Investigating the run-time searching behavior of the differential evolution algorithm applied to water distribution system optimization

by

Zheng, F., Zecchin A.C., Simpson A.R

Environmental Modelling & Software

Citation:
Zheng, F., Zecchin A.C., Simpson A.R, Investigating the run-time searching behavior of the differential evolution algorithm applied to water distribution system optimization, *Environmental Modelling & Software*, Available online 16 October 2014, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2014.09.022.

For further information about this paper please email Angus Simpson at angus.simpson@adelaide.edu.au

Investigating the run-time searching behavior of the differential evolution algorithm applied to water distribution system optimization

Feifei Zheng, Aaron C. Zecchin and Angus R. Simpson

Feifei Zheng: Corresponding author, Research Associate, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia. feifei.zheng@adelaide.edu.au. Tel: +618-8313-1113.

Postal address: N124, Engineering North, School of Civil, Environmental and Mining Engineering, University of Adelaide, North Terrace Campus, Adelaide, South Australia, 5005.

Aaron Zecchin: Lecturer, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia. aaron.zecchin@adelaide.edu.au.

Angus Simpson: Professor, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia. angus.simpson@adelaide.edu.au.

Abstract: In recent years, the differential evolution algorithm (DEA) has frequently been used to tackle various water resource problems due to its powerful search ability. However, one challenge of using the DEA is the tedious effort required to fine-tune parameter values due to a lack of theoretical understanding of what governs its searching behavior. This study investigates DEA's search behavior as a function of its parameter values. A range of behavioral metrics are developed to measure run-time statistics about DEA's performance, with primary focus on the search quality, convergence properties and solution generation statistics. Water distribution system design problems are utilized to enable investigation of the behavioral analysis using the developed metrics. Results obtained offer an improved knowledge on how the control parameter values affect DEA's search behavior, thereby providing guidance for parameter-tuning and hence hopefully increasing appropriate take-up of the DEA within the industry in tackling water resource optimization problems.

Keywords: differential evolution algorithm, evolutionary algorithms, search behavior, water distribution systems, optimization.

1. Introduction

In the water resource community, researchers and engineers often have to deal with various optimization problems. These include hydrological model calibration, the planning, design and operation of water resource systems (Nicklow et al., 2010). The optimization process tries to find the best solution of the given problem within the specified constraints. Optimizing water resource problems are often extremely difficult due to the highly nonlinear and complex decision spaces (Razavi et al., 2012). Although traditional deterministic optimization techniques have been attempted to solve these problems, the results have often been unsatisfactory (Zheng et al., 2011a).

Over the past two decades, there has been a move towards developing or applying various evolutionary algorithms (EAs) to deal with water resource optimization problems (Maier et al., 2014). The differential evolution algorithm (DEA), first proposed by Storn and Price (1995) as one type of EA, has especially received a deal of attention in recent years (details of DEA are given in Section 3). For example, Vasan and Raju (2007) introduced the DEA to optimize flow allocations of an irrigation system; Reddy and Kumar (2007), applied the DEA to reservoir system optimization; Vasan and Simonovic (2010), and Zheng et al. (2013) employed the DEA to optimize the design of water distribution systems (WDSs). More recently, Chichakly et al. (2013) designed watershed-based stormwater management plans using the DEA, and Joseph and Guillaume (2013) used the DEA to calibrate hydrological models. It has been reported in these studies that the DEA exhibited better performance in efficiently finding optimal solutions compared to other types of evolutionary algorithms (EAs), such as genetic algorithms (GAs) and ant colony optimization (ACO) algorithms. This shows that the DEA is promising for dealing with a broad array of water resource optimization problems.

Previous studies have also shown that DEA's search behavior is heavily dependent on the values of the control parameters *F* (the differential weight used in the mutation operator) and *CR* (the crossover probability used in the crossover operator), while it is not significantly affected by the varying population size *N* (Qin and Suganthan, 2005; Das and Suganthan, 2011). Using the same WDS case studies, Suribabu (2010) concluded that the performance of DEA is significantly better than GAs, while Marchi et al. (2014) stated that GAs gave better results overall than the DEA. This contradiction can be explained by the fact that different parameter values, including *F* and *CR*, were used in these DE applications. Zheng et al. (2011b) performed a sensitivity analysis of DEA's parameters (*F* and *CR*) in terms of affecting the final solutions based on two WDS case studies. Results in their study showed that the DEA was unable to solve the optimization problems effectively if inappropriate parameter values were used. This suggests that a set of appropriate parameter values is very critical in obtaining the satisfactory performance of the DEA, which is similar as other types of EAs such as genetic algorithms and the harmony search algorithm (Savic and Walters, 1997; Geem, 2006).

Suitable parameter values when using an EA approach are normally optimization problem-dependent due to the variation of fitness landscapes associated with different problems and problem types (Tolson et al., 2009). Typically, a trial-and-error approach is used to calibrate the parameter values for the DEA applied to given optimization problems in water resources (Reddy and Kumar, 2007). This results in a large computational overhead especially when dealing with real world optimization problems, for which a large number of decision variables are normally involved. The tedious effort required for tuning parameter values has been frequently claimed by practitioners as one of the main reasons for their reluctance to embrace EAs in practice (Geem and Sim, 2010).

In order to address this issue, two potential research directions have been adopted. The first direction is the development of parameter-free EAs. Wu and Walski (2005), for example, proposed a self-adaptive penalty approach within a GA to remove the pre-setting of the penalty multiplier parameter for pipeline optimization. Geem and Sim (2010) proposed a parameter-setting-free harmony search algorithm to optimize the design of WDSs. More recently, Zheng et al. (2013b) proposed a self-adaptive differential evolution algorithm (SADE) to optimize the design of WDSs, in which the two control parameter values *F* and *CR* were adapted along with the evolution of the solutions rather than being pre-specified to fixed values in advance.

The second research direction is the characterization of EA's run-time search properties as a function of the varying control parameter values, thereby providing guidance for fine-tuning parameter values. Traditionally, an EA's search performance is typically assessed based on end-of-run performance measures (i.e. statistics describing the least-cost solution found, and the time taken to find the least-cost solution, see discussions in the position paper by Maier et al., (2014)). A state-of-the-art example of the end-of-run performance analysis is the work by Hadka and Reed (2012), in which a diagnostic assessment framework was developed for evaluating the effectiveness, reliability, efficiency and controllability of multi-objective evolutionary algorithms (MOEAs). In contrast to the extensive research on the end-of run performance assessment, there has been few investigations into characterizing an algorithm's properties from the point of view of the underlying run-time searching performance. The only example of the run-time performance analysis is the work of Zecchin et al. (2012), who investigated the run-time search behavior of various ACO operators applied to WDS optimization problems.

In the context of a parametric study, the end-of-run statistics enable the determination of a direct relationship between an algorithm's parameter settings and overall performance. However, a consideration of the run-time behavioral statistics can provide more insight as to how the different values of the control parameters affect an EA's searching behavior in terms of the exploration (the ability to broadly explore the whole search space) and exploitation (the ability to intensively exploit the promising regions) within the search process (Maier et al., 2014). This insight should provide guidance not only for practitioners to select appropriate parameter values of EAs based on an available computational budget, but also for algorithm developers to understand more deeply the direct and measured impact of parameter variations on search behavior. For example, for applications with a limited computational budget, a set of parameter values should be selected in favor of exploitation. In contrast, if better quality solutions are preferred with relaxed computational constraints, the combination of the parameter values needs to possess more strength on the exploration ability. Building a fundamental understanding of EA's working principles, such as the run-time searching behavior, as opposed to focusing only on the end-of-run performance, is an important future research objective as stated in the position paper by Maier et al., (2014).

As previously outlined, Zheng et al. (2011b) conducted a parametric study on DEA's control parameters (*F* and *CR*), followed by a development of a self-adaptive DE algorithm (Zheng et al., 2013) in order to remove the tedious parameter tuning process. Both studies solely focused on the end-of-run statistics within the given computational budget, ignoring the algorithm's run-time searching properties. Therefore, the question still remains as to why certain algorithms or algorithms with certain parameterization outperformed others for the selected case studies, and how the internal operators and mechanisms alter the DEA's run-time searching behavior that lead up to the end-of-run performance. This paper is such an attempt to address this issue.

