

# Embracing Analytics in the Drinking Water Industry

---



# Embracing Analytics in the Drinking Water Industry

---

Edited by  
Juneseok Lee and Jonathan Keck



**Published by**

**IWA Publishing**  
**Unit 104–105, Export Building**  
**1 Clove Crescent**  
**London E14 2BA, UK**  
Telephone: +44 (0)20 7654 5500  
Fax: +44 (0)20 7654 5555  
Email: [publications@iwap.co.uk](mailto:publications@iwap.co.uk)  
Web: [www.iwapublishing.com](http://www.iwapublishing.com)

First published 2022  
© 2022 IWA Publishing

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the UK Copyright, Designs and Patents Act (1998), no part of this publication may be reproduced, stored or transmitted in any form or by any means, without the prior permission in writing of the publisher, or, in the case of photographic reproduction, in accordance with the terms of licenses issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of licenses issued by the appropriate reproduction rights organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to IWA Publishing at the address printed above.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for errors or omissions that may be made.

**Disclaimer**

The information provided and the opinions given in this publication are not necessarily those of IWA and should not be acted upon without independent consideration and professional advice. IWA and the Editors and Authors will not accept responsibility for any loss or damage suffered by any person acting or refraining from acting upon any material contained in this publication.

*British Library Cataloguing in Publication Data*

A CIP catalogue record for this book is available from the British Library

ISBN: 9781789062373 (Paperback)

ISBN: 9781789062380 (eBook)

ISBN: 9781789062397 (ePub)

This eBook was made Open Access in June 2022

© 2022 The Editors

This is an Open Access eBook distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). This does not affect the rights licensed or assigned from any third party in this book.



# Contents

---

<b>Preface</b> .....	xv
----------------------	----

## **Chapter 1**

<b><i>Introduction</i></b> .....	<b>1</b>
----------------------------------	----------

*Jonathan Keck and Juneseok Lee*

1.1	What is Analytics? .....	1
1.2	How can Analytics Help the Water Industry? .....	2
1.3	Effective Utility Management .....	2
1.3.1	Foundational element #1 – attributes of effectively managed water sector utilities .....	3
1.3.2	Foundational element #2 – keys to management success .....	3
1.3.3	Foundational element #3 – water utility measures .....	5
1.3.4	Foundational element #4 – water utility management resources .....	5
1.4	Effective Utility Management (EUM) and Water Analytics .....	6
1.4.1	Supply and demand management .....	6
1.4.2	Enterprise asset management .....	6
1.4.3	Distribution system modeling .....	7
1.4.4	Long-range planning .....	7
1.4.5	Systems optimization .....	8
1.5	Recommendations .....	8
1.5.1	Analytics leadership .....	8
1.5.2	Cultural importance .....	10
1.5.3	Capacity planning .....	10
1.5.4	Systems and structure .....	10
1.5.5	Tools and technology .....	10
1.5.6	Professional development and collaborative research .....	11
1.6	A Clear Future for Analytics .....	11
1.7	Roadmap of the Book .....	11
1.7.1	Planning section .....	11
1.7.2	Operations section .....	13
1.7.3	Management section .....	15
	References .....	17

**Part I: Planning** .....19**Chapter 2*****Water demand analysis | regression*** ..... 21*Stephanie A. Tanverakul*

Learning Objectives.....	21
2.1 Introduction.....	21
2.2 Principles of Regression.....	22
2.2.1 What is regression?.....	22
2.2.2 Basic regression equation – water demand and lot size example.....	22
2.2.3 OLS assumptions.....	25
2.2.4 Panel data regression.....	29
2.2.5 Multiple regression.....	30
2.3 Model Specification.....	32
2.3.1 Water use relationships.....	33
2.3.2 Data exploration.....	34
2.3.3 Level of aggregation.....	37
2.3.4 Data range and variation.....	37
2.3.5 Misspecification.....	37
2.4 Estimating Parameters.....	38
2.4.1 Panel regression – pooled, fixed effects, and random effects.....	38
2.4.2 Estimation example walk-through problem in R.....	39
2.5 Interpretation.....	45
2.5.1 Regression example – forecasting.....	45
2.5.2 Regression example – metering impacts.....	45
2.5.3 Presentation of results.....	47
2.6 Conclusion.....	48
References.....	48

