Department of Electrical and Computer Engineering

Fault Diagnosis and Security Monitoring in Water Distribution Systems

Demetrios G. Eliades

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the University of Cyprus

March, 2011

APPROVAL PAGE

Demetrios G. Eliades

Fault Diagnosis and Security Monitoring

in Water Distribution Systems

The present Doctorate Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering, and was approved on March 24, 2011 by the members of the Examination Committee.

Committee Chair _____
Dr. Christos Panayiotou

Research Supervisor _____
Dr. Marios M. Polycarpou

Committee Member _____
Dr. Elias Kyriakides

Committee Member _____
Dr. Symeon Christodoulou

Committee Member _____
Dr. Avi Ostfeld

Abstract

The efficient and sustainable management of water resources is a key challenge that is becoming more essential year after year. Water resources management involves the collection of water from various sources, the disinfection of water at treatment plants and the delivery of clean water to consumers through water distribution systems. Water distribution systems, in particular, have a significant role in sustaining vital societal functions; however, when a system fault occurs, such as a water contamination intrusion or a pipe break, these societal functions may be negatively affected. In previous years, various aspects of the security monitoring problem in water distribution systems have been examined; in addition, robust fault diagnosis algorithms have been developed within a system-theoretic framework. This thesis presents a formulation of a system-theoretic framework suitable for fault diagnosis and security monitoring in water distribution systems. First, a formulation of the monitoring and control problem of water distribution networks is presented, in a framework suitable for sensor placement and fault diagnosis. Based on the developed framework, the sensor placement problem is examined, to find suitable locations in a water distribution network where on-line quality sensors ought to be installed, for minimizing the risk of a severe damage on population, in case that a contaminant enters the network and is distributed with flow. Furthermore, the manual quality sampling scheduling problem is examined, to find where and when to take water samples for quality monitoring. Next, a disinfectant concentration regulation algorithm for water distribution networks is designed, using adaptive approximation to learn water demands. The detection of hydraulic leakage faults in District Metered Areas (DMA) is examined, by using a fault detection method based on learning the periodic consumption dynamics. Finally, the impact of a contamination detected in a water distribution system is evaluated, and its source area is isolated, using a methodology based on decision tree induction. The effectiveness of the proposed methodologies is illustrated with simulations using water distribution system models and historical hydraulic data.

Περίληψη

Η αποτελεσματική και αειφόρος διαχείριση των υδάτινων πόρων αποτελεί μια βασική πρόκληση η οποία καθίσταται, χρόνο με το χρόνο, ολοένα και πιο ουσιαστική. Η διαχείριση υδάτινων πόρων περιλαμβάνει τη συγκέντρωση νερού από διάφορες πηγές, την απολύμανση του νερού σε εργοστάσια επεξεργασίας και την παράδοση του καθαρού νερού στους καταναλωτές, μέσω των συστημάτων διανομής νερού. Τα συστήματα διανομής νερού διαδραματίζουν ένα ζωτικής σημασίας ρόλο στην διατήρηση υψηλού βιοτικού επιπέδου. Εντούτοις, όταν ένα σφάλμα εμφανιστεί στο σύστημα, όπως μόλυνση του νερού ή σπάσιμο κάποιου αγωγού, αυτό μπορεί να επιφέρει δυσμενείς επιπτώσεις στη λειτουργία της κοινωνίας. Στα προηγούμενα χρόνια έχουν εξεταστεί ερευνητικά διάφορες πτυχές του προβλήματος ελέγχου ασφάλειας στα συστήματα διανομής νερού, και επιπρόσθετα έχουν αναπτυχθεί εύρωστοι αλγόριθμοι για διάγνωση σφαλμάτων, μέσα στα πλαίσια της θεωρίας συστημάτων. Στη διατριβή αυτή διατυπώνεται ένα θεωρητικό πλαίσιο κατάλληλο για τη διάγνωση σφαλμάτων και τον έλεγχο ασφάλειας σε συστήματα διανομής νερού. Κατ' αρχάς, παρουσιάζεται μια διατύπωση του προβλήματος παρακολούθησης και ελέγχου των δικτύων διανομής νερού, σε ένα πλαίσιο κατάλληλο για την τοποθέτηση αισθητήρων και τη διάγνωση σφαλμάτων. Με βάση το προτεινόμενο θεωρητικό πλαίσιο, εξετάζεται το πρόβλημα τοποθέτησης αισθητήρων. Το πρόβλημα ορίζεται ως εξεύρεση των κατάλληλων σημείων σε ένα δίκτυο διανομής νερού στα οποία πρέπει να εγκατασταθούν αισθητήρες που να καταγράφουν την ποιότητα του νερού σε πραγματικό χρόνο. Στόχος είναι η ελαχιστοποίηση του κινδύνου σοβαρού αντίκτυπου στην κοινωνία, σε περίπτωση που μια μολυσματική ουσία εισέλθει στο δίκτυο διανομής νερού. Επιπρόσθετα, εξετάζεται το πρόβλημα προγραμματισμού των δειγματοληψιών στο δίκτυο διανομής, με σκοπό να εξευρεθεί που και πότε να διενεργούνται δειγματοληψίες νερού για ποιοτικό έλεγχο. Ακολούθως, σχεδιάζεται ένας αλγόριθμος ρύθμισης της συγκέντρωσης της απολυμαντικής ουσίας σε δίκτυα διανομής νερού, ο οποίος στηρίζεται στη χρήση τεχνικών προσαρμοστικής προσέγγισης για εκμάθηση της ζήτησης νερού από τους καταναλωτές. Στη συνέχεια, εξετάζεται το πρόβλημα ανίχνευσης υδραυλικών σφαλμάτων διαρροής σε διακρι-

τές ζώνες ελέγχου (District Metered Areas), με τη χρήση μεθόδου ανίχνευσης σφαλμάτων που βασίζεται στην εκμάθηση των περιοδικών δυναμικών κατανάλωσης. Τέλος, αποτιμάται ο βαθμός κινδύνου εξαιτίας μόλυνσης που έχει ανιχνευτεί σε ένα δίκτυο διανομής νερού, και απομονώνεται η περιοχή στην οποία βρίσκεται η πηγή του σφάλματος, με τη χρήση μεθοδολογίας βασισμένης στην επαγωγή δέντρων απόφασης. Η αποτελεσματικότητα των μεθοδολογιών που έχουν προταθεί στα πλαίσια της διατριβής εξετάζεται με τη χρήση προσομοιώσεων που χρησιμοποιούν μοντέλα δικτύων διανομής νερού και ιστορικά υδραυλικά δεδομένα.

Acknowledgments

First, I would like to express my warmest thanks and gratitude to my supervisor, Prof. Marios Polycarpou. As all great teachers, he has shown me through his example, how to be an independent researcher; through his tutoring, how to pursue analytical thinking, and through his vision, how to set and reach high goals.

I would like to express my thanks to the people I have collaborated with during these past years in various projects, to Stelios Neophytou, Michael Markou, Saikat Chakrabarti, Panagiotis Panagi, Christos Laoudias, Michalis Michaelides, Demetris Stavrou, Michalis Skitsas and George Milis as well as to the rest of the colleagues at the KIOS Research Center, who make coffee and lunch breaks something to look for.

In addition, I would like to express my thanks to Evripides Kyriakides of the Water Development Department of Cyprus, and to Bambos Charalambous of the Water Board of Limassol, who have helped me understand the water systems domain, as well as the difficulties faced by practitioners.

I would like to thank the faculty of the Electrical and Computer Engineering Department and in specific Christos Panayiotou and Elias Kyriakides for the fruitful discussions and recommendations for my research, as well as the KIOS administrative assistants, and in specific Skevi Chrysanthou, who has been there for me from the start.

During my PhD, I've had the opportunity to work at the University of Trieste in Italy, along with a great team of researchers led by Prof. Thomas Parisini; I thank them all for their hospitality. No acknowledgment by this author would be complete, without expressing my deepest friendship and gratitude to Riccardo Ferrari, who has helped me to forward my research in a time I felt I was getting no-where.

I would also like to acknowledge the Research Promotion Foundation of Cyprus and the University of Cyprus, for supporting my research during these years.

This thesis is dedicated to my wife, Maria, who is the pole star in my life's journey, and to Giorgos-Efstathios, who illuminates my life; both of you are *mon raison d'être*.

Publications

Published journal publications

1. D. Eliades, M. Polycarpou, and B. Charalambous, "A security-oriented manual quality sampling methodology for water systems," *Water Resources Management*, vol. 25, no. 4, pp. 1219–1228, Mar. 2010.

2. D. Eliades and M. Polycarpou, "A fault diagnosis and security framework for water systems," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 6, pp. 1254 –1265, Nov. 2010.

3. D. Eliades, M. Polycarpou and B. Charalampous,"Security issues in drinking water distribution networks," *Journal of Information Assurance and Security*, vol. 4, no. 6, pp. 500–508, 2009.

4. A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S.-T. Khu, D. Savic, D. Eliades, M. Polycarpou, S. R. Ghimire, B. D. Barkdoll, R. Gueli, J. J. Huang, E. A. McBean, W. James, A. Krause, J. Leskovec, S. Isovitsch, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. B. Trachtman, Z. Y. Wu, and T. Walski, "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," *ASCE Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 556–568, Nov./Dec. 2008.

5. S. Chakrabarti, E. Kyriakides, and D. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 12–19, Jan. 2009.

Published conference proceedings

1. D. G. Eliades and M. M. Polycarpou, "Fault isolation and impact evaluation of water distribution network contamination," in *Proc. IFAC World Congress*, 2011, p. 6, (to appear)

2. D. G. Eliades, M. M. Polycarpou, and B. Charalampous, "A security-oriented manual quality sampling methodology," in *Proc. European Water Resources Association*, Limassol, Cyprus, 2009, p. 8.