To facilitate the search behavior analysis, a range of behavioral measures are developed for the DEA in the current study. The primary run-time statistics of interest concern the population variance, the search quality, the convergence measures, the percentage of the time spent in the feasible and infeasible regions, and the percentage of improved solutions within each generation. The WDS design problem, as one typical type of complex optimization problems in water resources (Fu and Kapelan, 2011), is considered to analyze the search behavior of the DEA with respect to varying parameter values. Three WDS case studies with increased scales and complexity are used in the current study.

Various metrics have been developed to enable the non-dominant set comparison in the multi-objective EA (MOEA) domain. For example, to assess MOEA's final searching performance, Ang et al. (2002) proposed to plot the non-dominated solutions against their distance to the Pareto front and their distance between each other, while Hadka and Reed (2012) utilized a broad range of performance metrics including the hypervolume, the generational distance, the inverse generational distance, the additive epsilon indicator and the spread. These studies have made merit in developing metrics to evaluate MOEA's end-of-run performance, while they significantly differ to the focus of this study that attempts to develop various metrics to measure the DEA's run-time performance in the single-objective space.

Although the impact of different parameter values (*F* and *CR*) on DEA's performance is investigated in this study, it is not intended to derive a quantitive relationship between the case studies (different scales and complexity) and the appropriate parameter values. The aim of the present study is to: (1) develop and demonstrate the utility of metrics for analyzing the DEA's run-time search behavior; and (2) characterize the influence of varying parameter values (*F* and *CR*) on the DEA's searching behavior (exploration and exploitation) through an empirical numerical study, and compare the empirical results with prior theoretical results from Zharie (2002, 2009) using the complex WDS design problems. It is anticipated that such improved knowledge can provide qualitative guidance to design the DEA to possess various exploration and exploitation emphasis according to the problem scales and complexity as well as the available computational budgets.

The remainder of this paper is organized as follows. Section 2 briefly describes the WDS optimization problem, followed by a presentation of the DEA in Section 3. Section 4 presents the developed metrics for measuring the search behavior of DEAs and Section 5 describes the three WDS case studies used in the current study. Section 6 shows the results of search behavior analysis for each case study. Finally, the discussion and conclusions of this paper are outlined in Section 7.

2. Water distribution system optimization problem

Typically, a single-objective WDS design problem is to minimize the system total life cycle costs (pipes, tanks, valves and other components) while satisfying pressure head constraints at each node. Given a WDS design problem involving the selection of n pipe diameters $\mathbf{D} = [d_1, d_2, \cdots, d_n]^{\mathrm{T}}$, this problem can be defined as:

$$\text{Minimize} \qquad f = \sum_{j=1}^{n} c_j(d_j) \qquad (1)$$

Subject to:

$$\mathbf{H}(\mathbf{D}) \geq h_{\min} \qquad (2)$$

$$d_j \in \{A\} \qquad (3)$$

where f=network cost that is to be minimized; d_i=diameter of the pipe i; c_j is the cost function for pipe i associated with the choice of the decision variable d_j; n=total number of pipes in the network; $\mathbf{H}(\mathbf{D})$ represents the performance constraints of the design solution **D**, with $\mathbf{H}(\mathbf{D}) \geq h_{\min}$ showing that the design pressure head at each demand node is above (or equal to) its corresponding minimum allowable pressure head $h_{\min}$ (the determination of $\mathbf{H}(\mathbf{D})$ normally involves a hydraulic simulation model (EPANET2.0 in this study), which solves the nonlinear mass and energy balance equations for flows and heads in the network); and $\boldsymbol{A}$ = a set of commercially available pipe diameters.

3. Differential evolution algorithm

The DEA, first introduced by Storn and Price (1995), is a simple yet powerful EA for global optimization. There are three important operators involved in the process of the DEA, including the mutation operator, the crossover operator and the selection operator. The process of a typical DEA is outlined as follows (Storn and Price, 1995):

Initialization. The DEA is a population based stochastic search technique. Thus, an initial population with a population size N is required to start its search. The individual i ($i = 1, 2....., N$), at generation number G ($G = 0, 1....., G_{\max}$) can be described as a variable vector $X_{i,G} = \left[x_{1,i,G}, x_{2,i,G}, \cdots, x_{n,i,G}\right]^{\mathrm{T}}$, where $G_{\max}$ is the number of the maximum allowable generations, and n is the number of decision variables within the problem. Initial solutions (at $G = 0$) are typically generated by uniformly randomizing individuals within the total search space as

$$x_{j,i,0} = x_{j,\min} + Rand_{i,j}[0,1](x_{j,\max} - x_{j,\min}) \quad i=1, 2, \ldots N, j=1, 2, \ldots n \tag{4}$$

where $x_{j,\min}$ and $x_{j,\max}$ are respectively the minimum and maximum bounds of the j^{th} decision variable; and $Rand_{i,j}[0,1]$ represents a uniform distribution random variable in the range [0, 1], which is generated independently for each decision variable j in the i^{th} vector. $X_{i,G}$ is denoted as a target vector to be improved by the following three operators.

Mutation

The DEA is mainly differentiated from other EAs by its mutation approach, in that a mutant vector $V_{i,G} = \left[v_{1,i,G}, v_{2,i,G}, \cdots, v_{n,i,G}\right]^{\mathrm{T}}$, with respect to each individual $X_{i,G}$, is produced by adding the weighted difference (with differential weight *F*) between two random population members to third member from the current population. This is given as:

$$V_{i,G} = X_{r_1^i,G} + F(X_{r_2^i,G} - X_{r_3^i,G}) \tag{5}$$

where $X_{r_1^i,G}$, $X_{r_2^i,G}$, and $X_{r_3^i,G}$ are three different vectors randomly selected from the current population ($r_1^i \neq r_2^i \neq r_3^i$). These three indexes are randomly selected, without replacement, for each mutant vector.

Crossover

After the mutation, a trial vector $U_{i,G} = \left[u_{1,i,G}, u_{2,i,G}, \cdots, u_{n,i,G}\right]^{\mathrm{T}}$ is generated though selecting solution component values either from $X_{i,G}$ or $V_{i,G}$. The binomial crossover operator is mathematically given as:

$$u_{j,i,G} = \begin{cases} v_{j,i,G}, \text{ if } Rand_{i,j}[0,1] \leq CR \\ x_{j,i,G}, \text{ otherwise} \end{cases} \tag{6}$$

where *CR* (0≤ CR ≤1) is the crossover probability. As shown in Equation (6), if $Rand_{i,j}[0,1]$ is smaller than *CR*, the value $v_{j,i,G}$ in the mutant vector is copied to the trial vector $u_{j,i,G}$, otherwise, the j^{th} trial vector value is inherited from $X_{i,G}$.

Selection

After crossover, all the trial vectors are evaluated using the objective function $f(U_{i,G})$ and compared with their corresponding target vectors $f(X_{i,G})$. The vector with a lower objective function value (given the problem is a minimization problem) survives for the next generation. That is

$$X_{i,G+1} = \begin{cases} U_{i,G}, & \text{if } f(U_{i,G}) \leq f(X_{i,G}) \\ X_{i,G}, & \text{otherwise} \end{cases} \quad (7)$$

The use of the minimization problem in Equation (4) does not lose the generality of the DEA, as a maximization problem can be easily converted to a minimization problem (Das and Suganthan, 2011).

The mutation, crossover and selection operators are repeated aiming to seek the best variable vector X^* as such $f(X^*) < f(X)\,(f:\Omega \subseteq \Re^n)$ holds for all $X \in \Omega$, where Ω is the domain of the search space with the constraints satisfied ($\Omega = \Re^n$ for unconstrained optimization problems). In practice, it is difficult, if not impossible, to find the global optimal solution for a large-scale and complex optimization problem. A limited computational budget, such as a specified number of the maximum allowable generation $G_{\max}$, is normally used as the stopping criteria for the DEA to provide a near-global optimum.

