



Improved Evolutionary Algorithm Optimisation of Water Distribution  
Systems Using Domain Knowledge

by  
Weiwei Bi

Thesis submitted to School of Civil, Environmental & Mining Engineering  
of the University of Adelaide  
in fulfillment of the requirements for  
the degree of  
**Doctor of Philosophy**

Copyright© Weiwei Bi, December 2015.



Improved Evolutionary Algorithm Optimisation of Water Distribution Systems  
Using Domain Knowledge

By:

Weiwei Bi

Supervised by:

Graeme C. Dandy, *B.E. (Hons), MEngSc, Ph.D.*

*Professor, School of Civil, Environmental & Mining Engineering,  
The University of Adelaide*

Holger R. Maier, *B.E. (Hons), Ph.D.*

*Professor, School of Civil, Environmental & Mining Engineering,  
The University of Adelaide*

Thesis submitted in fulfillment of the requirements for the degree of  
**Doctor of Philosophy**

School of Civil, Environmental & Mining Engineering  
Faculty of Engineering, Computer and Mathematical Sciences  
The University of Adelaide  
North Terrace, Adelaide, SA 5005, Australia  
Telephone: +61 8303 6139  
Facsimile: +61 8303 4359  
Web: <http://www.adelaide.edu.au/directory/weiwei.bi>  
Email: [weiwei.bi@adelaide.edu.au](mailto:weiwei.bi@adelaide.edu.au)

Copyright© Weiwei Bi, December, 2015.



# Contents

<b>Contents</b>	<b>i</b>
<b>Abstract</b>	<b>v</b>
<b>Statement of Originality</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Acronyms</b>	<b>xii</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Objectives of research	3
1.2 Outline of the thesis	4
<b>Chapter 2. Journal Paper 1-Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge</b>	<b>7</b>
2.1 Introduction	9
2.2 Proposed prescreened heuristic sampling method for WDS design	12
2.3 Methodology	18
2.3.1 Sampling methods	19
2.3.2 Case studies	21
2.3.3 Genetic algorithm optimization	23
2.4 Computational Experiments	24
2.5 Results and discussion	27
2.5.1 Group 1 (G1) case studies	29
2.5.2 Group 2 (G2) case studies	31
2.5.3 Group 3 (G3) case studies	33
2.6 Summary and conclusions	34
<b>Chapter 3. Journal Paper 2- Impact of starting position and searching mechanism on evolutionary algorithm convergence rate</b>	<b>37</b>

<b>3.1 Introduction</b>	<b>40</b>
<b>3.2 Methodology</b>	<b>43</b>
3.2.1 Initialization approaches	44
3.2.2 Evolutionary algorithms and their parameterization	45
3.2.3 Case studies	46
3.2.4 Performance assessment	47
3.2.5 Performance explanation	48
<b>3.3 Computational Experiments</b>	<b>53</b>
<b>3.4 Results and discussions</b>	<b>54</b>
3.4.1 Impact of the starting positions and searching mechanisms	54
3.4.2 Relationship between observed performance and problem statistics	57
3.4.3 Relationship between observed performance and population diversity	60
3.4.4 Summary	61
<b>3.5 Conclusions</b>	<b>62</b>
<b>Chapter 4. Journal Paper 3- Use of domain knowledge to increase the convergence rate of evolutionary algorithms for optimizing the cost and resilience of water distribution systems</b>	<b>65</b>
<b>4.1 Introduction</b>	<b>67</b>
<b>4.2 The optimization problem</b>	<b>70</b>
<b>4.3 Proposed multi-objective prescreened heuristic sampling method</b>	<b>72</b>
4.3.1 Step 1: Identify initial solutions using domain knowledge related to cost	72
4.3.2 Step 2: Identify an initial front by adjusting the solutions obtained in Step 1 based on domain knowledge related to network resilience	73
4.3.3 Step 3: Generate initial MOEA population by sampling in the vicinity of the initial front identified in Step 2	75
<b>4.4 Methodology</b>	<b>76</b>
4.4.1 Methods for determining initial MOEA population	77
4.4.2 Multiobjective evolutionary algorithms	78
4.4.3 Case studies	79
4.4.4 Run-time performance metrics	79
<b>4.5 Computational experiments</b>	<b>81</b>
<b>4.6 Results and discussion</b>	<b>83</b>

<b>4.7 Summary and conclusions</b>	<b>90</b>
<b>Chapter 5. Conclusions and Recommendations for Future Work</b>	<b>93</b>
<b>5.1 Research Contributions</b>	<b>93</b>
<b>5.2 Research Limitations</b>	<b>95</b>
<b>5.3 Recommendations for Future Work</b>	<b>96</b>
<b>References</b>	<b>99</b>
<b>Appendix</b>	<b>107</b>





## **Abstract**

Water distribution systems (WDSs) are becoming increasingly complex and larger in scale due to the rapid growth of population and fast urbanization. Hence, they require high levels of investment for their construction and maintenance. This motivates the need to optimally design these systems, with the aim being to minimize the investment budget while maintaining high service quality. Over the past 25 years, a number of evolutionary algorithms (EAs) have been developed to achieve optimal design solutions for WDSs, representing a focal point of much research in this area.