**Chapter 3*****Water demand forecasting – machine learning*** ..... 51*Maria Xenochristou*

Learning Objectives.....	51
3.1 Introduction.....	51
3.2 Data.....	52
3.2.1 Data collection.....	52
3.2.2 Data cleaning.....	53
3.2.3 Feature engineering.....	54
3.2.4 Feature selection.....	54
3.2.5 Data transformations.....	56
3.3 Model Building.....	56
3.3.1 Model selection.....	56
3.3.2 Hyperparameter optimization.....	56
3.3.3 Training, validation, and testing.....	57
3.4 Model Evaluation.....	57
3.4.1 Model accuracy.....	57
3.4.2 Model interpretability.....	57

3.5	Model Deployment	58
3.6	Tools and Software	58
3.6.1	Prerequisites	58
3.6.2	Useful tools, packages, and APIs	59
3.7	Practical Examples	59
3.7.1	Installation	59
3.7.2	Example 1: A simple model for demand forecasting	60
3.7.3	Installing and loading R packages	60
3.7.4	Get and preprocess the data	62
3.7.5	Model training and testing	63
3.7.6	Questions	67
3.7.7	PDP and ICE plots	71
3.8	Conclusion	72
	References	72

## Chapter 4

<b><i>Water demand forecasting   time series data</i></b>	<b>75</b>
<i>Jarai Sanneh, A. Di Mauro, S. Venticinquè, G. F. Santonastaso, A. Di Nardo, Yi Wang and Juneseok Lee</i>	

Learning Objectives	75
4.1 Introduction	75
4.2 Time Series Data Analysis	76
4.2.1 ARIMA model	76
4.2.2 SARIMA model	77
4.2.3 Creating ARIMA/SARIMA models using econometric toolbox	77
4.2.4 Forecasting	87
4.2.5 Limitations	89
4.3 Machine Learning Time Series	90
4.3.1 Machine learning	90
4.3.2 Practice problems	93
4.4 Deep Learning Time Series	97
4.4.1 Deep learning models	97
4.4.2 Practice problems	99
4.5 Other Popular ML Techniques	102
4.5.1 Ensemble learning	102
References	120

## Chapter 5

<b><i>Use of cost-benefit analysis (CBA) in water infrastructure</i></b>	<b>123</b>
<i>Anita M. Chaudhry</i>	

Learning Objectives	123
5.1 Introduction	123
5.2 Contribution of CBA to Water Policymaking	123
5.2.1 Imperatives of water scarcity: demand management or supply enhancement?	123
5.2.2 CBA as a decision-making tool	125
5.2.3 Policy background	125
5.3 CBA Methods	126
5.3.1 Building a spreadsheet of the CBA model	126

5.3.2	Identifying and measuring the benefits . . . . .	126
5.3.3	Identifying and measuring the costs . . . . .	126
5.3.4	Time horizon and discount and interest rates . . . . .	128
5.4	CBA in Practice . . . . .	128
5.4.1	CBA of reservoir construction . . . . .	128
5.4.2	CBA of rainwater harvesting systems (RWHS) . . . . .	130
5.5	Conclusion . . . . .	131
	References . . . . .	132

## **Part II: Operations . . . . .133**

### **Chapter 6**

#### ***Water quality modeling and analysis . . . . . 135***

*Maria A. Palmegiani and Juneseok Lee*

	Learning Objectives . . . . .	135
6.1	Introduction . . . . .	135
6.2	EPANET and EPANET-MSX Software . . . . .	136
6.3	Creating an EPANET Network File . . . . .	140
6.4	Modeling Water Age and Single-Species Water Quality on EPANET . . . . .	145
6.5	Modeling Multiple Species Using EPANET-MSX . . . . .	147
6.6	Running EPANET-MSX Software and Calibrating Results to Sampled Data . . . . .	150
6.7	Model Statistical Verification . . . . .	153
6.8	Conclusion . . . . .	155
	References . . . . .	156