For the design of WDS problem as describe from Equations (1) to (3), the pipe diameters for each of the decision variables can only be selected from the predetermined discrete set ***A*** (see Equation 3). Therefore, the continuous values produced in the initialization and the mutation processes in the DEA were converted to the nearest discrete pipe diameters in *A* following Vasan and Simonovic (2010). The minimum pressure head $h_{\min}$ was taken as the constraints for each demand node, where EPANET2.0 was used to obtain the nodal pressure head for each candidate solution. A penalty approach was adopted (Zecchin et al., 2012) to handle constraints, in which the solutions with constraints (Equation 2) violated (referred as infeasible solutions) were penalized, followed by performing the selection operator (Equation 7).

4. Behavior analysis measures

This study develops a range of behavioral measures to investigate DEA's search behavior as a function of the two control parameters: *F* and *CR*. These measures concern the search quality, the convergence properties, and the run-time solution generation statistics such as the percentage time spent in the feasible and infeasible regions, and the percentage of improved solutions within each generation. Details of these measure metrics are given in following sub-sections.

4.1 Measures of the search quality

Search quality measures are used to characterize both the fitness of a searching population in the objective space, and the closeness of a searching population to the known optimum region in the decision space (Zecchin et al., 2012). For a single objective optimization problem, an important indicator for measuring the search quality is related to the best solution $f_{\min}(G)$ (given the minimization problem) found at each generation *G*, where

$$f_{\min}(G) = \min_{X \in \{X_{1,G},\ldots,X_{N,G}\}} f(X) \tag{8}$$

This measure has been widely used to assess the performance of the search algorithms (see Zecchin et al., 2012 for details). Equation (8) provides the information of the best known solution at each generation in the objective space.

Another fundamental measure of an algorithm's search quality is the production of solutions that are increasingly close to the global optimum X^* in the decision space. The closeness of the searching algorithm (DEA in this study) to the X^* can be quantitively measured using:

$$d_{\min}(G) = \min_{X \in \{X_{1,G},\ldots,X_{N,G}\}} \langle X, X^* \rangle \tag{9}$$

where $\langle X, Y \rangle$ is a topological distance metric between solutions *X* and *Y*. In contrast to the aforementioned variance diversity measures, typically for combinatorial problems an appropriate metric is the Hamming distance, which is given as

$$\langle X, Y \rangle = \left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle = \sum_{i=1}^{n} I(x_i, y_i) \tag{10}$$

and *I* is the indicator function, given by

$$I(x, y) = \begin{cases} 0 & x = y \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

The values in Equation (9) versus the generations provide insight into the search quality of an algorithm, and its ability to search in the near optimal region. The global optimum X^* may be unknown for some case studies. In such situations, the current best known solution can be adopted, which still enables an indication of how the searching approaches the promising regions that contain the good quality near-optimal solutions.

4.2 Measures of convergence

The run-time convergence measures are aimed at quantifying the spread, or distribution, of an EA's population of solutions through the decision space. In the following, first a range of general convergence measures are discussed, followed by the population diversity measures that are adopted to compare our numerical results with the theoretical convergence relationship derived by Zharie (2002, 2009).

4.2.1 General convergence measures

Convergence has often been defined in the objective space where the objective function values (normally consider the best solutions) are not further improved within a specified time-frame (Zheng et al., 2014). However, such a convergence measure may not be valid as a further improvement in the objective function value can be likely when a sufficiently large computational budget is allowed. Alternatively, convergence can be defined as being dependent on the topological distance between solutions within the decision space. Zecchin et al. (2012) proposed a measure in which the mean of the total pairwise Hamming distance between all solutions within the population was used to measure the extent of convergence. This convergence measure has been adopted in the current study. The mean population search distance can be defined as

$$d_{mean}(G) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\langle X_{i,G}, X_{j,G} \right\rangle \tag{12}$$

where $N(N-1)/2$ is the total number of pairs of the candidate solutions.

The mean population search distance in Equation (12) provides a quantitative measure of the spread of solutions over the search domain, with large and low values respectively corresponding to periods of high exploration (broad searching within the decision space) and exploitation (confined focused searching within the decision space). Variation of the mean search distance with the generation indicates the convergence behavior, and can provide insight into how this behavior is influenced by control parameter settings. For example, for a given EA with a certain control parameter setting, if the $d_{mean}(G)$ value decreases particularly quickly, this indicates that such a parameter setting focuses more on exploitation of the identified good information rather than exploration of the search space, and hence it is more likely to converge at a sub-optimal solution. Conversely, if the value of $d_{mean}(G)$ decreases at a very slow rate, this suggests that the given parameter setting is insufficient in exploiting the determined promising regions, and hence, results in a slow convergence.

An alternative measure of convergence is also proposed below. In addition to quantifying the spread of solutions within a decision space, it is also of interest to know whether the search has converged only in some dimensions of the decision space, whilst still broadly searching in the other decision variables. Such information provides meaningful knowledge on the convergence properties of the search process, as, for a given problem, the convergence speed of one decision variable may differ to the others due to the variation in the separability of a given dimension, and the sensitivity of the objective to a given variable. This metric is referred as the number of converged decision variables NC, which is mathematically expressed as:

$$NC(G) = \sum_{j=1}^{n} \left(1 - \mathrm{sgn}(\mathrm{Var}_{j,G})\right) \tag{13}$$

where sgn() is the signum function, and $\mathrm{Var}_{j,G}$ is the sample variance of the population for decision variable *j*, that is $\{x_{j,1,G}, \ldots, x_{j,N,G}\}$. The term $\left(1 - \mathrm{sgn}(\mathrm{Var}_{j,G})\right) = 1$ when $\mathrm{Var}_{j,G}$ =0, suggesting that all solutions in the current generation *G* have selected the same value for the j^{th} decision variable, and this decision variable has converged. It is noted that signum function used here only has two alternatives with 0 and 1 due to the non-negative of the variance value ($\mathrm{Var}_{j,G}$).

The measure $NC(G)$ gives the number of decision variables (or search space dimensions) that have been converged at generation G. It is straightforward to record which decision variables have converged, as such information can provide a better understanding of which decision variables are difficult to determine, allowing for the guided use of preconditioning or local search to facilitate in the determination of these variables. Other methods have previously been developed to measure the convergence status of the decision variables within the searching process, in order to perturb selected solutions to avoid premature convergence (Geem et al., 2001). For instance, some researchers used the pattern recognition approach to extract pattern(s) from the solutions in the later searching period, during which most solutions tend to resemble each other (i.e., many solutions have identical values for the same decision variables, see Michalski and Wojtusiak (2012) for details).

4.2.2. Measures of the population variance

The population variance can be characterized by the averaged variance of all solution components in the population $\mathrm{Var}_X(G)$ (e.g., Zharie, 2002; Das and Suganth, 2011), that is

$$\mathrm{Var}_X(G) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{Var}\left\{x_{j,1,G}, \ldots, x_{j,N,G}\right\} \tag{14}$$

Typically, a high value of $\mathrm{Var}_X(G)$ indicates that the search is spread broadly throughout the decision space, and for a relatively low value, the search is focused on small areas. It is noted that the $\mathrm{Var}_X(G)$ only measures the overall variance of the population, and cannot provide the searching status of each decision variable (i.e., it may suffer from scaling issues if some variables are much larger than others in magnitude). In the current study, all the decision variables have the identical diameter options for each case study. Therefore, the $\mathrm{Var}_X(G)$ can provide a proper measure of the DEA's population variance within each generation.

It is important to note that both the $d_{mean}(G)$ (Equation 12) and the $\mathrm{Var}_X(G)$ (Equation 14) can be used to indicate the convergence properties of the searching algorithm, but differing in that the former considers the total pairwise Hamming distance while the latter uses the variance of all solutions in the decision space. The main reason for the introduction of $\mathrm{Var}_X(G)$ is to compare the empirical searching variance (this study) with the theoretical work of Zaharie (2002, 2009).