One issue that hinders EAs' wide application in industry is their significant demand on computational resources when handling real-world WDSs. In recognition of this, there has been a move from aiming to find the globally optimal solutions to identifying the best possible solutions within constrained computational resources. While many studies have been undertaken to attain this goal, there have been limited efforts that use engineering knowledge to reduce the computational effort. The research undertaken in this thesis is such an attempt, as it aims to efficiently identify near-optimal solutions with the aid of WDS design knowledge.

This thesis presents a domain-knowledge based optimization framework that enables the near-optimal solutions (fronts) of WDS problems to be identified within constrained computing time. The knowledge considered includes (i) the relationship between pipe size and distance to the water source(s); (ii) the impact of flow velocities on optimal solutions; and (iii) the relationship between flow velocities and network resilience.

This thesis consists of an Introduction, three chapters that are based around a series of three journal papers and a set of Conclusions and Recommendations for Further Work.

The first paper introduces a new initialization method to assist genetic algorithms (GAs) to identify near-optimal solutions in a computationally efficient manner. This is attained by incorporating domain knowledge into the generation of the initial population of GAs. The results show that the proposed method performs better than the other three initialization methods considered, both in terms of computational efficiency and the ability to find near-optimal solutions.

The second paper investigates the relative impact of different algorithm initializations and searching mechanisms on the speed with which near-optimal solutions can be identified for large WDS design problems. Results indicate that EA parameterizations, that emphasize exploitation relative to exploration, enable near-optimal solutions to be identified earlier in the search, which is due to the “big bowl” shape of the fitness function for all of the WDS problems considered. Using initial solutions that are informed using domain knowledge can further increase the speed with which near-optimal solutions can be identified.

The third publication extends the single-objective method in the first paper to a two-objective problem. The objectives considered are the minimization of cost and maximization of network resilience. The performance of the two-objective initialization approach is compared with that of randomly initializing the population of multi-objective EAs applied to range of WDS design problems. The results indicate that there are considerable benefits in using the proposed initialization method in terms of being able to identify near-optimal fronts more rapidly.

Although all of the results obtained in this research have shown that the proposed method is effective for improving the efficiency of EAs in finding near-optimal solutions, only gravity fed water distribution systems with a single loading case were considered as case studies. One important area for future research is the extension of the proposed method to more complex WDSs which may include tanks, pumps and valves.

## **Statement of Originality**

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution. To the best of my knowledge and belief it contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying, subject to the provision of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder (s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signature

Date

---

## **Acknowledgements**

I would take this opportunity to thank my academic supervisors Professor Graeme C. Dandy and Professor Holger R. Maier for their continuous support and guidance. I appreciate the research freedom I have enjoyed under their supervision, and thank them deeply for always being available to provide constructive suggestions and discuss any issues with great patience.

I would like to acknowledge the support of my family: to my husband Feifei Zheng for his love, patience and continuous support; to my parents Mr. Keming Bi and Mrs. Yuqin Guo for their help and encouragements; and to my baby boy Zhuohao Zheng for the happiness he brings to me over the journey of my PhD candidature.

I acknowledge the research scholarship provided by the University of Adelaide, as well as the assistance from the colleagues and staff in the School of Civil, Environmental and Mining Engineering, The University of Adelaide.