### **Chapter 7**

#### ***Calibration and uncertainty analysis of hydraulic models . . . . . 159***

*Adell Moradi Sabzkouhi, Juneseok Lee and Jonathan Keck*

	Learning Objectives . . . . .	159
7.1	Introduction . . . . .	159
7.2	Uncertain Parameters in Pipe Network Analysis . . . . .	160
7.2.1	Pipe roughness coefficients . . . . .	160
7.2.2	Nodal demands . . . . .	161
7.2.3	Pipe diameters . . . . .	161
7.2.4	Leakage parameters . . . . .	161
7.2.5	Boundary conditions, tanks, valves and pump characteristics . . . . .	161
7.3	Review on Calibration Steps . . . . .	162
7.3.1	Identifying the intended use of the model . . . . .	162
7.3.2	Determining initial estimation of model parameters . . . . .	162
7.3.3	Collecting calibration data . . . . .	162
7.3.4	Evaluating model results . . . . .	162
7.3.5	Performing macro-level calibration . . . . .	163
7.3.6	Performing sensitivity analysis . . . . .	163
7.3.7	Micro-calibration . . . . .	163
7.4	Automatic Calibration . . . . .	163
7.4.1	Conceptual framework . . . . .	163



7.4.2	Dynamic link for the simulation-optimization model . . . . .	165
7.4.3	Mathematical statement of the problem . . . . .	165
7.5	Example 7.1: Calibration of Anytown Modified Network . . . . .	167
7.5.1	Optimization model: genetic algorithm . . . . .	169
7.5.2	Optimization model setting . . . . .	170
7.5.3	Model execution and results . . . . .	172
7.6	Parameter Uncertainty Analysis in Pipe Network Modeling . . . . .	173
7.6.1	Does UA search for the most pessimistic combination of parameters? . . . . .	174
7.6.2	Approaches for parameter UA . . . . .	176
7.6.3	Interval analysis (IA) for parameter UA . . . . .	176
7.7	Example 7.2: Interval Analysis for the ATM Network . . . . .	178
7.7.1	Producing the impact table . . . . .	179
7.7.2	Calculating the extreme pressure heads . . . . .	179
7.8	Conclusion . . . . .	179
	References . . . . .	184

## Chapter 8

### ***Optimal pump operation*** . . . . . **187**

*Adell Moradi Sabzkouhi, Juneseok Lee and Jonathan Keck*

	Learning Objectives . . . . .	187
8.1	Introduction . . . . .	187
8.2	A Brief Review on Pump Performance . . . . .	188
8.2.1	Head-flow characteristics . . . . .	188
8.2.2	Power-flow characteristics . . . . .	189
8.2.3	Efficiency-flow characteristics . . . . .	189
8.2.4	NPSH-flow characteristics . . . . .	190
8.2.5	System's curve and pump duty-point . . . . .	190
8.2.6	Affinity laws for rotational speed . . . . .	191
8.3	Main Considerations for Optimal Pump Operation . . . . .	193
8.3.1	<i>BEP</i> and minimum efficiency . . . . .	193
8.3.2	Pump discharge range . . . . .	193
8.3.3	Pump speed . . . . .	193
8.3.4	Pump switches and daily working hours . . . . .	193
8.4	Pump Operation Control . . . . .	194
8.4.1	Change in system curve . . . . .	194
8.4.2	Change in characteristic curve . . . . .	194
8.5	Optimal Pump Scheduling . . . . .	194
8.5.1	Simulation-optimization approach . . . . .	194
8.5.2	Optimization objectives . . . . .	196
8.5.3	Pump scheduling approaches: CSP vs. VSP . . . . .	196
8.5.4	Example 8.2 – CSP approach . . . . .	197
8.5.5	Example 8.3 – VSP approach . . . . .	202
8.6	VSP Scheduling; EWQMS Approach . . . . .	205
8.6.1	EWQMS for ATM network – A single objective optimization approach . . . . .	206
8.6.2	EWQMS for ATM network – A multi-objective optimization approach . . . . .	210
8.7	Conclusion . . . . .	211
	References . . . . .	213