Motivated by the theoretical work of Zaharie (2002), to enable a more detailed analysis of the DEA operators, three forms of population variance were considered, namely: (1) the variance after the mutation operator, $\mathrm{Var}_V(G)$; (2) the variance after the crossover operator $\mathrm{Var}_U(G)$; and (3) the variance after the selection operator, $\mathrm{Var}_X(G)$. Zaharie (2002) studied the impact of *CR* and *F* on Var_V(G) and Var_U(G), and derived theoretical expressions for the expected value of these variance as a function of the population variance at the end of the previous iteration

$$\mathrm{E}\left[\mathrm{Var}_V(G)\right] = \left(2F^2 + \frac{N-1}{N}\right)\mathrm{Var}_X(G-1) \tag{15}$$

$$\mathrm{E}\left[\mathrm{Var}_U(G)\right]=\left(2F^2CR-\frac{2CR}{N}+\frac{CR^2}{N}+1\right)\mathrm{Var}_X(G-1) \quad (16)$$

where it is seen that the DEA parameters (F and CR) influence the expected variances by providing a scaling factor to the parent population variance at generation G-1. It is important to note that the influence of the selection operator on the population variance is not amenable to theoretical analysis, and hence no theoretical estimates exist relating $\mathrm{E}\left[\mathrm{Var}_X(G)\right]$ to $\mathrm{Var}_X(G-1)$.

Typically, to ensure reasonable performance, DEA's population variance should be increased after the mutation and crossover operators, followed by a decrease owing to the selection action (Zaharie, 2009). The rationale behind this is that an algorithm with an expanded exploration after mutation and crossover is more likely to find promising regions (especially for large search spaces) and the variance reduction caused by the selection permits intensive searching in the identified small regions. Such an interactive process with appropriate balance (i.e., trade-off between exploration and exploitation) has been demonstrated to be effective in guaranteeing EA's performance (Zecchin et al., 2012). Motivated by this, DEA's parameter combinations that satisfy the equation

$$2F^2CR-\frac{2CR}{N}+\frac{CR^2}{N}=0 \quad (17)$$

can be considered to be critical since they result in a population whose variance remains constant overall. If the influence of the selection is removed (i.e., all the trial solutions are accepted), Equation (17) predicts that F will display a critical value, F_c, such that the population variance decreases when $F< F_c$ and increases if $F> F_c$. By solving Equation (17), F_c can be described as (Zaharie, 2009)

$$F_c = \sqrt{\frac{1 - CR/2}{N}} \tag{18}$$

The F_c establishes a lower limit for F in the sense that smaller value will induce premature convergence even on a flat objective function landscape (Zaharie, 2009). One contribution of this study is to validate these theoretical predictions (Equations 15, 16 and 18) using the large and complex WDS design problems.

4.3 Solution generation statistics

Constraints are often involved in water resource optimization problems, contributing to the complexity of the search spaces. A candidate solution that satisfies all constraints is a feasible solution, otherwise it is termed an infeasible solution. The efficiency with the feasible solution found is also an important assessment criteria for searching algorithms especially when dealing with complex water resource optimization problems. Following Zecchin et al (2012), in the current study, we use the percent of the solutions (*PF*%) in the feasible region to measure DEA's search behavior and investigate how this varies throughout the algorithm's total search time. For the WDS design problems considered in the current study, the feasible solutions need to satisfy Equation (2), where the nodal pressure heads should be no less than the minimum allowable pressure head $h_{\min}$.

An additional measure providing an important statistic of the DEA's behavior is the percentage of improved solutions found by the DEA within each iteration (*PI*%). The DEA constructs solutions through the processes of mutation, crossover and selection, where only improved solutions are selected, and all others are rejected (see Equation 7). Consequently, considering the percentage of improved solutions provides a measure of the effectiveness of the mutation and crossover operators in terms of generating improved solutions.

It is highlighted that the six measure metrics were designed to measure the DEA's run-time searching behavior from different aspects, thereby offering a comprehensive assessment on how parameter settings influence the evolution of DEA's search, such as convergence behavior and productive solution improvement stages throughout the searching process. As anticipated, and observed in the results, although the two solution-quality measures have a high correlation with one another, their measurements are in different spaces (i.e. $f_{\min}(G)$ is an objective space measure, and $d_{\min}(G)$ is a decision space measure). Similarly, the three convergence measures possess a high correlation as they are all indicators of convergence as shown in the results, but, again they are measuring different aspects: the mean population distance $d_{mean}(G)$ deals with the Hamming distance metric and considers pairwise differences between all solutions, the variance measure uses a Euclidian distance metric in the decision space to facilitate comparison with the theoretical work of Zaharie (2002, 2009), and the *NC(G)* measure considers the number of decision space dimensions for which there is no exploratory activity. This study adopts this array of measures to attempt to give a broad description of run-time behavior, where relationships between the measures are discussed. However, a focused comparative study of the measures themselves is not the focus of this work.

5. Case studies

Three different WDS case studies with different sizes and complexity were optimized by the DEA with varying control parameter values in the current study. These case studies were the Hanoi Problem with 34 decision variables (HP_{34}), the ZJN network with 164 decision variables (ZJN_{164}) and the Balerma network with 454 decision variables (BN_{454}). Note that the subscript number is the number of decision variables for each case study. Details of the HP_{34}, ZJN_{164}, and BN_{454} case studies can be found in Fujiwara and Khang (1990), Zheng et al., (2011a) and Reca and Martínez (2006) respectively. The current best known solution for HP_{34} was first reported by Reca and Martínez (2006) with a cost of \$6.081 million. Zheng et al. (2011a) found the current best known solutions for the ZJN_{164} and BN_{454} case studies with cost of \$7.082 million and €1.923 million respectively. The total search space and the current best known solution for each case study are given in Table 1.

The behavioral metrics given in Section 4 were used to analyze the behavior of the DEA with varying control parameter values (F and CR). The population sizes (N) used for each case study was given in Table 1, which was selected based on the guidance provided in Zheng et al. (2013). The current best known solutions for these case studies were assumed to be the global optimums to enable the behavior analysis given in Equation (9).

The computational experiments were designed as follows: for each case study, a value of CR=0.5 was first used to explore DEA's search behavior as a function of varying the differential weight F including $F \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ using the developed metrics; then the value of F that yielded robust results within the given computational budget was used to investigate the search properties of the DEA with respect to the varying CR with $CR \in \{0.2, 0.4, 0.6, 0.8\}$. To understand the interactions between F and CR, we also investigated the DEA's performance for each $F \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ paired with each $CR \in \{0.1, 0.5, 0.9\}$ using the proposed metrics. It is highlighted that the focus of this paper is to explore DEA's searching behavior as a result of the varying parameter values rather than empirically calibrating the parameters for each case study.

For each parameter set, 20 runs were performed with different random number seeds for each case study. A preliminary analysis of the raw time-series behavioral data demonstrated that, despite some variation across different seeds, the overall trends were similar. As such, the averaged results over the 20 runs are presented in this paper in order to provide a more statistically meaningful characterization of the searching behavior. However, it is important to note that Kollat and Reed (2006) found that when comparing the end-of-run performance (not the run-time behavior as did in this paper) of multi-objective algorithms, performance difference across random number seeds for a single parameterization could be more influential than variations in the parameter settings.

In the current study, a penalty cost approach was used to handle constraints, where the penalty cost was set to equal the value of the penalty factor R ($R=10^8$ in this study), multiplied by the maximum violation of the pressure constraints (Zecchin et al., 2012). Although DEA's performance was affected by the penalty multiplier, for large R this performance difference is insignificant, and as such the use of the identical $R=10^8$ was adopted for each parameter set to enable a fair comparison. The detailed investigation on the penalty multiplier values is beyond the scope of the current study.