## List of Figures

Figure 2.1 WDS used to illustrate the result of network division of the PHSM (The red dot lines represent the distance boundary used to assign diameters) .....	14
Figure 2.2 Flowchart of the algorithm for adjusting pipe diameters based on flow velocity .....	15
Figure 2.3 Flowchart of the assessment process .....	19
Figure 2.4 Results of the GAs with the four sampling methods applied to case studies in Group 1 (G1 in Table 2.3).....	30
Figure 2.5 Results of the GAs with the four sampling methods applied to case studies in Group 2 (G2 in Table 2.3).....	32
Figure 2.6 Results of the GAs with the four sampling methods applied to case studies in Group 3 (G3 in Table 2.3).....	34
Figure 3.1 Flowchart of the assessment process, where $N$ is the population size, $L$ is the number of decision variables, $P_c$ and $CR$ are the crossover probabilities for the genetic algorithm (GA) and differential evolution (DE), respectively, and $P_m$ and $F$ are their mutation probabilities. The subscript of each case study indicates the number of decision variables. ....	44
Figure 3.2 Illustration of the spatial correlation statistics using a hypothetical case. ....	49
Figure 3.3 Illustration of the dispersion metric using a hypothetical case. Dots represent solutions in the two-dimensional domain and the red dots indicate the 30 best solutions across different sample sizes. The blue “+” indicate the local optima and the dashed circles show the promising regions. ....	51
Figure 3.4 Results for GAs for which initial solutions are obtained using the RS method (black lines), and the PHSM (red lines). Different line types represent different mutation probabilities $P_m$ . Left panel: Deviation of the mean cost from the best known solution (DMO%), with the horizontal grey lines showing 5% deviation. Right panel: Standardized average population diversity SPD (%) with the horizontal grey lines indicating complete convergence. ....	55
Figure 3.5 Results for DEs for which initial solutions are obtained using the RS method (black lines), and the PHSM (red lines). Different line types represent different mutation weighting factors $F$ . Left panel: Deviation of the mean cost to the best known solution (DMO%), with the horizontal grey lines showing 5% deviation to the best-known solution. Right panel: Standardized average population diversity SPD (%) with the horizontal grey lines indicating complete convergence. ....	56
Figure 3.6. (a) Fitness function statistics (spatial correlation) of the four WDS problems considered; (b) Change in normalized pairwise distance of the top $m=100$ solutions with increase in sample size $N$ . The grey line represents the values for the benchmarking NYTP.....	58

Figure 3.7. Stylized representation of the cross-section of the fitness function of the WDS design problems considered. ....	58
Figure 3.8. Results of the dynamic correlations between DMO% and SPD%. The correlation at generation $G$ was estimated using DMO%[1: $G$ ] and SPD%[1: $G$ ]. The results shown are for random initialization. Similar results are obtained for initialization with the PHSM. ....	61
Figure 4.1 Flowchart of the proposed methodology for adjusting the pipe diameters obtained from step 1 based on flow velocity in order to identify good initial solutions in relation to both cost and network resilience .....	74
Figure 4.2 Distribution of samples for different values of $a$ .....	76
Figure 4.3 Flowchart of the assessment process. The subscript of each case study indicates the number of decision variables. ....	77
Figure 4.4 Approximate fronts of the proposed MOPHSM (red '+') and the random initialization approach (black circles) obtained using NSGA-II. The grey triangles are the best-known fronts.....	85
Figure 4.5 Approximate fronts of the proposed MOPHSM (red '+') and the random initialization approach (black circles) obtained using Borg. The grey triangles are the best-known fronts.....	86
Figure 4.6 Run-time performance metrics of the proposed MOPHSM (red dashed line) and the random initialization approach (black solid lines) for NSGA-II. The horizontal grey line in the left panel indicates 95% of the best-known front hypervolume (near-optimal fronts).....	88
Figure 4.7 Run-time performance metrics of the proposed MOPHSM (red dashed line) and the random initialization approach (black solid lines) for Borg. The horizontal grey line in the left panel indicates 95% of the best-known front hypervolume (near-optimal fronts).....	89

## List of Tables

Table 1.1 Publication information.....	4
Table 2.1 An example to illustrate the application of Step 3 of the PHSM .....	18
Table 2.2 Details of the seven case studies .....	22
Table 2.3 Parameters values of GAs for each case study.....	25
Table 2.4 Computational overhead analysis for the proposed sampling method (PHSM) .....	27
Table 2.5 Cost of the best solution found by each sampling method for each case study .....	28
Table 4.1 MOEA parameters used for each case study.....	82

## List of Acronyms

ABS	Average of the best solutions
BBS	Best of the best solutions
DE	Differential evolution
DMO	Deviation of the mean cost relative to the best known solution
EA	Evolutionary algorithms
GA	Genetic algorithm
KLSM	Kang and Lansey's sampling method
LHS	Latin hypercube sampling
MOEA	Multi-objective evolutionary algorithms
MOPHSM	Multi-objective prescreened heuristic sampling method
NYTP	New York tunnel problem
PCX	Patent-centric crossover
PHSM	Prescreened heuristic sampling method
RS	Random sampling
SBX	Simulated binary crossover
SPD	Standardized average population diversity
SPX	Simplex crossover
UM	Uniform mutation
UNDX	Unimodal normal distribution crossover
WDS	Water distribution systems