3. D. G. Eliades, M. M. Polycarpou, and B. Charalampous, "A security-oriented manual quality sampling methodology for water systems," in *Proc. Electrical and Computer Engineering Student Conference*, Thessaloniki, Greece, 2009, p. 6.

4. D. G. Eliades and M. M. Polycarpou, "Security of water infrastructure systems," in *Critical Information Infrastructure Security*, ser. Lecture Notes in Computer Science, R. Setola and S. Geretshuber, Eds. Berlin / Heidelberg, Germany: Springer, 2009, vol. 5508, pp. 360–367.

5. D. G. Eliades and M. M. Polycarpou, "Security issues in drinking water distribution networks," in *Proc. Computational Intelligence in Security for Information Systems*, 2009, pp. 69–76.

6. C. Laoudias, D. Eliades, P. Kemppi, C. Panayiotou, and M. Polycarpou, "Indoor localization using neural networks with location fingerprints," in *Proc. International Conference on Artificial Neural Networks*, ser. Lecture Notes in Computer Science, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds., vol. 5769. Limassol, Cyprus: Springer Berlin / Heidelberg, 2009, pp. 954–963.

7. D. G. Eliades, E. C. Kyriakides, and M. M. Polycarpou, "Enhanced robustness in model predictive control of water quality using adaptive estimation methods," in *Proc. Computing and Control for the Water Industry*, Leicester, UK, 2007, p. 8.

8. D. G. Eliades and M. M. Polycarpou, "Multi-objective optimization of water quality sensor placement in drinking water distribution networks," in *Proc. European Control Conference*, Kos, Greece, 2007, pp. 1626–1633.

9. S. Chakrabarti, D. Eliades, E. Kyriakides, and M. Albu, "Measurement uncertainty considerations in optimal sensor deployment for state estimation," in *Proc. IEEE International Symposium on Intelligent Signal Processing*, 2007, p. 6.

10. D. Eliades and M. Polycarpou, "Iterative deepening of Pareto solutions in water sensor networks," in *Proc. ASCE Water Distribution Systems Analysis*, Cincinnati, USA, 2006, paper 114, p. 14.

11. D. G. Eliades, A. L. Symeonidis, and P. A. Mitkas, "Genecity: A multi agent simulation environment for hereditary diseases," in *Proc. IEEE International Conference on Computer Systems and Applications*, Dubai, UAE, 2006, pp. 529–536.

12. D. G. Eliades, A. L. Symeonidis, and P. A. Mitkas, "Genecity: A multi agent simulation environment for hereditary diseases," in *Proc. European Workshop on Multi Agent Systems*, Brussels, Belgium, 2005, pp. 137–147.

Unpublished

1. D. G. Eliades and M. M. Polycarpou, "Leakage fault detection in a district metered area of a water distribution system," 2011, (in preparation).

2. D. G. Eliades and M. M. Polycarpou, "Water contamination impact evaluation and source-area isolation using decision trees," 2011, (submitted for review).

Other academic publications

1. A. C. Emilianides and D. G. Eliades, *Electronic Voting: Reality and Opportunities*, A. C. Emilianides, Ed. Nicosia, Cyprus: Legal Publications Dikeonomia, 2007, (in greek).

2. N.-G. Eliades and D. G. Eliades, *HAPLOTYPE ANALYSIS: Software for analysis of haplotype data*, Forest Genetics and Forest Tree Breeding, Georg-August University Goettingen, Germany, 2009. [Online]. Available: `http://www.uni-goettingen.de/en/134935.html`

3. D. G. Eliades, "An Intelligent Language Tutoring System (ILTS) for Anglophobe learners of passive voice in French," M.Sc thesis, University of Edinburgh, 2005.

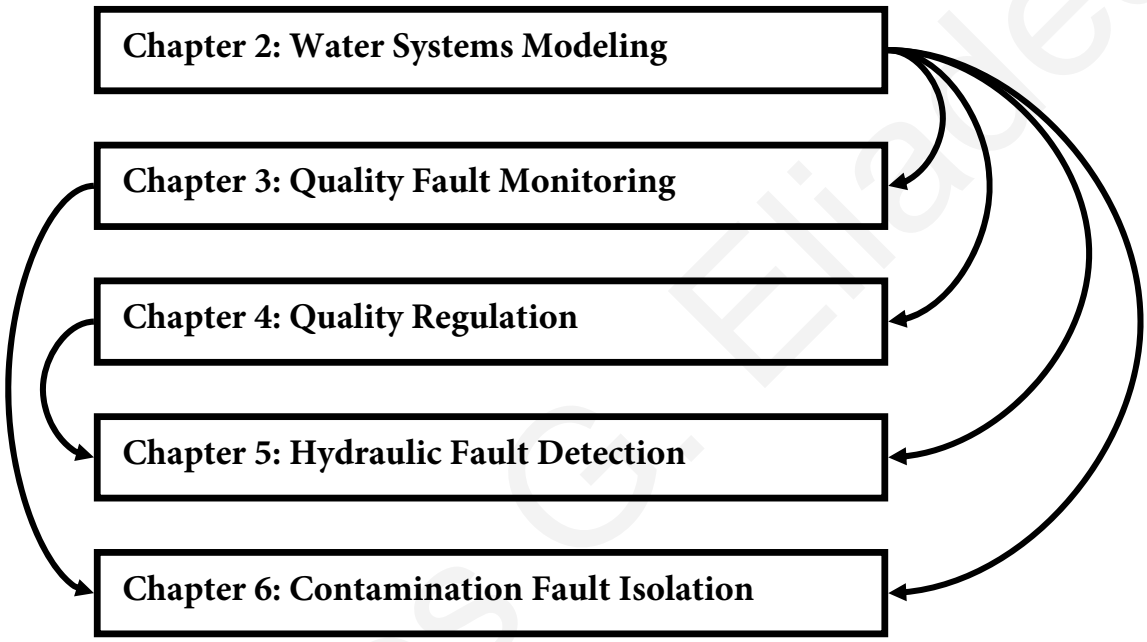
4. D. G. Eliades, "Epidemiological simulation of hereditary diseases," Diploma thesis, Aristotle University of Thessaloniki, 2004.

Other publications

1. D. G. Eliades and C. Laoudias, "Interview with Prof. Christos Papadimitriou," Synchrone Apopsi Magazine, September/October 2010. (in greek)

Contents

THESIS ROAD MAP



Nomenclature

α_f	Flow exponent
α_m	Minor loss coefficient
α_r	Pipe resistance coefficient
$\beta(k)$	Time profile of a fault
χ	Binary vector
Δt	Sampling time step
Δz	Length of a single cell in a pipe
$\delta^{(i)}$	The i-th time delay in $\mathcal{D}$
$\delta_w(\cdot)$	Difference of the average night flow with the minimum average night flow for previous days
ϵ_c	Concentration detection threshold
ϵ_d	Approximation residual error
ϵ_k	CUSUM threshold for fault start time
η_c	Penalizing factor for the quality regulation objective function
$\gamma(\cdot)$	Function to compute impulse coefficients
$\Gamma(k)$	Impulse responce matrix for predictive controller
γ_0	Expected change in the average of the offset DC term, set by the designer
$\hat{\theta}$	Parameter vector estimation of a basis function
$\hat{\theta}_0(k)$	The offset DC term of the Fourier series

$\hat{k}_0$ Estimated fault start time

$\hat{l}_0$ Estimated leakage start day

$\hat{p}(k)$ Estimated normalized demand pattern vector

$\hat{q}_d(k)$ Total water consumption at a DMA

$\hat{r}_t(\cdot)$ Estimate of the trend and seasonal signal

$\hat{s}(k)$ Estimation of the weekly periodic DMA inflow signal

$\hat{Y}(k)$ Predicted output vector

κ, κ_0 Reaction coefficients

Λ Set of the risk-level labels

$\lambda(k)$ Courant number in a pipe

$\mathbb{X}$ Set of feasible sensor placements

$\mathcal{A}$ Set of edge indices

$\mathcal{A}_i$ Set of pipe indices connected to node i

$\mathcal{A}_i^+$ Set of pipe indices of pipes which outflow water to node i

$\mathcal{B}_m$ Set of 2-tuples corresponding to node indices and time instances for manual sampling

$\mathcal{C}$ Set of scenario indices

$\mathcal{C}_0$ Set of extreme-fault scenario indices

$\mathcal{D}$ Set of discrete time delays in stopping the systems

$\mathcal{E}$ Set of fault-location matrices

$\mathcal{F}$ Finite set of contaminant injection signal structures

$\mathcal{F}^*$ The set of all contaminant injection signal structures

$\mathcal{G}_1$ Set of detection times by an on-line sensor

$\mathcal{G}_2$ Set of detection times by manual sampling

$\mathcal{G}_3$ Set of fault occurrence times

$\mathcal{H}$ Set of fault scenarios

$\mathcal{J}$ Set of overall-impact matrices

$\mathcal{J}^*$ The set of all pairs corresponding to a contamination fault's node indices and start times

$\mathcal{J}^0$ Finite set of pairs of node indices and start time

$\mathcal{K}_h$ Set of discrete times of historical hydraulic data without known leakages

$\mathcal{K}_p$ Set of the most-recent time instances set prior to the time the leakage fault begins

$\mathcal{L}_s$ Set of the candidate sensing node indices

$\mathcal{N}(\cdot)$ The normal distribution

$\mathcal{N}_f(\cdot)$ Set of discrete times corresponding to night-flows

$\mathcal{Q}^*, \mathcal{Q}$ Set of parameter matrices

$\mathcal{S}^*$ The set of all contamination fault functions

$\mathcal{S}^0$ Finite set of contamination fault functions

$\mathcal{T}$ Set of discrete times used in simulation

$\mathcal{T}_m$ Set of discrete time instances for manual sampling

$\mathcal{T}_p$ Set of period indices

$\mathcal{V}$ Set of finite volume cell indices

$\mathcal{V}_m$ Set of node indices for manual sampling

$\mathcal{V}_s$ Set of candidate nodes for installing sensors

$\mathcal{W}$ Nodes serving consumers

$\mathcal{W}_l$ Set of demand nodes in the l-th DMA

$\mathcal{X}$ Set of sensing node indices

$\mathcal{Y}$ Set of node location indices and/or manual sampling node locations and sampling times