6. Results and Discussion

The presentation of the results is structured as follows. Firstly the influence of the differential weight *F* and the crossover probability *CR* on the changes in population variance, as induced by the operations of mutation and crossover, are discussed in Section 6.1. Within this section, comparisons between our empirical results and the theoretical work of Zaharie (2002, 2009) are presented. Following this the independent influence of *F* and *CR* on the search behavior as measured by the run-time metrics from Section 4 is explored in Sections 6.2 and 6.3. Within each parameter themed section, the discussion covers first the search quality measures, followed by the convergence measures and finally the solution generation statistics. The interaction between *F* and *CR* is covered in Section 6.4, followed by a discussion of the main practical findings in Section 6.5

6.1 Influence of *F* and *CR* on changes in population variance

The population variance measure results presented in Figure 1 are normalized by the variance of the seeded solutions at G = 1 for the HP_{34} case study. Two general trends can be observed from this figure. Firstly, a mild increase in the population variance occurs before a continual reduction, implying that the DEA initially expands its exploration before convergence. Secondly, the mutation operation (black lines) serves to increase the population diversity, and that the actions of crossover (red lines) and selection (blue lines) decrease the population variance, that is $\mathrm{Var}_V(G) > \mathrm{Var}_U(G) > \mathrm{Var}_X(G)$.

As seen in Figure 1, an increase in the differential weight *F* offers larger population diversity after mutation, encouraging greater exploration. This increase in diversity is expected as *F* controls the amount that a randomly selected trial vector is perturbed by the difference between two other chosen vectors (see Equation 5). Consequently, a larger

F will result in a larger perturbation of the selected vector, and will serve to encourage a broader exploration of the search space.

It is also observed from Figure 1 that a higher value of *CR* maintains greater population variance (red lines) after crossover relative to the parent population (blue lines). This can be explained that low values of *CR* imply that an increased number of solution components are inherited from the less diverse parent population than the more diverse mutant population. These observed results are qualitatively consistent with the theoretical predictions of Zaharie (2002) that the expected variance after mutation and crossover increases for larger *F* and *CR* (Equations 15 and 16). Similar observations were made for other parameter sets and other case studies. The variance metric itself provides only convergence-based information (i.e. how diverse the population of solution is) but provides no information about the quality of the search (for example a rapid decrease in the variance can be caused by either a premature convergence or a quick identification on the global optimum). Therefore, other metrics, such as the solution quality, have to be used to assist the diagnostic assessment of an algorithm's searching behavior.

Figure 2 show the surface plots of the errors (percentage relative difference %) between the theoretical variance ratios in Equations 15 and 16 and the experimental results for all parameter combinations used for the three case studies. Considering Figure 2, it can be concluded that the differences between the theoretical predictions and the experimental observations are consistently minor for all *F* and *CR* values applied to the three case studies (within 2%), suggesting an excellent quantitative agreement. This indicates that the theoretical predictions (Equations 15 and 16) can be used to quantitatively determine the changes in population variance as a function of the *F* and *CR* values, intuitively meaning that larger values of F and CR produce larger variances in solutions after mutation and crossover operators.

Following the work of Zaharie (2009), we turned off the selection operator (e.g., all trial solutions are accepted) to validate the theoretical lower limit of the F (F_c in Equation 18) using the three case studies. Figure 3 shows the parameter values (N, F and CR), the theoretical values of F_c (black lines) and the population variances for each case study. It is seen from Figure 3 that the population variances with F_c are overall constant against generations for each case study despite expected random fluctuations, while for $F>F_c$ and $F<F_c$, the population variances respectively increase and decrease. These experimental results are again matched well with the theoretical lower bounds of the F value (Equation 18).

In practice, the objective function landscapes are seldom flat, and hence F must be larger than F_c to counteract the additional reduction in variance that selection induces. Such an improved knowledge can provide guidance for selecting the appropriate parameter values when applying DEAs to water resource optimization problems.

6.2 Influence of differential weight *F*

Numerical results for the study of the influence of F on the DEA's behavior are given in Figures 4 to 6. As shown in these figures, the DEA's improvement in search quality for increasing generation number is observed by the decreasing objective function value in subfigures (a) ($f_{\min}(G)=\min_{X\in\{X_{1,G},\ldots,X_{N,G}\}} f(X)$), and the decreasing distance to the known optimal solution in subfigures (b) ($d_{\min}(G)=\min_{X\in\{X_{1,G},\ldots,X_{N,G}\}} \langle X,X^{*}\rangle$). Despite the existence of many local minima for WDS design problems arising from the nonlinear constraints in Equation (2), the highly correlated pattern within these figures (i.e. objective function cost is highly correlated to distance from the global minima) implies that the objective surface possesses a large valley type structure, with the local minima clustered within a small sub-region of the decision space.

A notable general feature observed within the quality metrics is that, in most cases, the DEA achieves the majority of its solution improvement within the very early stages of the search. A consideration of subfigures (a) and (b) in Figures 4 to 6 shows that a decreasing F drives a faster initial solution improvement, which is consistent with the increased exploitative behavior expected for low F. For the small case study (Figure 4), the explorative higher values of F resulted in an improved performance, as premature convergence to sub-optimal solutions was observed at lower F values. In contrast, the lower values of F improve DEA's performance for the larger case studies for the given run-times (Figures 5 and 6), as the rate of solution improvement for the higher values of F was significantly slower due to the relatively larger focus on exploration.

The convergence of the DEA is represented by the decreasing values in the mean population distance ($d_{mean}(G) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left\langle X_{i,G}, X_{j,G} \right\rangle$ in subfigures c), and by the increase in the number of converged variables ($NC(G) = \sum_{j=1}^{n} \left(1 - \mathrm{sgn}(\mathrm{Var}_{j,G})\right)$ in subfigures d). The metric $d_{mean}(G)$ (Equation 17) represents the overall population spread, while the $NC(G)$ (Equation 18) provides the number of converged decision variables in the search space. As shown in Figures 4 to 6, an inverse similarity of the pattern within subfigures (c) and (d) is observed, suggesting that both measures are equally indicative assessment of the search convergence. Compared to $d_{mean}(G)$, the metric $NC(G)$ offer greater insight in the search properties since it can provide the information on the number of converged decision variables as well as which decision variables have converged. Such improved understanding can provide guidance to apply preconditioning methods (e.g. local search or engineering judgment) to the decision variables that are difficult to converge, thereby enhancing the efficiency of the whole search process. This is especially useful when dealing with difficult or large-scale optimization problems.

As observed in subfigures (c) and (d) in Figures 4 to 6, the rate of convergence decreases for increasing F, which is consistent with the theoretical predictions of a greater explorative behavior for higher F. This inability to converge for high F is most markedly observed for the larger case studies (Figures 5 and 6). For the smaller case study, the fast convergence led to only sub-optimal solutions being found. In contrast, in the larger case studies, the fast convergence associated with the lower values of F enabled the DEA to exploit good information and drive the search into the near optimal regions (see subfigures (a) and (b)) using the given computational budget. However, it should be noted that the use of the larger F may be more likely to find further improved solutions (or global optimum) for large-scale optimization problems if a sufficiently large computational budget is allowed (Zaharie, 2002).

As observed in subfigures (e) of Figures 4 to 6, the DEA tended to generate mainly infeasible solutions within the very early stages of the search followed by an increase in the search effort to exclusively focus in the feasible region. For the ZJN_{164} and BN_{454}, the convergence to the feasible region is very rapid. However, consistent with previous findings (Zecchin et al., 2012), an increased search effort was required to locate and focus the search within the feasible region for the small-scale, but difficult, HP_{34} case study, owing to the notably small feasible region for this problem. From Figures 4 to 6 (subfigures (e)), it can also be observed that the fast convergence associated with the low values of *F* serves to more rapidly drive the search effort into the feasible region.

Subfigures (f) in Figures 4 to 6 show the percentage of improved solutions (*PI*%) against the generations for each parameter sets applied to the three case studies. The value of *PI*% determined during the selection operation within the DEA typically varied from around 50% in the early stages of the search, following a relatively rapid decline to less than 10% in the intermediate to longer stages of the search. It was also detected that a lower *F* can produce a larger value of *PI*% in the early stages of the search for all case studies, followed by a quick decline or even *PI*% =0 in the later stage. In contrast, a mildly increased emphasis on mutation (higher *F*) was observed to lead to a sustained, albeit low, solution improvement in later stages of the search. Despite the lower values of percentage of improved solutions in these later stages, this residual exploration for high *F* values was demonstrated to be an effective strategy for improving the solution quality for the difficult (small feasible region) HP_{34} and the large and more complex BN_{454}, as observed by the solution quality measures in subfigures (a) and (b) demonstrates.