**Chapter 9*****Hydraulic transients in pipe systems* . . . . . 215***Juneseok Lee, Lu Xing and Lina Sela*

Learning Objectives . . . . .	215
9.1 Introduction . . . . .	215
9.2 Numerical Method Considering Initial and Boundary Conditions . . . . .	216
9.2.1 Reservoir . . . . .	218
9.2.2 Junctions . . . . .	218
9.2.3 Discretization error . . . . .	220
9.2.4 Truncation error . . . . .	220
9.2.5 Consistency . . . . .	220
9.2.6 Convergence . . . . .	220
9.2.7 Stability . . . . .	220
9.2.8 CFL (Courant Friedrich Lewy) stability condition . . . . .	220
9.2.9 Example . . . . .	220
9.3 Other Phenomenon of Interests   Cavitation and Column Separation . . . . .	221
9.3.1 Discrete vapor cavity model (DVCM) . . . . .	225
9.3.2 Short term pressure peaks following cavity collapse . . . . .	226
9.4 Transient Simulations in Water Distribution Networks: TSNet . . . . .	226
9.4.1 TSNet . . . . .	226
9.4.2 Use of Python, Spyder and Anaconda . . . . .	226
9.4.3 Example application . . . . .	227
9.4.4 Create and set up a transient model . . . . .	229
9.4.5 Valve closure . . . . .	229
9.4.6 Pump shutdown . . . . .	231
9.4.7 Pipe burst . . . . .	233
9.4.8 Other applications . . . . .	233
References . . . . .	234

**Chapter 10*****Innovative methods for optimal design of water network partitioning* . . . . . 237***Armando Di Nardo and Giovanni Francesco Santonastaso*

Learning Objectives . . . . .	237
10.1 Introduction . . . . .	237
10.2 Advantages of WNP . . . . .	239
10.3 Water Balance . . . . .	239
10.4 Water Pressure Management . . . . .	240
10.5 Water Contamination Protection . . . . .	241
10.5.1 Clustering and dividing . . . . .	242
10.5.2 Innovative methods for optimal WNP design . . . . .	245
10.5.3 WNP with SWANP® software . . . . .	247
10.5.4 Python code to design an optimal WNP . . . . .	247
References . . . . .	253

**Chapter 11*****Reliability analysis using optimization* . . . . . 255***Sangamreddi Chandramouli*

Learning Objectives . . . . .	255
11.1 Introduction . . . . .	255
11.1.1 Brief history of pipe networks . . . . .	255
11.1.2 Development of water supply engineering . . . . .	255
11.1.3 Brief description of optimization techniques . . . . .	256
11.1.4 Brief description of reliability concept . . . . .	257
11.2 Concept of Fuzzy Set Theory . . . . .	257
11.2.1 Brief description of fuzzy set theory . . . . .	257
11.2.2 Membership functions . . . . .	258
11.2.3 Types of fuzzy sets and fuzzy functions . . . . .	259
11.3 Reliability Analysis of Water Supply Pipe Networks . . . . .	260
11.3.1 Definition of reliability index . . . . .	260
11.3.2 Ranges of residual pressures and satisfaction levels based on fuzzy logic . . . . .	261
11.4 Reliability Based Optimization of Pipe Networks . . . . .	264
11.4.1 Description of objective function . . . . .	264
11.4.2 Description of constraints . . . . .	264
11.4.3 Incorporation of reliability index into optimization . . . . .	265
11.5 Description of Epanet Toolkit Functions . . . . .	265
11.5.1 Introduction . . . . .	265
11.5.2 A typical usage of the toolkit functions to analyse a distribution system . . . . .	266
11.5.3 Input file format with examples . . . . .	266
11.5.4 Linking of EPANET tool kit functions in MATLAB (input, output, opening and closing) . . . . .	266
11.6 Process of Optimization Using Genetic Algorithms in MATLAB Using GA Tool Kit Functions . . . . .	267
11.6.1 Genetic algorithms . . . . .	267
11.6.2 Description of GA tool kit functions . . . . .	267
11.7 Implementation of Reliability Based Optimization for Pipe Network Design Using MATLAB and EPANET Toolkit Functions . . . . .	268
11.7.1 Development of coding in MATLAB – step by step process . . . . .	268
11.8 Study on a Standard Benchmark Network . . . . .	269
11.8.1 Description of network . . . . .	269
11.8.2 Input file preparation for the standard benchmark network . . . . .	271
11.8.3 Developing the code in matlab . . . . .	272
11.8.4 Screenshots for the program implemented in matlab . . . . .	275
11.8.5 Analysis of the results . . . . .	278
11.8.6 Comparison of reliability Index of a simple pipe network based on binary logic and fuzzy logic . . . . .	281
11.8.7 Summary of results and conclusions . . . . .	284
11.9 Limitations of the Present Study and Scope for Future Research . . . . .	285
11.10 Practice Problems . . . . .	286
11.10.1 Practice problem-1 . . . . .	286
11.10.2 Practice problem 2 . . . . .	289
References . . . . .	292