$\mathcal{Y}_0$ Set of worst-case fault scenario indices

$\mathcal{Y}_p$ Parent set in evolutionary algorithm

μ^j Node index for manual sampling at the j-th sampling iteration

ν^j Contamination flag at the j-th sampling iteration

Ω Overall-impact matrix

ω Argument in the sinusoidal terms of the Fourier series

ω_s Estimation of the weekly periodic DMA inflow signal

$\overline{\delta}(k)$ Outlier detection threshold

$\overline{\Omega}$ Maximum scenario impact

$\overline{\phi}(k)$ Estimate of the average leakage outflow

$\overline{\Theta}$ Bounds of parameter matrices

$\overline{\theta}_0$ Theoretical average of the offset DC term of the Fourier series

$\overline{C}_p$ Concentration threshold for polluted water

$\overline{d}_i(k)$ Daily average demand outflow at node i

$\overline{h}_s$ CUSUM threshold for fault detection

$\overline{h}_w$ Threshold for night-flow leakage detection

$\overline{r}$ Maximum redundancy vector

$\Phi(\cdot)$ Standard normal cumulative distribution function

$\phi(\cdot)$ Quality fault

$\phi_h(\cdot)$ Hydraulic fault

$\psi(k)$ Overall-impact metric

σ, σ^2 Standard deviation and variance

τ_d Discrete duration time of a contaminant injection signal

$\mathbf{0}$ Vector of zeroes

$\mathbf{1}$	Vector of ones
Θ	Fault amplitude matrix of the basis function
θ	Parameter vector of a basis function
$\theta^*(k)$	Unknown parameter vector
θ_r	Parameter vector for trend and seasonal signal approximation
$\underline{C}$	Concentration lower bound
$\underline{r}$	Minimum redundancy vector
$\underline{u}, \overline{u}$	Lower and upper input concentration
$\underline{y}, \overline{y}$	Lower and upper output concentration
$\xi(k)$	Impact state vector
$\zeta(k)$	Vector of basis functions
$A(k)$	Advection-reaction state matrix
a_1, a_2	Miscellaneous parameters and coefficients
a_γ	Constant characterizing the volumetric rate of water consumption per person
a_λ	Period index
a_μ	Average daily consumption per person
a_ϕ	Magnitude of a contaminant injection signal
a_B	Evolution rate of the time profile of a fault
a_C	Parameter for CVaR metric
a_D	Discharge coefficient
a_E	Exponent term which depends on the leakage type
a_F	Adaptive law design parameter
A_G	Adaptive law gain parameter matrix
a_K	Fast reaction percentage

a_P Cross-sectional area in a pipe

a_R Redundancy normalization factor

a_S Probit slope parameter

a_W Average body mass

a_Ψ Median infective dose

$B(k)$ Advection-reaction input vector

$b(k)$ Water velocity vector in pipes

$b^{(i)}$ The i-th sampling 2-tuple, comprised of a node index and a sampling discrete time

C Binary output matrix corresponding to sensor locations

$C(t)$ Concentration of a substance

$C(z,t)$ Substance concentration at distance z and time t in a pipe

$C_t^0(z), C_z^0(t)$ Boundary conditions for advection equation in a pipe

$C_T^+(t)$ Concentration of storage tank inflow in continuous time

$C_A(t)$ Algebraic variables in reaction

c_i Index of the quality sensor node i in $\mathcal{V}_s$

$d_j^*(k)$ Real consumption demand outflow

$d_i(k)$ Consumer demand outflow at node i

E^i Fault-location matrix

$e_q(k)$ Estimation error

$e_r(k)$ Normalized flow estimation error

F Injected contaminant location matrix

$F^{(i)}(\cdot)$ The i-th objective function

$f_\lambda(\cdot)$ Function to compute the impact label set

$f_\mu(\cdot)$ Function to compute the next node for manual sampling

$f_{\psi}(\cdot)$ Function to compute the overall-impact metric

$f_{\xi}(\cdot)$ Non-negative function which characterizes the impact increase

$f_A(\cdot)$ Mass-balance function

$f_E(\cdot)$ Function to compute the impulse response matrix

$f_e(\cdot)$ Function to compute the entropy of a set

$f_h(\cdot)$ Function to compute the headloss

F_i The i-th objective function

$f_l(\cdot)$ Leakage fault function

$f_p(\cdot)$ Function to compute total nodal demands in a DMA

$f_R(\cdot)$ Concentration change rate function

$f_u(\cdot)$ Quality and hydraulics controller

$f_w(\cdot)$ Function to calculate the impact increase between two time instances

$f_{T_i}(\cdot)$ Function to compute the cross-sectional area tank i for certain water levels

$f_{u_c}(\cdot), f_{u_h}(\cdot)$ Quality and hydraulic controller

H Binary coverage matrix

$h^{(i)}$ The i-th fault scenario

$h_i(k)$ Hydraulic head at node i

H_p, H_c Prediction and control time horizons

I Identity matrix

$J(\cdot)$ Redundancy objective function

$J_c(\cdot)$ Objective function of the predictive controller

k Discrete time

$K(\cdot)$ The stopping-time matrix

k_0 Fault start time

k_d Fault detection time

$k_d^{(i,j)}(\epsilon_c)$ Discrete time of detection of the fault i starts at the sampling location j, with concentration above ϵ_c

k_s Activation time of leakage fault detection algorithm

L^j Impact label set

l_d Day of leakage detection

M^j Contamination fault propagation binary matrix

M_0 Minimum leakage detection window for night-flow leakage detection

M_c Number of nodes with disinfection booster actuators

M_d Number of days in detection window for night-flow analysis

M_m Number of manual samplings

M_s Number of on-line sensors

M_w Number of days to compute the difference $\delta_w(\cdot)$

M_y Number of nodes with on-line chlorine monitoring

$n(k)$ Multiplicative uncertainty term of DMA inflow

N_μ Number of samples for computing DMA inflow median

N_ξ Number of impact metrics considered

N_b Number of all possible samplings in $\mathcal{B}_m$

N_f Number of multiple objectives considered

N_h Number of nodes in coverage matrix

N_k Number of time delays considered

N_m Number of nodes (junctions and tanks)

N_n Number of nodes and finite volume cells

N_q Number of parameter matrices in $\boldsymbol{Q}$

N_s Number of candidate nodes for installing sensors

N_t Number of discrete time instances for manual sampling

N_v Number of fault scenarios in $\mathcal{H}$

N_w Number of nodes serving consumers

N_x Number of parents in evolutionary algorithm

N_y Number of previous inputs to be considered in the predictive controller

N_z Number of linearly parameterized basis functions

O^j Maximum impact metric vector

$p(k)$ Normalized demand pattern vector

$P_1(k), P_2(k), P_3(k)$ Intermediate matrices/vectors in the quality regulation objective function

$Q(t)$ Water flow in a pipe

$q_T^+(k), q_T^-(k)$ Storage tank inflow and outflow in discrete time

$Q_T^+(t), Q_T^-(t)$ Storage tank inflow and outflow in continuous time

$q_\mu(k)$ DMA inflow median

$q_j(k)$ Water flow in pipe j

$q_{in}^{(l)}(k), q_{out}^{(l)}(k)$ Total inflow and outflow in the l-th DMA

$r(k)$ Reference signal vector

R_c Diagonal "installation cost" matrix

$r_c(k), r_h(k)$ Reference quality and hydraulic signal vector

R_s Diagonal "significance" matrix

$r_t(k)$ DMA inflow trend/seasonal signal

s A contamination fault function

$S(k)$ The CUSUM metric

$s(k)$ DMA inflow weekly periodic signal

t Continuous time

$T(\cdot)$ Propagation-time matrix

t_a, t_b Discrete times at which night-flow period begins and ends

T_H Number of discrete times in one hydraulic period

T_S Number of samples within a weekly period

T_{max} Time for detecting contamination through other means

$U(k)$ Future inputs vector

$u(k)$ Control input vector

$u_c(k), u_h(k)$ Quality and hydraulic control input vector

$v(k)$ Storage tank volume in discrete time

$V(t)$ Storage tank volume in continuous time

v_i The i-th node index in $\mathcal{V}$

$w(\cdot)$ Average night-flow

W^j Normalized impact damage vector

w_i The i-th demand node

$x(k)$ Concentration state vector

$x_T^+(k)$ Concentration of storage tank inflow in discrete time

$x_i(k)$ Average concentration in the i-th finite volume cell

$x_i^+(k)$ Ghost cell concentration

$Y(k)$ Future responses vector based on past inputs

$y(k)$ Measured output vector

$y_c(k)$ Output vector of monitored concentrations

$y_h(k)$ Output vector of monitored hydraulics

z Distance in a pipe

Z^j Information gain vector computed for the j-th sampling iteration

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Water as a Human Right

Water is essential for life and for human development; it is, however, unfortunate that even today, approximately 884 million people lack access to safe water. For this reason, the United Nations General Assembly has proclaimed the years 2005–2015 as the international decade for action "Water for Life" [165]. In addition, the UN General Assembly adopted in 2010 a resolution which recognized the right to safe and clean drinking water as a human right that is essential for the full enjoyment of life and all human rights; it is the responsibility of each state to ensure that access to safe, clean and affordable drinking water is realized [166].

The supply of water is not only important for the everyday human needs, but in addition, for sustainable development, for the production of energy and for industrial and agriculture processes. Nevertheless, the increasing population growth and the improvement in living conditions, as well as economic development lead to new challenges for water resources management and water quality research. This thesis addresses some of these challenges.