Considering the solution quality plots (subfigures (a) and (b) of Figures 4 to 6) and the *PI*%, it can be concluded that lower values of *F* are more likely to be trapped by local optimal solutions, although they can produce better quality solutions and large $PI_{\%}$ values in the early stages of the search. This is because the search of the DEA with a lower value of *F* is dominated by exploitation, resulting in premature convergence. In contrast, the

DEA with larger F values possess higher likelihood to find better solutions due to the greater exploration, but at expense of dramatically increased computational overheads, especially for the larger case studies (Figures 5 and 6).

Based on all measures presented in Figures 4 to 6, it was found that DEA's performance (solution quality and the convergence speed) is more sensitive to the selected F values when the size of the optimization problem becomes larger. For the HP_{34} case study, the measures of distance to the global optimum ($d_{min}(G)$, equation (9)) and the mean population distance (d_{mean} (G), equation (12)) decreases with the increasing generations even if an extremely large F=0.9 was used, while these two measures stay approximately constant for the large case studies ZJN_{164} when $F \geq 0.7$ and BN_{454} when $F \geq 0.5$ within the given computation-frame.

6.3 Influence of crossover probability *CR*

Based on results in Section 6.2, we selected the F values that produced the best final solution quality (solution quality measure) within the given computational budget for the case studies to enable the influence study of varying CR on DEA's searching behavior. The results are given in Figures 7 to 9. As observed in subfigures (a) and (b) of these three figures, higher values of CR drive a faster improvement in the solution quality, but leading to premature convergence. The rapid solution improvement in the early searching stages for high values of CR indicate the effectiveness of the increased emphasis on exploration in the crossover operator in these early stages. However, despite the initial slower solution improvement rates, lower values of CR tended to yield better quality solutions in the later stages of the search, indicating the importance of an exploitative crossover operator in these later stages, as for low *CR* values, the DEA searches more intensively in the neighborhood close to the parent population.

It is observed that DEA's performance in terms of the solution quality exhibits a low sensitivity to varying *CR* values as long as appropriate *F* values have been given, especially for the large case studies (ZJN_{164} and BN_{454}). This suggests that DEA's search quality is more controlled by the mutation operator compared to the crossover operator when dealing with large and complex search spaces.

From subfigures (c) and (d) in Figures 7 to 9 it is seen that, consistently for all case studies, the convergence speed increased for higher values of *CR*. This finding, on the surface, contradicts the theoretical analysis from Zaharie (2002), in which an increase in *CR* was demonstrated to enhance the population variance and hence would be expected to slow down the convergence. Such a counter intuitive finding was also observed in the numerical studies of Montgomery and Chen (2010), who found for a range of test functions, higher *CR* led to faster convergence rates.

To explain this paradox, it is important to note that the increase in population variance (i.e., diversity) caused by the crossover operator occurs prior to the selection operator. As a consequence, this diversity increase cannot be directly translated to an increase in the diversity of the next generation of solutions. Our experiments results showed that such diversity increase associated with larger values of *CR* was significantly reduced through the selection operator. This can be explained by that larger *CR* encourages larger exploratory moves that are less likely to consistently generate improved solutions, and consequently, the selection operator drives the search toward the already found sub-optimal solutions quickly, resulting in premature convergence (Montgomery and Chen, 2010). In contrast, smaller exploratory moves (small perturbations about the parent population) are associated with smaller *CR* values, which, although more likely to produce improved solutions in later generations, will result in a slower and more gradual convergence of the algorithm. As to be expected, the rapid convergence (large *CR* values) typically led to sub-optimal solutions, where the slower and more gradual convergence for the lower *CR* values ($CR \leq 0.4$) typically led to improved longer term solutions.

Figures 8 and 9 (subfigure (e)) show that the faster converging DEAs (large *CR*) drive the search effort more rapidly into the feasible region, except for the case of the HP_{34} (Figure 7) where the most rapid convergence appears to retain a high percentage of the search effort in the infeasible region until the later stages of the search. This anomaly can be understood by considering the combination of the small feasible region in the HP_{34} with the larger exploratory moves associated with the larger CR=0.8. Namely that the exploratory moves still tended to generate infeasible solutions quite late in the search due to the small feasible region associated with the HP_{34}, whilst for the low *CR*, the smaller exploratory moves in the neighborhood of the parent population tended to yield a higher percentage of feasible solutions.

The trends in the percentage of improved solutions (*PI*%) observed in subfigures (f) in Figures 7 to 9 supports the explanation of the highly convergent behavior for large *CR* and the corresponding slow convergence for low *CR*. Namely, the smaller exploratory moves associated with lower values of *CR* result in a higher percentage of improved solutions found albeit at the cost of convergence rate and a slower overall rate of solution improvement.

It is interesting to note that there is a sudden increase in the values of *PI*% immediately before DEA's convergence as shown in subfigures (f) in Figures 7 to 9 (also Figure 10 in the next section), with the occurrence time dependent on the values of *CR* (i.e. larger CR values tend to have this phenomenon earlier due to the faster convergence). This is especially apparent for large case studies ZJN_{164} and BN_{454}. This phenomenon can be attributed to the fact that the search is restricted within a particular region, in which a near globally optimal solution is located, and no better solutions are found after a long time exploration. As such, due to the selection pressure, all the population members converge to this identified optimal solution quickly and hence a large percentage of improved solutions (*PI*%) can be expected for this short process.

6.4 Interaction between *F* and *CR*

The parametric behavioral analysis presented in Sections 6.2 and 6.3 represents a detailed study of the individual parameters under a *ceteris paribus* assumption. In this section, we extend our study by considering the interaction between the two parameters *F* and *CR*, and how their combined variation influenced the DEA's searching behavior. Within the previous sections, the parameter considered was varied throughout its feasible range, whilst the other parameter was held constant at typical standard values. In the following, to study the interactive effects between *F* and *CR*, the DEA's behavior for low and high values of *F* with a range of *CR* values is considered. These parameter combinations were applied to the three case studies and their performance was measured using the three selected metrics only: the search quality (minimum cost), the convergence

($d_{mean}(G)$), and the percent of improved solutions (*PI*%). Figures 10 and 11 respectively show the results of *F*=0.1 and 0.9 with three different *CR* values (*CR* = 0.1, 0.5, and 0.9).

Considering the low *F*=0.1 (Figure 10), it was found that the lower values of *CR* were able to find better quality solutions in the later searching stages, while the large *CR* values tended to prematurely converge in a rapid speed for each case study. Based on the obtained knowledge in Sections 6.2 and 6.3, this is expected as both small *F* and large *CR* were exploitation encouraging, resulting in quick convergence to sub-optimal solutions. In contrast, the lower *CR* values with relatively stronger explorative ability were able to counteract the powerful exploitative behavior associated with the small *F*, leading to a better balance between exploration and exploitation, and accordingly an improved performance (solution quality) in the later searching phases. As shown in Figure 10, a DEA with a small *F* combined with a large *CR* is more likely to produce improved solutions (i.e., higher *PI*%) at the initial searching stage due to its great exploitative ability, but followed by a rapid decrease in *PI*% caused by the premature convergence. As for the results for the tuned *F* values in Figures 7-9, the larger *CR* values tend to converge faster when a rather low *F* value (*F*=0.1) was used for the three case studies as shown in Figure 10.

When a very strong explorative searching is used (*F*=0.9), a large value of *CR* is expected to offer a relatively better trade-off between exploration and exploitation, since larger *CR* values are demonstrated to be more exploitation emphasizing, as in Section 6.3. This is reflected by the fact that relatively larger *CR* (e.g., *CR*=0.5 and 0.9) yielded better quality solutions, faster convergence speed, and a slightly larger percent of improved solutions compared to *CR*=0.1 for the relatively small HP_{34} case study in the later generations (see Figure 11).