**Chapter 12*****Water network tool for resilience* . . . . . 295***Lucinda-Joi Chu-Ketterer, Jonathan Burkhardt, Katherine Klise  
and Terranna Haxton*

Learning Objectives . . . . .	295
12.1 Introduction . . . . .	295
12.2 Resilience of Drinking Water Systems . . . . .	296
12.2.1 Disasters . . . . .	296
12.2.2 Measuring resilience . . . . .	297
12.2.3 Challenges with modeling system resilience. . . . .	297
12.3 Water Network Tool for Resilience . . . . .	298
12.3.1 Overview. . . . .	298
12.3.2 Installation and requirements . . . . .	298
12.3.3 Units . . . . .	299
12.3.4 Available solvers. . . . .	299
12.3.5 Examples and demos. . . . .	300
12.4 Disaster Scenarios . . . . .	302
12.4.1 Pipe breaks . . . . .	302
12.4.2 Segment isolation. . . . .	302
12.4.3 Earthquakes . . . . .	303
12.4.4 Fires. . . . .	303
12.4.5 Loss of source water . . . . .	303
12.4.6 Power outage . . . . .	303
12.4.7 Other scenarios . . . . .	304
12.5 WNTR Resilience Metrics . . . . .	304
12.5.1 Water service availability . . . . .	304
12.5.2 Todini index . . . . .	304
12.5.3 Modified resilience index . . . . .	304
12.5.4 Additional metrics . . . . .	305
12.6 Tutorials . . . . .	305
12.6.1 Pipe break. . . . .	305
12.6.2 Segment isolation. . . . .	309
12.6.3 Fire flow . . . . .	314
12.6.4 Earthquake . . . . .	317
12.7 Conclusions . . . . .	322
12.8 Disclaimer. . . . .	322
References. . . . .	322

**Part III: Management. . . . . 325****Chapter 13*****Optimal replacement time of water mains* . . . . . 327***Juneseok Lee*

Learning Objectives . . . . .	327
13.1 Introduction . . . . .	327
13.1.1 Performance management . . . . .	327
13.1.2 Failure mode analysis . . . . .	328
13.1.3 Operations and maintenance . . . . .	329
13.1.4 Risk analysis. . . . .	329

13.1.5	Capital specification	329
13.1.6	Prioritization	329
13.2	Optimal Replacement	329
13.3	Practical Examples	331
13.3.1	Example 1	331
13.3.2	Example 2	331
13.3.3	Solution	331
13.4	Conclusions	334
	References	334

## Chapter 14

<b><i>Water mains replacement decision using GIS analytics</i></b>	<b>335</b>
<i>Diego Martinez Garcia</i>	