1.2 Water Systems

The process of collecting, cleaning, disinfecting and delivering water to consumers can be considered as a *water system*. A water system can be described as a system comprised of multiple subsystems, due to its large-scale and decentralized nature, as well as due to the operation and managerial independence of its components. Overall, as seen in Fig. 1.1, within a water system, water from some reservoirs, wells and rivers is transferred to certain treatment plants and from there, transferred to some urban water distribution systems. Water can be

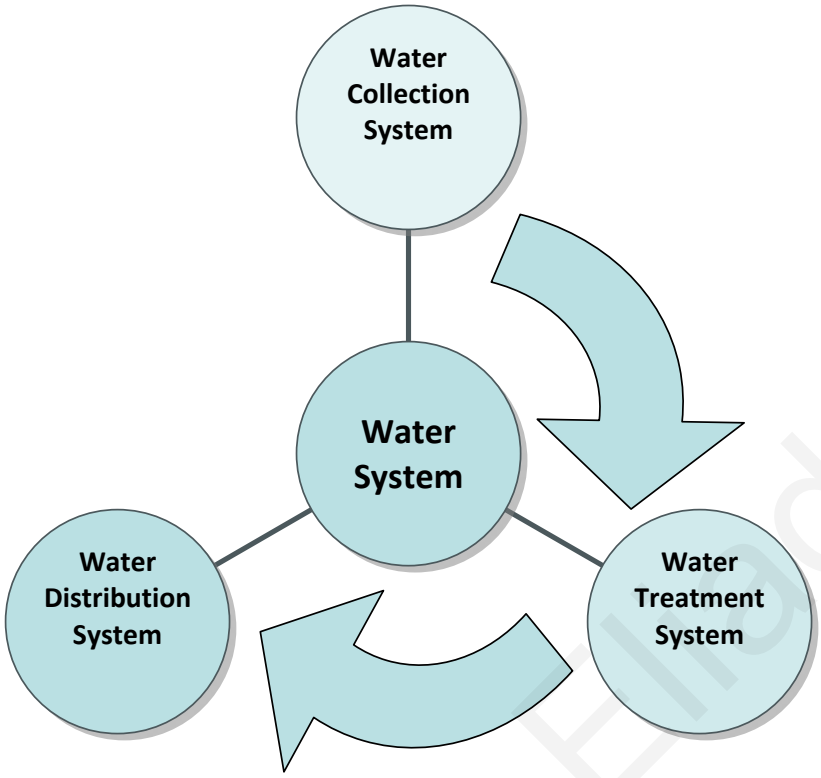


Figure 1.1: In a water system, water is collected, transferred for treatment and then transferred for distribution and consumption.

collected from various sources, such as from reservoirs, lakes, rivers or may be pumped from the ground; in the case of reservoirs, some of the water outflow may be used to produce hydroelectricity. Collected water is then transferred to the water treatment plants for processing and disinfection. Water treatment plants are factories which clean and disinfect water, passing it through a series of chemical processes while moving through different tanks. After water has been disinfected, it is transported to water utilities for distribution.

Water utilities are responsible for distributing drinking water to consumers through their drinking water distribution systems, which they must monitor and control so that the hydraulic and quality parameters of the system are within certain desired limits. Water distribution networks are large-scale systems, which are comprised of pipe networks and dynamical elements such as water storage tanks, pumps and valves to control pressures and flows in the system, as well as sensors measuring various hydraulic and quality water characteristics. Due to the large-scale nature of water distribution systems, there may be parts in the water distribution network which are decoupled for better monitoring and control. The present thesis focuses on water distribution networks and how they are monitored, controlled and secured.

Maintaining water quality within the regulations specified by the World Health Organization (WHO) [189], the European Commission [67], or the U.S. Environmental Protection

Agency (EPA) [172], is an important challenge faced by water utilities which supply water to consumers through the drinking water distribution network.

However, maintaining water quality is not an easy task, as faults may occur in the system affecting quality. For instance, leakages, pipe bursts and malfunctioning pumps and valves may downgrade water delivery or even cause quality faults. Quality faults, on the other hand, may be due to certain chemical, biological or radioactive substances which travel along the flow of water and may exhibit certain dynamics. In practice, disinfectants such as chlorine are used in prescribed concentrations to maintain the drinking water quality, by preventing bacteria growth and neutralizing chemical agents [178]. In specific, according to the WHO, a free chlorine residual concentration must exist in drinking water distribution systems, with minimum target concentration $0.2\ \frac{mg}{L}$ at the point of delivery and $0.5\ \frac{mg}{L}$ for high-risk circumstances [189]. In general, it is common practice to supply water with a few tenths of a milligram per litre of chlorine residual. Water providers are required to control and monitor the hydraulics and the water quality in the water distribution network which they operate, in order to deliver adequate disinfected water to all consumers. To satisfy this, water providers have to collect hydraulic and quality data at various locations in the network (either manually or by using sensors) and control the system appropriately. Through this, water providers are able to detect faults related to the hydraulic dynamics (pressures, flows) or quality dynamics (such as disinfectant and contaminant concentration).

Hydraulic faults, such as leakages, pipe bursts, malfunctioning pumps and valves, may interrupt water consumption or may deteriorate water quality, due to contaminant infiltration in the system. In general, water contamination faults in water distribution systems may be due to natural or accidental events.

One example of an accidental contamination was the widely covered "Nokia Water Crisis", a large-scale water contamination which occurred in the town of Nokia, Finland, in November 2007 [154]. At a water treatment plant, a drinking-water pipe was connected to a waste-water pipe through a valve. A worker had accidentally opened the valve, and as a result, within two days 400 000 litres of waste-water were injected into the drinking water distribution network. Figure 1.2 depicts the source and the extent of the contamination, which had affected more than 12 000 people, as well as some local industries. Hundreds were hospitalized, while authorities were forced to impose a complete ban on all water usage.

The case of the "Nokia Water Crisis" raises a number of questions, from a research point of view. For instance, in what way could the contamination fault have been detected early enough in order to minimize the extent of the contamination impact? How could the source

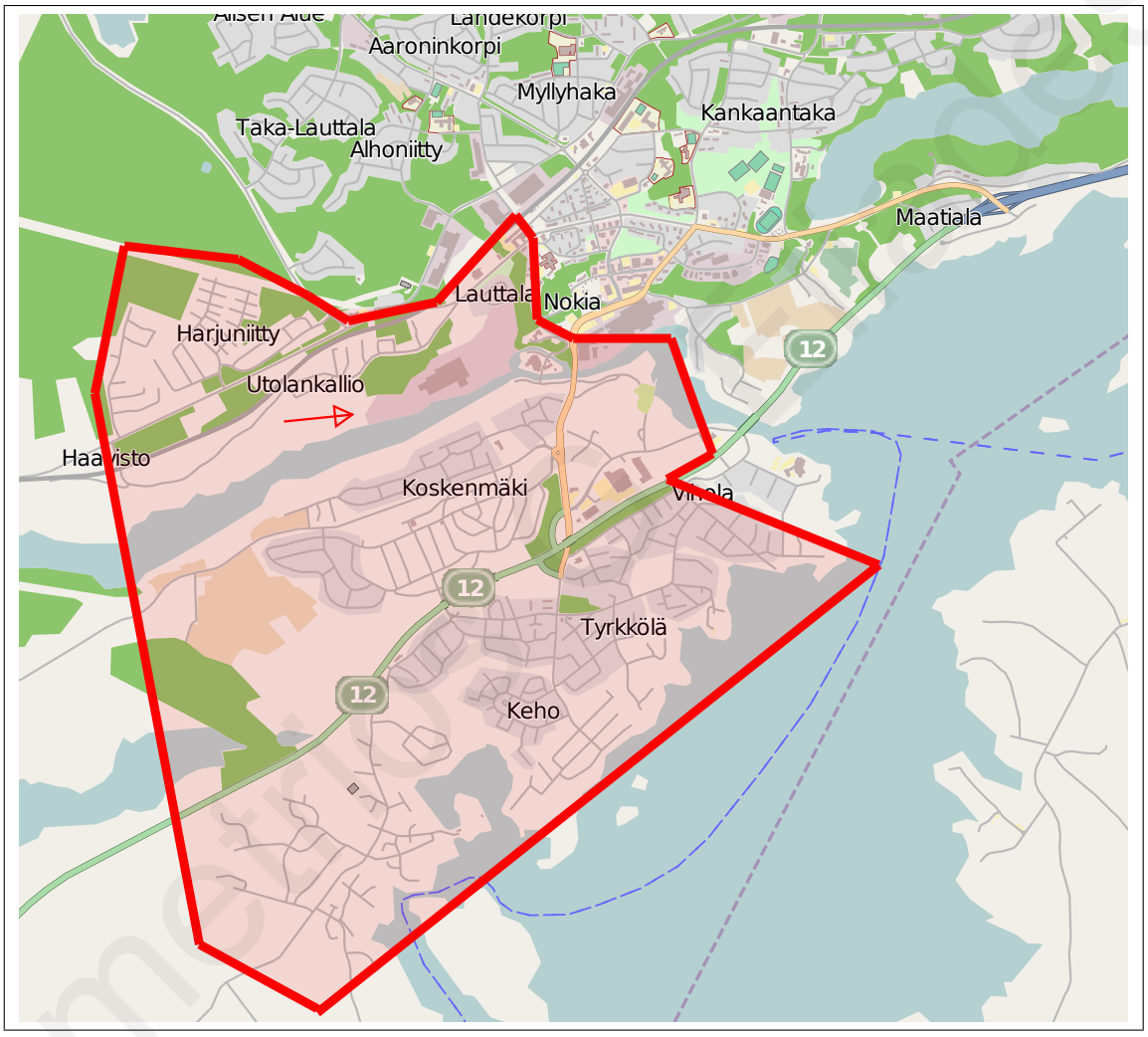


Figure 1.2: Map of Nokia, Finland. The red arrow indicates the starting location of the contamination; the contaminated area is covered with a red layer.

location of this contamination be isolated, and how could the level of risk be evaluated, as early as possible? Furthermore, how would the problem be seen from a completely different perspective, in case the contamination had in fact been a malicious attack on the system? Those are some of the questions which this thesis elaborates on.