Interestingly, for the two large case studies ZJN_{164} and BN_{454} (also the early searching stage of the HP_{34} case study), *CR*=0.1 produced better quality solutions than *CR*=0.5 and 0.9 due to its higher likelihood to find improved solutions within a small neighborhood of the parent population (low CR values are associated with small exploratory perturbations on the parent population). This is because a high *CR* value combined with a large *F*=0.9 produces too large explorative moves at the initial searching phase, and hence is less likely to find the promising regions especially for large scale optimization problems. This results in a slow convergence speed as well as a slow improvement in the solution quality as demonstrated in Figure 11. Therefore, when a very large *F* value is used (e.g., *F*=0.9), a relatively smaller *CR* value is more appropriate in offering good quality solutions within a limited computational time-frame, although a large *CR* is more likely to find better solutions if the computational resource is sufficient.

A notable observation made from Figures 10 and 11 is that the crossover probability (*CR*) becomes more important to DEA's performance when an improper *F* value is used, compared to its reduced importance in the case of a tuned *F*, as demonstrated in Section 6.3. This suggests that the interaction strength between the two control parameter values is *F* value dependent. As shown in Figure 10, a relatively low *CR* value was able to appreciably enhance DEA's performance when a very low *F* was used. In the case of a very large *F* being assigned, a relatively small *CR* significantly improved DEA's searching effectiveness in the early stages, especially for the large case studies (Figure

11). However, a variation of *CR* had limited influence on DEA's searching quality when an appropriate *F* was used as illustrated in Section 6.3.

6.5 Behavior result discussions

Based on the comprehensive analysis of DEA's search behavior using the six developed metrics, the results obtained can provide guidance for parameter tuning or design of adaptive parameter algorithms as outlined below.

The best parameter sets for a given optimization problem are dependent on the computational budget. If a larger computational resource is allowed, a relatively larger value of *F* combined with a lower value of *CR* is expected to produce better quality solutions. Care is needed to determine appropriate *F* for very large case studies as the sufficient computation time-frame associated with large values of *F* for finding good quality solutions can be extremely large (see Figure 11). As such, a low to middle range ($F \leqslant 0.5$) is recommended for large and complex optimization problems.

If the computational budget is limited, a milder mutation strategy with low values of *F* (if say *F*= 0.2 or 0.3, which emphasis on local search) and a moderate value of *CR* (e.g., *CR*=0.5 or 0.6) can be used to emphasize the DEA's exploitative behavior in the early stages of the search, thereby offering sub-optimal solutions quickly. Such parameter combinations can be typically used to deal with realistic optimization problems in water resources (if say a WDS design problem with 10000 decision variable), since, for these complex problems, finding reasonable solutions of low cost with the given time-frame is more important than locating the global optimum that requires significant computational overheads.

Although it is difficult, if not impossible, to quantitively establish the relationship between the scales/complexity of the case study and the best parameter values, the improved understanding of DEA's searching behavior suggests that a relatively lower value of *F* (e.g., *F*= 0.2 or 0.3) combined with a relatively larger *CR* (e.g., *CR*=0.5 or 0.6) are mostly likely to provide good near-optimal solutions for a larger case study (i.e., a larger search space) within the computational budget that is typically available in practice.

The observed run-time search behavior of the DEA can provide guidance to design more powerful algorithms (e.g. self-adaptive DEA). For example, a self-adaptive DEA may dynamically balance the exploration and exploitation by adjusting the values *F* and *CR* values throughout the entire search process: strengthen the exploitation in the early stage using a smaller *F* and a larger *CR*, while force a stronger explorative searching behavior with a larger *F* and a lower *CR* in the later stage.

7. Summary and Conclusions

The research presented in this paper provided a detailed study of the behavior of DEAs as influenced by the controlling parameters of *F* (differential weight) and *CR* (crossover probability) applied to the classical civil engineering water distribution design problem. The run-time behavioral metrics have considered the solution quality measures, convergence measures, and other search properties such as percentage of effort spent in the feasible region, and the generation-wise percentage of improved solutions (from one generation to the next). Three case studies with ranging from 34 to 454 decision variables have been used to enable the investigation of DEA's behavior analysis. These include HP_{34}, ZJN_{164} and BN_{454}, where the subscript represents the number of decision variables.

The six developed measure metrics have effectively characterized DEA's run-time searching behavior, thereby providing great insight on how the control parameters (*F* and *CR*) alter the DEA's searching performance. Such improved understanding can provide useful guidance in determining the appropriate parameter values with reduced computational effort relative to the traditional trial-and-error approach (detailed discussion are given in Section 6.5). In addition to the practical implications, this study also offers important new findings and contributions, which are summarized in the following.

1. Excellent agreement between predicted and observed population variance as well as the lower bound of the mutation parameter *F* has been found in this study, indicating the practical utility of the theoretical work of Zaharie (2002, 2009). Such improved knowledge can provide guidance in selecting appropriate parameter values for DEAs applied to complex engineering optimization problems. To authors' knowledge, this is the first time that these theoretical results have been validated using complex water resource optimization problems.
2. The interaction strength between the *F* and *CR* parameters is varied as a function of the *F*, where the *CR*'s impact is limited in terms of the searching quality (see Figures 7-9) when a proper *F* value is used. For an inappropriate *F* value, the influence of *CR* becomes more significant (Figures 10 and 11). This indicates that DEA's performance is more dominated by the parameter *F*. This finding is important knowledge for the fine-tuning of DEA's parameters. For example, when the computational resource is limited, the practitioners may fix using a moderate

crossover strength (e.g. *CR*=0.5), and only tune the values of *F* to ensure a well performing DEA.

3. It was found that a very high *CR* value (*CR*>0.8) often reduce DEA's diversity with a rapid speed, resulting in a high likelihood of premature convergence. In contrast, a very low *CR* (*CR*<0.2) is typically slow in convergence, although it is more likely to offer better solutions when the computational budget is sufficient. Therefore, a moderate value of $CR \in [0.3, 0.6]$ (combined with a tuned *F*) is highly likely to provide effective trade-offs between exploration and exploitation within the typically available computational budget. This finding significantly differs to the previous work that a *CR*=0.9 was considered as the *de facto* standard crossover strength for DEAs (Zaharie, 2009) and GAs (Savic and Walters, 1997; Reca and Martínez, 2006).

The developed metrics in this study transcend the specific algorithms and can be used to analyze the search behavior of any types of evolutionary algorithms (EAs). It is expected that the improved knowledge of EAs' working principles obtained using these metrics can enhance their appropriate take-up within the industry in handling various water resource optimization problems. One important future study is to investigate the sensitivity of the DEA's run-time searching performance as a function of varying formulations of water network design problems.

References

Ang, K.H., Chong, G., Li, Y., 2002. Visualization technique for analyzing non-dominated set comparison. Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and

Learning, vol. 1, edited by L. Wang et al., 36-40, Nanyang Technical University, Orchid Country Club, Singapore.

Chichakly, K.J., Bowden, W.B., Eppstein, M.J., 2013. Minimization of cost, sediment load, and sensitivity to climate change in a watershed management application. Environmental Modelling & Software 50(0) 158-168.

Das, S., Suganthan, P.N., 2011. Differential Evolution: A Survey of the State-of-the-Art. Evolutionary Computation, IEEE Transactions on 15(1) 4-31.

Fujiwara, O., Khang, D.B., 1990. A two-phase decomposition method for optimal design of looped water distribution networks. Water Resour. Res. 26(4) 539-549.

Fu, G., Kapelan, Z., 2011. Fuzzy probabilistic design of water distribution networks. Water Resour. Res. 47(5) W05538.

Geem, Z.W., Kim, J.H., Loganathan, G., 2001. A new heuristic optimization algorithm: harmony search. Simulation 76(2) 60-68.

Geem, Z. W., 2006. Optimal cost design of water distribution networks using harmony search. Engineering Optimization 38(3) 259 - 277.

Geem, Z.W., Sim, K.-B., 2010. Parameter-setting-free harmony search algorithm. Applied Mathematics and Computation 217(8) 3881-3889.

Hadka, D., Reed, P., 2012. Diagnostic Assessment of Search Controls and Failure Modes in Many-Objective Evolutionary Optimization. Evolutionary Computation 20(3) 423-452.

