasible solution at initialisation. On the other hand, it is a lot harder for an EA to locate the feasible search space for the BIN problem, due to less surplus hydraulic head and pipe sizing options. The HDH models therefore aid the algorithm locate the feasible region more quickly, thus boosting overall algorithm performance.

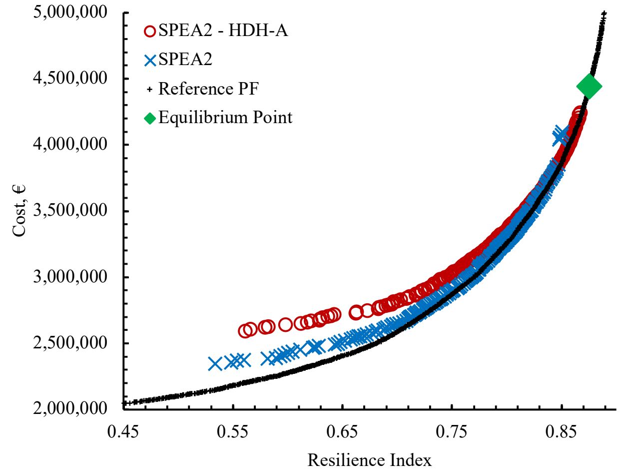

Figure 11: Non-dominated Fronts for the Balerma Irrigation Network Problem – SPEA2 & SPEA2 – HDH-A

4 CONCLUSIONS

This paper presents several HDH based algorithms for the optimisation of multi-objective WDN design problems. The derived heuristics are generated automatically from the interactions of an expert engineer solving small WDN design problems, through the use of random forest machine learning algorithms. An improved interactive visualisation system for the efficient capture of human interactions was also presented. The system is designed to maximise productivity and reduce user fatigue with the view of producing higher quality interactions from engineers. The derived heuristics are integrated into EAs through their mutation operators using a variety of techniques, including a method for combining HDH models using an adaptive process which controls HDH model application strength by monitoring hypervolume convergence. The results show that the performance on an EA can be enhanced through the integration of captured domain specific knowledge, not only in terms of performance metrics but also solution feasibility. Although the use of singular HDH models within the mutation operator of an EA improves performance, adaptively combining the HDH approaches yields improved and more consistent results on large-scale multi-objective WDN design problems.

This work paves the way for the development of a new type of IEA that has the capability of automatically learning from human experts whilst minimising user fatigue. This algorithm would open up the potential for an effective interface between human expert and evolutionary algorithm resulting in improved, more engineering feasible solutions to real-world problems. There is also scope with further development to apply these methods to different domains and a wider set of problems beyond WDN design.

ACKNOWLEDGMENTS

The work presented in this paper was developed as part of the Human-Computer Optimisation for Water Systems Planning and Management (HOWS) project funded by the Engineering and Physical Sciences Research Council (EPSRC) - grant EP/P009441.

REFERENCES

[1] Marchi, A., Salomons, E., Ostfeld, A., Kapelan, Z., Simpson, A.R., Zecchin, A.C., Maier, H.R., Wu, Z.Y., Elsayed, S.M. and Song, Y. 2013. Battle of the water networks II. *Journal of water resources planning and management*. 140, 7 (2013), 04014009.

[2] Bi, W., Dandy, G.C. and Maier, H.R. 2015. Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. *Environmental Modelling & Software*. 69, (Jul. 2015), 370–381. DOI:https://doi.org/10.1016/j.envsoft.2014.09.010.

[3] Kang, D. and Lansey, K. 2011. Revisiting optimal water-distribution system design: issues and a heuristic hierarchical approach. *Journal of Water resources planning and management*. 138, 3 (2011), 208–217.

[4] Keedwell, E. and Khu, S.-T. 2006. A novel evolutionary meta-heuristic for the multi-objective optimization of real-world water distribution networks. *Engineering Optimization*. 38, 03 (2006), 319–333.

[5] Zheng, F., Simpson, A.R. and Zecchin, A.C. 2011. A combined NLP-differential evolution algorithm approach for the optimization of looped water distribution systems. *Water Resources Research*. 47, 8 (2011).

[6] Johns, M.B., Keedwell, E. and Savic, D. 2014. Adaptive locally constrained genetic algorithm for least-cost water distribution network design. *Journal of Hydroinformatics*. 16, 2 (2014), 288–301.

[7] McClymont, K., Keedwell, E. and Savic, D. 2015. An analysis of the interface between evolutionary algorithm operators and problem features for water resources problems. A case study in water distribution network design. *Environmental Modelling & Software*. 69, (Jul. 2015), 414–424. DOI:https://doi.org/10.1016/j.envsoft.2014.12.023.

[8] Takagi, H. 2001. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*. 89, 9 (2001), 1275–1296.