1.3 Critical Water Infrastructures and Security

According to the European Commission, critical infrastructures are those physical facilities and networks, which, if disrupted, would significantly affect the health, safety, security or economic well-being of citizens [68]. Due to their vital role, water systems are considered among the critical infrastructures, along with power and telecommunications systems. Critical infrastructure security has received significant attention within the past few years, especially due to various terrorist acts around the globe.

Malicious faults in water distribution systems can be part of an informed attack seeking to cause economic losses and affect dramatically the served population's health. Contamination could be performed by some criminals, at the most neuralgic location in the drinking water distribution network, and at the most critical time instance, as an attempt to cause the greatest "damage".

A large number of terrorist and criminal threats or attacks on water infrastructures has been recorded in the last decades [76]. More specifically, in 1976, a biologist in Germany threatened to contaminate water supplies with Anthrax, unless he was paid $8.5 million. In 1984, followers of the Indian guru Rajneeshee contaminated water and food supplies with biological agents in The Dalles, Oregon, USA, to serve their political agenda. The followers gained access to the town's water system maps, and tried to inject contaminants to water tanks [32]. In 2000, a cyber-attack on a waste-water management system in Queensland, Australia, caused the release of millions of sewage to parks and contaminated rivers, while in 2003, the Al-Qaida called for the poisoning of drinking water in American and Western cities [76].

Protecting the critical infrastructure from malicious acts, natural hazards and accidental faults requires the development of fault diagnosis and security frameworks; this involves the synergy of various research fields, such as modeling, control, risk management and optimization. In this thesis we address a number of modeling, optimization and security issues related to the critical water infrastructure.

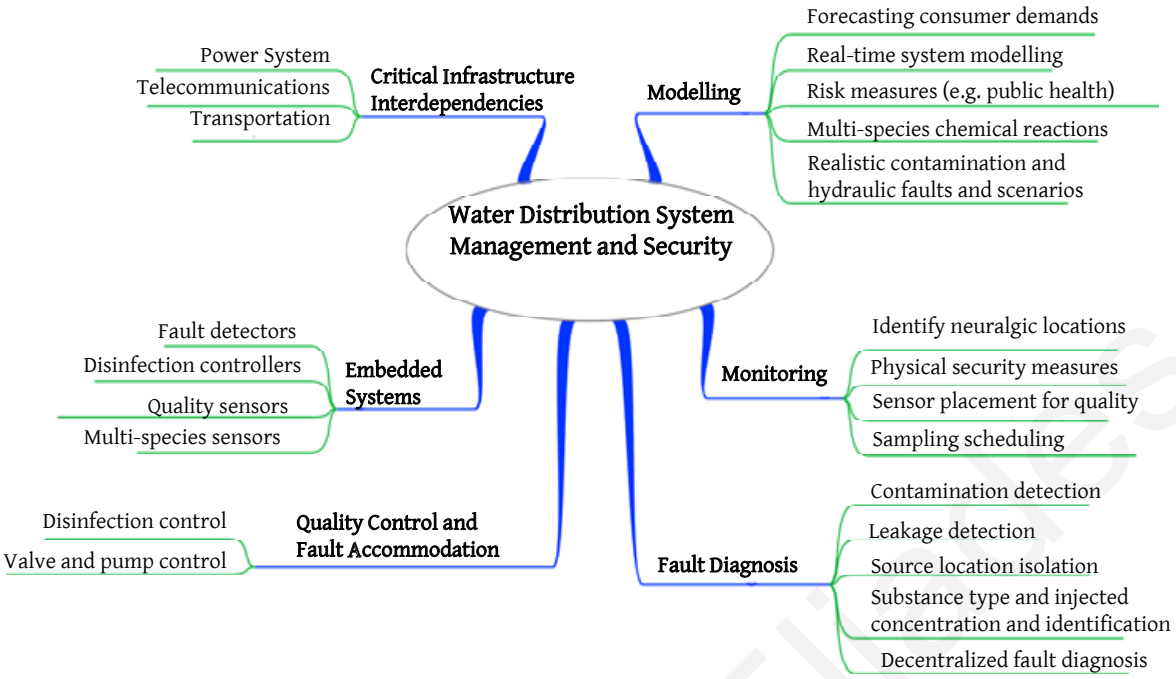


Figure 1.3: Overview of some of the key issues related to water distribution system management and security: system modelling, quality monitoring, hydraulic and quality fault diagnosis, quality control and contamination fault accommodation, embedded devices for fault diagnosis and control, interdependence analysis with other critical infrastructures. In this thesis the problems of system modelling, quality monitoring and control and the hydraulic and quality fault diagnosis are addressed.

1.4 Research Challenges in Water Systems

Due to their complexity and large-scale nature, water systems provide various challenges that can be addressed within a systems and control framework. Figure 1.3 provides an overview of some of the key issues related to water distribution system management and security: modelling and monitoring of water systems are deemed essential for quality control, hydraulic and quality fault diagnosis and fault accommodation to return the system into normal operation. Furthermore, embedded systems provide the means for constructing specialized sensors and controllers based on the developed algorithms. Finally, a key issue is to model the interdependencies between water systems with respect to other types of critical infrastructures, such as power and telecommunication networks, since control actions and faults occurring at a certain critical infrastructure may affect dramatically the operation of another. For instance, telecommunication faults may cause loss of monitoring and control of some pumps and valves in the network, and on the other hand, hydraulic faults, such as pipe breaks, may cause power losses due to power station flooding.

1.5 Thesis Motivation

Fault diagnosis describes the process of detecting and isolating a fault, by applying various techniques to monitor the changes in the states, learn the fault dynamics, and fault accommodation describes the process of adapting the input so that the system returns to safe operation. In addition, security monitoring describes the supervision of the distribution network, aiming at minimizing the potential economic losses and the damages inflicted on the consumers, as a result of a fault or a malicious attack. Security monitoring and fault diagnosis in water distribution systems belong to an area of increasing research interest, and require the synergy of various fields, such as control systems, water engineering, optimization and risk management.

The risk of hydraulic and quality faults in water distribution systems which may cause water losses or water contamination, with the danger of deteriorating the consumers' health, constitutes the motivation of this research.

In the previous years, various aspects of the security monitoring problem in water distribution systems were examined. In addition, robust fault diagnosis algorithms were developed within a system-theoretic framework. An open research area is the formulation of a system-theoretic framework suitable for fault diagnosis and security monitoring in water distribution systems; that is the problem we investigate.

The general thesis contributions are: a) formulates the problem of monitoring and control in water distribution networks, using a mathematical framework which is suitable for the sensor placement and the fault diagnosis problem; b) provides algorithms to find the locations in a water distribution network where quality sensors should be installed, as well as where and when to conduct manual quality sampling, in order to monitor the water chemical characteristics; c) to regulate the disinfectant concentration in a water distribution network within some desired levels; d) designs algorithms to isolate the source area and evaluate the impact of the contamination; and e) designs algorithms to detect hydraulic faults in a water distribution network, using measurements from standard monitoring sensors.

1.6 Thesis Outline

In the following, we present the outline of each chapter and the most important contributions, together with any relevant publications.

Chapter 2: Water Systems Modeling

This chapter provides background information on the modeling of the hydraulic dynamics in water distribution systems, as well as on modeling the advection-reaction quality dynamics in distribution networks. In addition, we define the hydraulic and quality control problem, as well as present models describing the hydraulic faults, the quality faults and the impact dynamics. The models presented serve as the basis for the development of a mathematical framework suitable for fault diagnosis and security in water systems.

The contribution of this work is the formulation of the advection-reaction and quality fault dynamics, coupled with the impact dynamics, into a state-space representation, which is used as a framework for the optimization and control problems addressed in this thesis.

Part of the work in this chapter has been published and presented in peer-reviewed journals [58, 61, 66] and conferences [59, 60, 63, 64].

Chapter 3: Quality Fault Monitoring

This chapter provides background information on the quality sensor placement and manual sampling scheduling problem in water distribution systems. The problem of quality sensor placement for maximum redundancy, when the steady-state hydraulic dynamics can be expressed as a graph, is formulated and solved as an integer quadratic optimization program. This work is extended when the system dynamics are available, for the sensor placement and manual sampling scheduling problem, as well as for the problem for identifying the most neuralgic locations in a water distribution network. We formulate an optimization problem in which one or more risk-objective functions are minimized to compute the locations where to install quality sensors, or, in the case of the manual sampling problem, compute the locations and time where and when to perform manual sampling. The methodology is demonstrated on an illustrative network, on a benchmark network, as well as on part of a water distribution network in Cyprus.

The contributions of this work are the presentation of a rigorous mathematical formulation and the solution of the maximum redundancy sensor placement problem, as well as the security-oriented sensor placement problem considering risk-objective functions; the formulation is extended for the security-oriented manual sampling scheduling problem.

The work presented in this chapter has been published and presented in peer-reviewed journals [58, 61, 66] and conferences [55, 59, 60, 62–64, 123]. In addition, the maximum redundancy formulation has been applied to power transmission networks for Phasor Measurement Units location selection; part of the results have been published in a peer-reviewed journal and a conference [33, 34].

Chapter 4: Quality Regulation

This chapter addresses the issue of designing algorithms for the regulation of the spatial-temporal distribution of chlorine residual in drinking water distribution networks, which are influenced by unknown time-varying water demands. A solution methodology is presented using model predictive control principles combined with on-line adaptive forecasting of the system hydraulics as they are driven by consumer demands. The periodic nature of water demands allows the use of Fourier series with coefficients which change adaptively, for forecasting future flows at various nodes in the distribution network. The objective is to regulate chlorine residuals, which act as a disinfectant for enhanced water quality, within certain lower and upper concentration bounds. Simulation results on a distribution network are used to illustrate the performance of the proposed chlorine regulation algorithm under unknown water demands and the trade-offs involved in the chlorine regulation methodology.