Joseph, J.F., Guillaume, J.H.A., 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. Environmental Modelling & Software 46(0) 292-298.

Kollat, J.B., Reed, P.M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Advances in Water Resources 29(6) 792-807.

Maier et al., 2014. Evolutionary algorithms and other metaheuristics in water resources: current Status, research Challenges and future Directions, submitted to Environmental Modelling & Software.

Marchi, A., Dandy, G., Wilkins, A., Rohrlach, H., 2014. Methodology for Comparing Evolutionary Algorithms for Optimization of Water Distribution Systems. Journal of Water Resources Planning and Management 140(1) 22-31.

Michalski, R.S., Wojtusiak, J., 2012. Reasoning with unknown, not-applicable and irrelevant meta-values in concept learning and pattern discovery. Journal of Intelligent Information Systems 39(1) 141-166.

Montgomery, J., Chen, S., 2010. An analysis of the operation of differential evolution at high and low crossover rates, IEEE World Congress on Computational Intelligence, 881-888.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., Engineering, E.Z.A.T.C.o.E.C.i.E., Water, R., 2010. State of the art for genetic algorithms and beyond in water resources planning and management. Journal of Water Resources Planning and Management 136(4) 412-432.

Qin, A.K., Suganthan, P.N., 2005. Self-adaptive differential evolution algorithm for numerical optimization, Evolutionary Computation, 2005. The 2005 IEEE Congress on, pp. 1785-1791 Vol. 1782.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Numerical assessment of metamodelling strategies in computationally intensive optimization. Environmental Modelling & Software 34(0) 67-86.

Reca, J., Martínez, J., 2006. Genetic algorithms for the design of looped irrigation water distribution networks. Water Resour. Res. 42(5) W05416.

Reddy, M.J., Kumar, D.N., 2007. Multiobjective diffential evolution with application to reservoir system optimization. Journal of Computing in Civil Engineering 21(2) 136-146.

Savic, D.A., Walters, G.A., 1997. Genetic algorithms for least-cost design of water distribution networks. Journal of Water Resources Planning and Management 123(2) 67-77.

Storn, R., and Price, K., 1995. Differential Evolution-A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Space, Technical Report, International Computer Science Institute, Berkeley, CA.

Suribabu, C.R., 2010. Differential evolution algorithm for optimal design of water distribution networks. Journal of Hydroinformatics 12(1) 66-82.

Tolson, B.A., Asadzadeh, M., Maier, H.R., Zecchin, A., 2009. Hybrid discrete dynamically dimensioned search (HD-DDS) algorithm for water distribution system design optimization. Water Resour. Res. 45(12) W12416.

Vasan, A., Raju, K., 2007. Application of Differential Evolution for Irrigation Planning: An Indian Case Study. Water Resources Management 21(8) 1393-1407.

Vasan, A., Simonovic, S.P., 2010. Optimization of water distribution network design using differential evolution. Journal of Water Resources Planning and Management 136(2) 279-287.

Wu, Z.Y., Walski, T., 2005. Self-adaptive penalty approach compared with other constraint-handling techniques for pipeline optimization. Journal of Water Resources Planning and Management 131(3) 181-192.

Zaharie, D., 2002. Critical values for control parameters of differential evolution. Proceedings of the 8-th International Mendel Conference on Soft Computation, 62-67.

Zaharie, D., 2009. Influence of crossover on the behavior of differential evolution algorithms. Applied Soft Computing 9(3) 1126-1138.

Zecchin, A.C., Simpson, A.R., Maier, H.R., Marchi, A., Nixon, J.B., 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., Simpson, A.R., Zecchin, A.C., 2011a. A combined NLP-differential evolution algorithm approach for the optimization of looped water distribution systems. Water Resour. Res. 47(8) W08531.

Zheng, F., Simpson, A.R., Zecchin, A.C., 2011b. Parametric analysis of differential evolution algorithm applied to water distribution system optimization 11th International Conference on Computing and Control for the Water Industry (CCWI 2011): Exeter, UK.

Zheng, F., Zecchin, A., Simpson, A., 2013. Self-Adaptive Differential Evolution Algorithm Applied to Water Distribution System Optimization. Journal of Computing in Civil Engineering 27(2) 148-158.

Zheng, F., Zecchin, A., 2014. An efficient decomposition and dual-stage multi-objective optimization method for water distribution systems with multiple supply sources. Environmental Modelling & Software 55(0) 143-155.

Figure captions

Figure 1. Search diversity normalized variance measures for the DEA applied to the HP_{34}. Results are averaged from 20 runs with different random number seeds. Black, red and blue lines represent $Var_V(G)$, $Var_U(G)$ and $Var_X(G)$ respectively.

Figure 2. Surface plots of the percentage differences (%) between the theoretical population variance ratios (Equations 15 and 16) and the experimental results for the three case studies using various *F* and *CR* values. Results are averaged from 20 runs with different random number seeds for each parameter set.

Figure 3. The population variance variations versus generations for the three case studies. Note that the values of F_c (black lines) are derived based on Equation (18). These results are obtained without selection pressure (i.e., all trial solutions are accepted).Results are averaged from 20 runs with different random number seeds for each parameter set.

Figure 4: Behavioral metrics for DEAs with different values of *F* applied to the HP_{34} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are N =100, *CR*=0.5, and *F*=0.1 (black), 0.3(red), 0.5 (blue), 0.7 (green), 0.9 (orange). Results are averaged from 20 runs with different random number seeds.

Figure 5: Behavioral metrics for DEAs with different values of *F* applied to the ZJN_{164} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are N =300, *CR*=0.5, and *F*=0.1 (black), 0.3(red), 0.5

(blue), 0.7 (green), 0.9 (orange). Results are averaged from 20 runs with different random number seeds.

Figure 6: Behavioral metrics for DEAs with different values of *F* applied to the BN_{454} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are *N* =500, *CR*=0.5, and *F*=0.1 (black), 0.3(red), 0.5 (blue), 0.7 (green), 0.9 (orange). Results are averaged from 20 runs with different random number seeds.

Figure 7: Behavioral metrics for DEAs with different values of *CR* applied to the HP_{34} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are *N* =100, *F*=0.7, and *CR* = 0.2 (black), 0.4 (red), 0.6 (blue), 0.8 (green). Results are averaged from 20 runs with different random number seeds.

Figure 8: Behavioral metrics for DEAs with different values of *CR* applied to the ZJN_{164} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are *N* =300, *F*=0.3, *CR* = 0.2 (black), 0.4 (red), 0.6 (blue), 0.8 (green). Results are averaged from 20 runs with different random number seeds.

Figure 9: Behavioral metrics for DEAs with different values of *CR* applied to the BN_{454} case study: (a) $f_{\min}(G)$; (b) $d_{\min}(G)$; (c) $d_{mean}(G)$; (d) $NC(G)$; (e) *PF*%, and; (f) *PI*%. Parameter values are *N* =500, *F*=0.3, and *CR*= 0.2 (black), 0.4 (red), 0.6 (blue), 0.8 (green). Results are averaged from 20 runs with different random number seeds.

Figure 10: Behavioral metric results for DEAs applied to the three case studies. Parameter values are *F*=0.1, and *CR* = 0.1 (black), 0.5 (red), and 0.9 (blue). Results are averaged from 20 runs with different random number seeds. Note that the DEA with *CR*=0.9 and *F*=0.1 was unable to find the feasible solutions for the HP_{34} case study (i.e., the blue line is missing)

Figure 11: Behavioral metric results for DEAs applied to the three case studies. Parameter values are *F*=0.9, and *CR* = 0.1 (black), 0.5 (red), and 0.9 (blue). Results are averaged from 20 runs with different random number seeds.

Table 1 Case studies and the DEA parameter values.

Case study	No. of decision variables	The size of the total search space	Cost of the current best known solution (million)	DEA Population size (N)	Maximum allowable generations
HP_{34}	34	2.865×10^{26}	$6.081	100	10000
ZJN_{164}	164	9.226×10^{187}	$7.082	300	10000
BN_{454}	454	1×10^{454}	€ 1.923	500	10000