[9] Brintrup, A.M., Ramsden, J.J. and Takagi, H. 2008. Ergonomic chair design by fusing qualitative and quantitative criteria using interactive genetic algorithms. (2008).

[10] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A. and Singh, A. 2009. State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*. 136, 4 (2009), 412–432.

[11] Singh, A., Minsker, B.S. and Valocchi, A.J. 2008. An interactive multi-objective optimization framework for groundwater inverse modeling. *Advances in water resources*. 31, 10 (2008), 1269–1283.

[12] Deb, K., Agrawal, S., Pratap, A. and Meyarivan, T. 2000. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *International Conference on Parallel Problem Solving From Nature* (2000), 849–858.

[13] Breiman, L., Friedman, H, J., Olshen, A, R. and Stone, J, C. 1984. *Classification and Regression Trees*. Routledge.

[14] Beale, H.D., Demuth, H.B. and Hagan, M.T. 1996. Neural network design. *Pws, Boston*. (1996).

[15] Todini, E. 2000. Looped water distribution networks design using a resilience index based heuristic approach. *Urban water*. 2, 2 (2000), 115–122.

[16] Prasad T. Devi and Park Nam-Sik 2004. Multiobjective Genetic Algorithms for Design of Water Distribution Networks. *Journal of Water Resources Planning and Management*. 130, 1 (Jan. 2004), 73–82. DOI:https://doi.org/10.1061/(ASCE)0733-9496(2004)130:1(73).

[17] Rossman, L.A. 2000. EPANET 2: users manual. (2000).

[18] Johns, M.B., Walker, D.J., Keedwell, E. and Savić, D.A. 2018. Interactive Visualisation of Water Distribution Network Optimisation. (2018).

[19] Walker, D.J., Johns, M.B., Keedwell, E. and Savić, D.A. 2018. Towards Interactive Evolution: A Distributed Optimiser for Multi- objective Water Distribution Network Design. (2018).

[20] Sherali, H.D., Subramanian, S. and Loganathan, G.V. 2001. Effective Relaxations and Partitioning Schemes for Solving Water Distribution Network Design Problems to Global Optimality. *Journal of Global Optimization*. 19, 1 (Jan. 2001), 1–26. DOI:https://doi.org/10.1023/A:1008368330827.

[21] Alperovits, E. and Shamir, U. 1977. Design of optimal water distribution systems. *Water resources research*. 13, 6 (1977), 885–900.

[22] Fujiwara, O. and Khang, D.B. 1990. A two-phase decomposition method for optimal design of looped water distribution networks. *Water resources research*. 26, 4 (1990), 539–549.

[23] Johns, M., Mahmoud, H., Walker, D., Ross, N., Keedwell, E.C. and Savic, D. 2019. Augmented Evolutionary Intelligence: Combining Human and Evolutionary Design for Water Distribution Network Optimisation. (2019).

[24] Mahmoud, H.A., Johns, M., Keedwell, E. and Savic, D. 2019. Generalising human heuristics in augmented evolutionary water distribution network design optimisation. (2019).

[25] Walker, D.J. 2018. Generating Heuristics to Mimic Experts in Water Distribution Network Optimisation. *WDSA / CCWI Joint Conference Proceedings*. 1, (Jul. 2018).

[26] Breiman, L. 2001. Random forests. *Machine learning*. 45, 1 (2001), 5–32.

[27] Zitzler, E., Laumanns, M. and Thiele, L. 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK-report*. 103, (2001).

[28] Farmani, R., Savic, D.A. and Walters, G.A. 2005. Evolutionary multi-objective optimization in water distribution network design. *Engineering Optimization*. 37, 2 (2005), 167–183.

[29] Wang Qi, Guidolin Michele, Savic Dragan and Kapelan Zoran 2015. Two-Objective Design of Benchmark Problems of a Water Distribution System via MOEAs: Towards the Best-Known Approximation of the True Pareto Front. *Journal of Water Resources Planning and Management*. 141, 3 (Mar. 2015), 04014060. DOI:https://doi.org/10.1061/(ASCE)WR.1943-5452.0000460.

[30] Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A. and Toth, P. 2012. On the optimal design of water distribution networks: a practical MINLP approach. *Optimization and Engineering*. 13, 2 (2012), 219–246.

[31] Reca, J. and Martínez, J. 2006. Genetic algorithms for the design of looped irrigation water distribution networks. *Water resources research*. 42, 5 (2006).

[32] Zitzler, E. and Thiele, L. 1998. Multiobjective optimization using evolutionary algorithms—a comparative case study. *International conference on parallel problem solving from nature* (1998), 292–301.

[33] Coello, C.A.C. and Sierra, M.R. 2004. A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. *Mexican International Conference on Artificial Intelligence* (2004), 688–697.

[34] Mann, H.B. and Whitney, D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*. (1947), 50–60.