The contributions of this work is the design and implementation of a quality regulation algorithm which learns the unknown demands, while using existing water distribution system models to compute the input signals.

The work presented in this chapter has been published and presented in a peer-reviewed conference [54].

Chapter 5: Hydraulic Fault Detection

In this chapter, we formulate the problem of leakage detection in a systems engineering framework, and propose a solution methodology to detect leakages in a class of distribution systems. In particular, we consider the case when water utilities use flow sensors to monitor the water inflow in a District Metered Area (DMA). The goal is to design algorithms which analyze the discrete inflow signal and determine as early as possible whether a leakage has occurred in the system.

The DMA inflow signal is normalized to remove long-term trends and seasonal effects, and two different algorithms are presented for leakage detection. The leakage fault detection algorithm presented in this work is based on learning the unknown, time-varying, weekly

periodic DMA inflow dynamics, with the use of an adaptive approximation methodology to update the coefficients of a Fourier series; as detection logic we utilize the CUSUM algorithm. For reference and comparison, we present a solution methodology based on night-flow analysis, using the normalized DMA inflow signal. To illustrate the solution methodology, we present results based on real hydraulic data measured at a DMA in Limassol, Cyprus.

The contribution of this work is the design and implementation of an automated leakage detection algorithm, which utilizes measurements which are already available in some DMA, within a adaptive approximation-based framework.

A journal paper on the results presented in this chapter is in preparation [56].

Chapter 6: Contamination Fault Isolation

When a contamination is detected in a water distribution network, expanded manual sampling can be performed at the nodes of the network, so as to determine the extent of the contamination. Choosing where to perform expended sampling can be a challenging task, due to the large-scale nature of the distribution network and the partially unknown hydraulic dynamics.

In this chapter we propose a computational methodology to select a sequence of nodes to perform expanded sampling. The goal is to evaluate the water contamination impact, and isolate the source-area of the contamination, with as few samples as possible. We consider that the water utility has a number of fixed quality sensors installed in the distribution network, and that manual quality sampling can be conducted by a contamination response team at any feasible location in the distribution network. After the triggering of a contamination alarm by a quality sensor, and upon its verification as an actual contamination by the utility operator, a manual sampling scheduling scheme is computed. The scheduling scheme gives guidelines with regard to which nodes the contamination response team should sample, in order to isolate the source-area and to evaluate its possible impact, as quickly as possible. The proposed method is based on decision tree induction; the conditional terms of the decision tree indicate where expanded manual sampling should be conducted, with a certain order, aiming at evaluating the possible fault-impact and at isolating the source region. To illustrate the solution methodology, we present results based on a simplified network and a real water distribution system benchmark.

The contributions of this work are the mathematical formulation, the design and the implementation of an expanded manual sampling scheduling algorithm; as added benefit, the

proposed algorithm is in accordance to the guidelines specified by the EPA as part of the *consequence management plan* when a contamination is detected in a water distribution network.

A journal paper on the results presented in this chapter is in a peer-review stage [57], and part of the results have been accepted to appear in a peer-reviewed conference [65].

Chapter 7: Conclusions

In this chapter we present some concluding remarks, as well as directions for future research.

Chapter 2

Water Systems Modeling

In this chapter we provide background information on the modeling of the hydraulic dynamics in water distribution systems, as well as on modeling the advection-reaction quality dynamics in distribution networks. In addition, we define the hydraulic and quality control problem, as well as models describing the hydraulic faults, the quality faults and the impact dynamics. The models presented serve as the basis for the development of a mathematical framework suitable for fault diagnosis and security in water systems.

2.1 Description of Water Distribution Networks

A water distribution network is the infrastructure responsible for delivering drinking water to consumers. Water enters the network after it has been collected from rivers, lakes, dams or underground sources and has been cleaned at treatment plants. Water distribution networks are comprised of pipes which are connected to storage tanks, reservoirs or other pipes using junctions, starting from the facilities of the water provider and reaching all the consumers. Water is supplied to consumers through various points in the network, namely the outflow nodes. Valves are usually installed to some of the pipes to reduce flow or pressure, or to isolate or close part of the network. Pumps deliver energy to the system by increasing the pressure at some locations. Both valves and pumps are considered as hydraulic actuators, which may be controlled through automated or manual feedback signals. Tanks which are connected to the network, fill or empty according to a time schedule or are regulated through feedback controllers. Demand is the water outflow due to consumer requests. Although such requests occur randomly throughout the day, in the macro-level they have some common characteristics, such as approximate periodicity or consumption patterns, which can be both learned and predicted.

In a water distribution network, hydraulic and quality parameters are usually measured through a Supervisory Control And Data Acquisition (SCADA) system. Hydraulic monitoring is quite common for water utilities, which measure flow and pressure at various points of the network in order to observe consumption behaviour and detect leaks. Quality monitoring, on the other hand, is more recent and involves performing mostly manual sampling or installing quality sensors at various locations, to determine the chemical concentrations of various chemical species such as chlorine (used for disinfection) or certain contaminants.

Modeling methodologies for hydraulic and quality dynamics and their faults have received significant attention during the last decade [78, 148], and are still an area of active research. Details regarding hydraulic and quality modeling can be found in [25, 80, 88, 99].

2.2 Modeling Hydraulics

The hydraulic analysis problem in water distribution networks is defined as the problem of computing the hydraulic head at each junction and the flows at each pipe. To solve this problem, the topology of the network and pipe characteristics, the control inputs, as well as the demand at each node, need to be known. In general, structural information of the network is available by the water utilities; however, pipe characteristics may require field measurements, and nodal demands at each discrete can only be estimated using historical data and other hydraulic measurements available, if no online demand sensors are used by the utility to monitor each consumer.

In general, a set of ordinary differential equations can be used to describe the dynamic relation of water flow in pipes and the differences in the hydraulic heads [99]. However, in practice, approximation of the actual hydraulic dynamics are considered in discrete time, in steady state, and by using an iterative optimization algorithm (e.g. gradient descent), the heads and flows are computed, so that the conservation of mass and energy is satisfied, e.g. [161].

To establish the notation, consider a water distribution network composed of pipes, junctions and water storages. The topology of this network can be represented as a graph with edges corresponding to pipes, and nodes corresponding to junctions and water storages. At discrete time k with sampling time Δt, let $d_i(k)$ be the consumer demand outflow at the i-th junction node, and let $q_j(k)$ correspond to the flow in the j-th pipe connected to junction i ($j \in \mathcal{A}_i$ where $\mathcal{A}_i$ is the set of pipe indices which are connected to the i-th node, assuming that inflows have a positive sign and outflows have a negative sign). In accordance to the principle of mass conservation, the sum of all the pipe inflows and pipe outflows must equal to

the demand (Kirchhoff's junction rule),

$$\sum_{j \in \mathcal{A}_i} q_j(k) = d_i(k).$$

Furthermore, in accordance to the principle of energy conservation, the flow-headloss relationship across each link in the network must be balanced. Let $h_i(k)$ be the *hydraulic head*, i.e. a measurement of water pressure expressed in length units, at the i-th node. For water moving from node j (higher head) to node i (lower head) with flow $q_l(k)$ in the l-th pipe, the flow-headloss relationship is given by

$$h_j(k) - h_i(k) = f_h(q_l(k)) \tag{2.1}$$

where $f_h(\cdot)$ is a nonlinear function, such that $f_h(q_l(k)) = \alpha_r q_l(k)^{\alpha_f} + \alpha_m q_l(k)^2$, which depends on the pipe resistance coefficient α_r, the flow exponent α_f and the minor loss coefficient α_m. These parameters are computed using empirical methods; for example, by considering the *Hazen-Williams* headloss relation, the flow exponent is $\alpha_f = 1.852$ and the resistance coefficient α_r is calculated using a nonlinear function which takes as arguments the pipe diameter, the pipe length and a unitless roughness coefficient which depends on the pipe material and has been computed empirically. The minor loss coefficient α_m is given empirically by the pipe fitting type [142]. Therefore, for a water distribution network, the set of hydraulic equations is constructed, and at each discrete time, a gradient optimization algorithm is solved using the current demand flows, control inputs and tank head [161].

Tanks are dynamic elements in the system and can be considered as nodes in the water distribution network; the head state of the i-th water tank node is given by

$$h_i(k+1) = h_i(k) + \frac{\sum_{j \in \mathcal{A}_i} q_j(k)}{f_{T_i}(h_i(k))} \Delta t,$$

where the tank head $h_i(k)$ corresponds to the relative water level plus the tank elevation, and function $f_{T_i}(\cdot)$ computes the cross-sectional area of the i-th tank at a certain height. Initial tank heads are typically known.

Currently, a number of off-the-shelf software are used to perform the hydraulic analysis in water distribution networks, such as the open-source EPANET [142]. To capture the time-variance of flows and pressure due to consumer water demands, these systems perform "extended-period simulations", i.e. at discrete time k, solve the steady-state equations, compute the state of the dynamic elements in the network for $k+1$ (i.e. the tanks), apply any control inputs and at discrete time $k+1$, repeat the procedure.