Learning Objectives	335
14.1 introduction	335
14.2 Data	336
14.2.1 Water main failures	336
14.2.2 Water mains	337
14.2.3 Base map	337
14.3 Multilinear Regression Model	337
14.3.1 Description of linear model	337
14.3.2 Age based linear model	338
14.3.3 R linear regression model	340
14.4 Hot Spot Analysis of Water Main Failures	342
14.4.1 Hot spot analysis tool	342
14.4.2 Interpretation of results	344
14.5 Spatiotemporal Assessment of Water Main Failures	345
14.5.1 Introduction	345
14.5.2 Emerging hot spot analysis tool	346
14.5.3 Interpretation of results	348
14.6 Spatiotemporal Based Business Risk Exposure Analysis	352
14.6.1 Likelihood of failure	352
14.6.2 Consequence of failure	352
14.6.3 BRE matrix integrating LOF/COF	354
14.7 Practice Problems	355
14.7.1 R linear regression model	355
14.7.2 Create grid to extract aggregated data	359
14.7.3 Hot spot analysis of water main failure	365
14.7.4 Create space time cube of water main failures	366
14.7.5 Emerging Hot spot analysis of water main failures	369
14.7.6 Spatiotemporal based business risk exposure	373
References	378

## Chapter 15

<b><i>Decision Analysis</i></b>	<b>381</b>
<i>Eftila Tanellari and Juneseok Lee</i>	

Learning Objectives	381
15.1 Nonmarket Valuation	381
15.2 Contingent Valuation	382
15.2.1 Analysis of CV data	383

15.3	Conjoint Analysis.....	386
15.3.1	Analysis of CA data.....	387
15.3.2	Analytical Hierarchical Process (AHP).....	391
15.4	Conclusions.....	396
	References.....	396
<b>Chapter 16</b>		
	<b><i>Non-revenue water, what are their determinants?</i></b> .....	<b>399</b>
	<i>Gamze Güngör-Demirci and Juneseok Lee</i>	
	Learning Objectives.....	399
16.1	Introduction.....	399
16.2	Regression Model Specification.....	400
16.2.1	Example.....	401
16.2.2	Water utility example.....	401
16.2.3	Interpretations.....	405
16.3	Conclusion.....	406
	References.....	406
<b>Chapter 17</b>		
	<b><i>Water utility performance measurements using data envelopment analysis</i></b> .....	<b>409</b>
	<i>Gamze Güngör-Demirci and Juneseok Lee</i>	
	Learning Objectives.....	409
17.1	Introduction.....	409
17.2	Methods.....	409
17.2.1	Efficiency calculation by DEA.....	409
17.2.2	Input and output variables.....	410
17.2.3	Bias correction for efficiency scores by bootstrapping.....	410
17.2.4	Exogenous variables.....	410
17.3	Examples.....	411
17.3.1	Results and interpretations.....	414
17.4	Conclusions.....	415
	References.....	415
	<b>Index</b> .....	<b>417</b>

# Preface

---

The purpose of this book is to introduce ‘*analytics*’ to practicing water engineers so that they can incorporate the covered subjects, approaches, and detailed techniques within their daily operations, management, and decision-making processes. Also, undergraduate students as well as early graduate students who are in water and environmental systems concentration areas will be exposed to established analytical techniques, along with many methods that are currently considered to be new or emerging and maturing.

This book covers a broad spectrum of water industry analytics topics in an easy-to-follow manner. The overall background and context are motivated by (and directly drawn from) actual water utility projects that we have worked on over numerous recent years. Many chapter authors are the editor’s previous students and collaborators that have worked together. We strongly believe that the water industry should embrace and integrate data-driven fundamentals and methods into their daily operations and decision-making process(es) in an effort to replace more traditional and established ‘rule-of-thumb’ and (arguably) weaker heuristic approaches – and an analytics viewpoint, approach, and culture is key to this industry transformation. Analytics can support numerous aspects of water utility planning, operations, and management, and the organization of this book naturally follows pace by including three principal sections – planning, operations, and management.

Water is essential for human well-being and survival, and throughout the water industry, it is becoming increasingly imperative that in-house analytics capability and championship be developed and integrated to address the current and transitional challenges we face. Again, one of our main contentions is that analytics will contribute substantially to future efforts aimed at providing innovative solutions that make the water industry more sustainable and resilient. We sincerely hope that this book provides a range of learning experiences that help to share and expand this view.

**Juneseok Lee, Editor**  
Manhattan College  
**Jonathan Keck, Editor**  
Water First, LLC