2.2.1 Conservation Equations Example

To illustrate the conservation equations, consider the network in Fig. 2.1. Each junction i has a demand flow $d_i(k)$ at discrete time k. The arrows indicate the flow of water in pipes. The

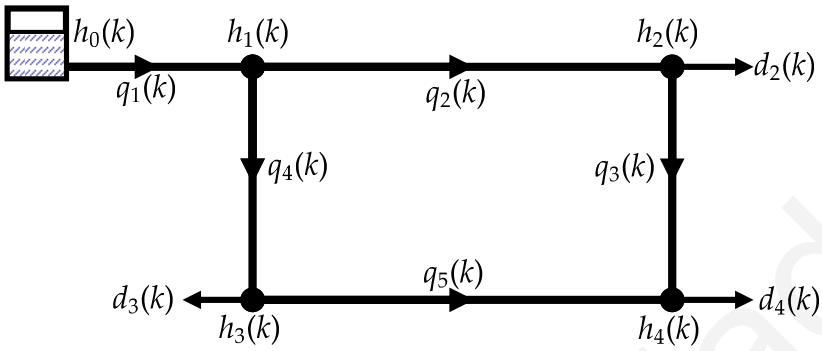


Figure 2.1: A simple water distribution network with four nodes; water is supplied by a tank.

mass conservation equations are given by

$$\begin{aligned} q_1(k) - q_2(k) - q_4(k) &= 0 \\ q_2(k) - q_3(k) &= d_2(k) \\ q_4(k) - q_5(k) &= d_3(k) \\ q_5(k) + q_3(k) &= d_4(k). \end{aligned}$$

For the conservation of energy, the equations are given by

$$\begin{aligned} h_0(k) - h_1(k) &= f_h(q_1(k)) \\ h_1(k) - h_2(k) &= f_h(q_2(k)) \\ h_2(k) - h_4(k) &= f_h(q_3(k)) \\ h_1(k) - h_3(k) &= f_h(q_4(k)) \\ h_3(k) - h_4(k) &= f_h(q_5(k)). \end{aligned}$$

To compute the headloss function $f_h(\cdot)$, certain parameters are considered known or are computed empirically. In addition, the demand flows at each node $d_i(k)$, for $i \in \{1, ...4\}$, are known, as well as the initial tank head $h_0(0)$. In this example we have constructed nine equations with nine unknowns. At time k, an optimization algorithm is used to compute the unknown parameters; then the new tank head is recalculated and the problem is solved for the next discrete time $k + 1$.

2.2.2 Challenges in Hydraulic Modeling

From a controls viewpoint, variations in demand flows are considered as time-varying disturbances, which affect flows in pipes and pressures throughout the network. In general, consumer demands are influenced by weather conditions, season, population growth, change of habits, even changes due to the response actions after a contamination . In practice, consumer demands are rarely measured online for each node; this information, however, is necessary to solve the hydraulic equations. Some information is acquired when water utilities measure the consumed water volume for a period of some months, for billing purposes. From those data, an average daily consumption demand could be calculated for each junction. Time varying consumer demands can be further estimated by using some flow measurements from within the network, and calculate water demand estimations.

Furthermore, in the hydraulic model discussed, we assume that some information of the system is known, such as the pipe characteristics, the initial tank levels and the pump flow/pressure equation. In addition, demands are assumed to be independent with respect to the pressure at the point of outflow; thus the hydraulic solver discussed is entirely demand-driven [160]. In some research, extensions to the demand-driven hydraulic model have been proposed, for pressure-driven analysis [75].

2.3 Modeling Quality Dynamics

Quality dynamics in water systems corresponds to the concentration of various contaminant or disinfectant substances, as well as other water chemical parameters, such as pH or turbidity; in this work, by water quality we refer to the concentration of some chemical substance in the water distribution system. Contaminants and disinfectants travel along the water flow within the pipe network, according to the advection and reaction dynamics. Advection is the transport mechanism of a substance in a fluid, which can be modelled as a hyperbolic partial differential equation and can be solved using various numerical methods [101]. Advection dynamics describes how a substance concentration propagates in space and time; reaction dynamics describes the change in the substance concentration due to decay, growth, or reaction with other substances. Advection and reaction dynamics are coupled to describe the quality dynamics. In addition, water quality in storage tanks is computed dynamically, and it depends on the inflows and their quality, the outflows, the tank volume and its quality.

2.3.1 Advection Dynamics

When a substance enters a pipe in which water flows, the substance moves along with that flow. Inside a pipe, and by neglecting axial dispersion and the substance reaction dynamics, the first-order hyperbolic partial differential equation which describes the change in space and time of the substance concentration in water, is given by

$$\frac{\partial C(z,t)}{\partial t} + \frac{Q(t)}{a_P}\frac{\partial C(z,t)}{\partial z} = 0, \tag{2.2}$$

where $C(z,t)$ is the substance concentration in water at continuous time t and at distance z along a certain pipe, with water flow $Q(t)$ and cross-sectional area of the pipe a_P. The boundary conditions are given by $C(z,0) = C_t^0(z)$ and $C(0,t) = C_z^0(t)$.

In water distribution system quality modeling, two main methodologies have been considered for discretizing the set of hyperbolic partial differential equations describing the advection dynamics within the pipe network, the *Eulerian* and the *Lagrangian* approaches [142–144]. In general, the Eulerian schemes assume that the water moves between a fixed grid point (finite differences) or volume segments (finite volumes), with a constant time-step [144]; a finite volume methodology was used in EPANET version 1.1 [141]. On the other hand, the Lagrangian method considers variable-sized water segments, with a constant time-step, unless an event has occurred; the event-driven method is implemented in EPANET version 2.0 [142, 143].

Next we provide more intuition to the formulation of a mathematical model describing the advection in water distribution systems, by presenting the *Finite Volume Method* [101]. This numerical method can be employed to approximate the set of hyperbolic partial differential equations which describe the advection dynamics. This requires to virtually segment all the pipes in the network, into a number of finite volume cells, while the Courant-Friedrichs-Lewy (CFL) condition is required for the convergence of the solution [101]. Consider a substance moving within a pipe which is segmented into a finite number of volume cells; for the i-th finite volume set, we define $x_i(k)$ as the average concentration in that cell, such that

$$x_i(k) = \frac{1}{\Delta z}\int_{z_i}^{z_i+\Delta z} C(l, k\Delta t)dz,$$

where Δt is the length of a hydraulic discrete time-step, which is a design parameter and may depend on the available sensors, Δz is the width of a single cell and z_i is the distance of node v_i from the pipe inflow point. Both Δt and Δz are assumed to be fixed within a certain pipe.

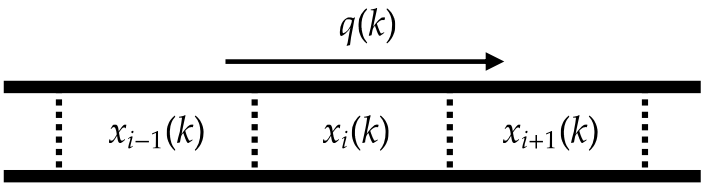


Figure 2.2: A pipe segmented into virtual finite volumes

Numerical Approximation Schemes

Various numerical approximation schemes can be used for solving the hyperbolic differential equation (2.2), such as the *leapfrog* or the *Lax-Wendroff*, which is second order accurate for smooth solutions [101]. Consider a pipe which is segmented into a number of finite volume cells, and the discrete flow $q(k) = Q(k\Delta t)$; for the i-th finite volume which is not at the boundaries of the pipe, as in Fig. 2.2. The *leapfrog* numerical scheme is given by

$$x_i(k+1) = x_i(k) - \lambda(k)\left(x_{i+1}(k) - x_{i-1}(k)\right),$$

and the *Lax-Wendroff* scheme is given by

$$x_i(k+1) = \frac{\lambda(k)}{2}[1+\lambda(k)]x_{i-1}(k) + [1-\lambda(k)^2]x_i(k) - \frac{\lambda(k)}{2}[1-\lambda(k)]x_{i+1}(k),$$

where $\lambda(k) = \frac{q(k)}{a_P}\frac{\Delta t}{\Delta z}$ is the Courant number, for a_P the cross-sectional area of the pipe; this must satisfy the CFL condition, i.e. $\frac{|q(k)|}{a_P}\frac{\Delta t}{\Delta z} \leq 1$ for that pipe to guarantee stability in the solution [101]. The direction of the flow does not affect these specific approximations; however in the boundary cells we need to reformulate the equations to capture the network behaviour.

Boundary Conditions

Boundary conditions need special treatment, since, depending on the flow direction and the numerical method selected, the concentration of a finite volume outside the pipe may be needed for calculating the state. A technique is to virtually extent the pipe by adding *ghost cells* at the ends, with some virtual concentrations [101]. These cells will be used to compute the boundary states in the pipe. The choice of what values to place in these ghost cells is not related to the numerical solution methodology. At each new time-step, we know the internal states (or initial conditions) and apply a boundary condition procedure to determine the values of these virtual cells.

In the case of junctions, if the water flows from the last cell into the junction, then we consider that the ghost cell concentration $x_i^+(k)$ has the same concentration as the i-th cell, i.e. $x_i^+(k) = x_i(k)$.

In the opposite case, we need to compute the concentration at the junction node, as a weighted sum of the concentrations which inflow. Let $\mathcal{A}_i^+$ be the set of pipe indices which deliver water to the i-th node; we assume that all inflows have positive values. However, if the water flows from the junction into the cell, we need to compute the overall concentration by considering the inflows; thus the overall concentration is given by the inflow-weighted sum of concentrations. Let $x^+_{(T,j)}(k)$ be the outflow concentration of the pipe $j \in \mathcal{A}_i^+$. The concentration at the i-th junction node is given by

$$x_i(k) = \left[\sum_{j\in\mathcal{A}_i^+} q_j(k) x^+_{(T,j)}(k)\right] \cdot \left[\sum_{j\in\mathcal{A}_i^+} q_j(k)\right]^{-1}, \tag{2.3}$$

as described in [142].

Remark: According to the Finite Volume method, the network must be segmented into a finite number of volume cells; the number of finite cells as well as the time step considered are crucial to guarantee stability in the approximation. In general for this method, assuming that the time step Δt does not change, it is considered that an optimization algorithm is solved at each discrete time, in order to compute a new Δz for each pipe. A different approach is to solve a nonlinear optimization problem in which a pre-determined number of finite volumes is distributed at each pipe, so that the minimum time step for which stability is guaranteed, is computed.

2.3.2 Reaction Dynamics

In this section we provide some background on the reaction dynamics in water distribution systems. The reaction dynamics characterize how the concentration of one or more substances changes, when reacting, decaying or growing within a finite volume of water. Single-species reaction dynamics are widely used in research, to describe the rate of decay or growth of a substance [142]. Recently, there has been interest in modeling multiple-species reactions, which involves coupled sets of differential and algebraic equations, such that

$$\begin{aligned} \frac{dC(t)}{dt} &= f_R(C(t), C_A(t)) \\ 0 &= f_A(C(t), C_A(t)), \end{aligned}$$

where $C(t) = C(t,0)$ is the concentration of one or more chemical species, $f_R(\cdot)$ is the function describing the concentration change rate, $C_A(t)$ is a vector of algebraic variables and $f_A(\cdot)$ is the algebraic function which describes the mass-balance relation [150].

Reaction Models in Water Systems

In most of the research, single-species reaction dynamics are considered. Let $C(t)$ correspond to the concentration of a single substance within a finite water volume. Some typical reaction dynamics are:

- No reactions $\frac{dC(t)}{dt} = 0$, e.g. for fluoride
- Linear decay $\frac{dC(t)}{dt} = -\kappa C(t)$, $\kappa > 0$, e.g. for radioactive materials
- Linear growth $\frac{dC(t)}{dt} = \kappa C(t)$, $\kappa > 0$, e.g. for trihalomethanes

Linear decay in specific, is commonly used for modeling chlorine dynamics, even though the dynamics are more complex since they are coupled with the concentration of other substances reacting with chlorine. In the next subsection we present the dynamics of chorine, one of the most common chemical substances used for disinfection in water distribution systems.

Chlorine Reaction Modeling

Chlorine is commonly used as a water disinfectant, due to its ability to deactivate various pathogen substances; in addition it has low cost and it is easy to store, transport and use [175]. Throughout the water distribution network, a detectable chlorine residual is required so that the various mirco-organisms and chemical agents are below certain thresholds set by the World Health Organization and governments (e.g. at a minimum of $0.2\ \frac{mg}{L}$) [189].

When chlorine is injected into water (e.g. as gas), it produces hypochlorous acid (HOCl) and hypochlorite ion (ClO^-), which react with natural organic matter floating in water or residing on the pipe/tank walls, disinfecting the drinking water [17]. Chlorine reacts with organic compounds and other substances naturally present in the water flowing within the distribution networks and at the pipe walls.

The actual chlorine reaction dynamics in most of the cases are not known, and as a result, empirical models are utilized [85]. A common assumption in water research literature is that chlorine dynamics are first-order linear, such that $\frac{dC(t)}{dt} = -\kappa C(t)$, where $C(t)$ is the chlorine concentration within a finite water volume and $\kappa > 0$ is the reaction rate coefficient, which depends on the bulk reaction coefficient (initial water quality), the wall reaction coefficient (pipe material) and the mass transfer coefficient (chlorine transfer from bulk water to pipe walls) [142].

In practice, the chlorine reaction rate κ in some water volume is calculated off-line by using pipe condition information, and by measuring the concentration of a water sample in a bottle,

at various time instances; since the dynamics are considered to be linear, the slope of the log-graph of the normalized concentration measured at each time step indicates an appropriate value for κ. It is important to note that this decay rate can be affected by exogenous parameters, such as temperature [142].

In order to provide a more accurate mathematical model of the chlorine dynamics in drinking water, a number of empirical studies have been conducted [21,41,129,175]. Various chlorine reaction dynamics have been considered in research in addition to the first-order linear model, such as:

- the i-th power order, $\frac{dC(t)}{dt} = -\kappa C^i(t)$,
- the first-order with a stable component $\frac{dC(t)}{dt} = -\kappa(C(t)-\underline{C})$, where $\underline{C}$ is a concentration lower bound,
- the parallel first-order model, $\frac{dC(t)}{dt} = -\kappa a_K C(t) - \kappa_0(1-a_K)C(t)$, where $0 \leq a_K \leq 1$ is the percentage of the reacting chlorine concentration which decays with rate κ (fast reaction), whereas the remaining concentration decays with rate κ_0 (slow reaction).

In some studies, the parallel first-order model was shown to better capture the chlorine dynamics [175], although the first-order linear in some studies adequately described the actual dynamics [129].

The above dynamics, however, do not consider explicitly the actual chemical reaction dynamics of chlorine with organic matter. Following an analytical methodology, the chlorine dynamics can be expressed as $\frac{dC(t)}{dt} = -\kappa C(t) - \kappa_0 C^2(t)$, where reaction rates κ, κ_0 depend on the stoichiometry constants and the initial conditions of both chlorine and the reacting substance; in some studies this model was found to capture the dynamics of chlorine in drinking water with more accuracy than other models [21,85].

More comprehensive models have been proposed, based on the chemical characteristics for the reaction dynamics and the disinfection by-products in [109]. In addition, models describing chlorine reactions with contaminants (such as sodium arsenite and organophosphate) have been proposed [51,164,194].

2.3.3 Advection-Reaction Dynamics

In this section we formulate the dynamic advection-reaction equations in pipes and tanks, which describe a water distribution system.

Quality Modeling in Pipes

By coupling the advection and reaction dynamics, we formulate a non-homogeneous equation to describe the concentration change in time and space within a certain pipe, such that

$$\frac{\partial C(z,t)}{\partial t} + \frac{Q(t)}{a_P}\frac{\partial C(z,t)}{\partial z} = f_R(C(z,t)), \tag{2.4}$$

where $C(z,t)$ is the substance concentration in water at continuous time t and at distance z along a certain pipe, with water flow $Q(t)$ and a cross-sectional area of the pipe a_P; $f_R(\cdot)$ is the concentration change rate due to reactions.

To solve the advection-reaction equation numerically, we can use various methods, such as the single finite-difference *unsplit* method, or the *franctional step* method, which solves separately the advection and the reaction dynamics [101].

To illustrate the unsplit method, let's assume that the forward Euler method is used for discretization and that the reaction dynamics are described by linear decay (2.4); therefore the i-th finite volume state is given by

$$x_i(k+1) = x_i(k) - \lambda(k)\left[x_i(k) - x_{i-1}(k)\right] - \kappa x_i(k)\Delta t,$$

where $\kappa > 0$ and $0 < \lambda(k) < 1$.

For the fractional step method, (2.4) is segmented into two parts,

$$\begin{aligned}\frac{\partial C(z,t)}{\partial t} + \frac{Q(t)}{a_P}\frac{\partial C(z,t)}{\partial z} &= 0,\\ \frac{dC(z,t)}{dt} &= -\kappa C(z,t);\end{aligned}$$

by using the second-order Runge-Kutta method for the standard differential equation and the Lax-Wendroff scheme for discretizing advection, the i-th finite volume is given by

$$\begin{aligned}x_i(k+1) = \Bigg[&\frac{\lambda(k)}{2}[1+\lambda(k)]x_{i-1}(k) + [1-\lambda(k)^2]x_i(k) -\\ &-\frac{\lambda(k)}{2}[1-\lambda(k)]x_{i+1}(k)\Bigg](1-\kappa\Delta t + \frac{1}{2}\kappa^2\Delta t^2).\end{aligned}$$

In the advection-reaction algorithm implemented in EPANET 2.0, at each discrete time the reactions are performed to compute the new concentrations within each water segment; then advection of the segments is performed [142].

Quality Modeling in Water Tanks

At least three types of tank models are considered in tank water quality modeling [142]:

- the continuous stirred-tank reactor (CSTR) model, at which the chemicals are perfectly mixed and uniformly spread;
- the plug-flow reactor model, at which there is no mixing of water between the different water parcels assumed to travel along the flow in the tank;
- the two-compartments mixing model, at which the tank is segmented into two perfectly mixed compartments.

The continuous stirred-tank reactor model is considered a reasonable assumption for various tanks [142, 145].

Let $C(t)$ be the concentration in a water storage tank, which supplies water to a water distribution network. The water which fills the tank may be supplied from within the water distribution network, or it may be supplied from the treatment facilities through the transport system. For the CSTR, the quality concentration dynamics are given by

$$\frac{d}{dt}(V(t)C(t)) = Q_T^+(t)C_T^+(t) - Q_T^-(t)C(t) + f_R(C(t)),$$

where $V(t)$ is the tank's volume, $C_T^+(t)$ is the substance concentration of the water which flows into the tank, $Q_T^+(t)$ is the tank inflow, $Q_T^-(t)$ is the tank outflow, and $f_R(C(t))$ a reaction term, such that if linear decay with reaction rate $\kappa > 0$ is considered, $f_R(C(t)) = -\kappa C(t)$.

By using the forward difference scheme, the volume state $v(k) = V(k\Delta t)$ at discrete time k is given by

$$v(k+1) = v(k) + \left(q_T^+(k) - q_T^-(k)\right)\Delta t,$$

where $q_T^+(k) = Q_T^+(k\Delta t)$ and $q_T^-(k) = Q_T^-(k\Delta t)$. Let $x(k) = C(k\Delta t)$ correspond to the water quality in the tank and let $x^+(k) = C_T^+(k\Delta t)$ be the water quality of the water going into the tank with inflow $q_T^+(k)$, as depicted in Fig. 2.3. The tank water quality dynamics are given by

$$\begin{aligned} x(k+1) &= \frac{\left(q_T^+(k)x^+(k) - q_T^-(k)x(k)\right)\Delta t + v(k)x(k) - \kappa\Delta t x(k)}{v(k+1)} \\ &= \left[\frac{q_T^+(k)\Delta t}{v(k+1)}\right]x^+(k) + \left[\frac{v(k) - (q_T^-(k) - \kappa)\Delta t}{v(k+1)}\right]x(k) \end{aligned}$$

2.4 Hydraulic and Quality Control

The feedback control problem in water systems can be defined as the problem of computing at each discrete time k, the input vector $u(k)$, for the pumps, valves as well as the concentration of the injected disinfectant at each booster station, so that the measured hydraulic and

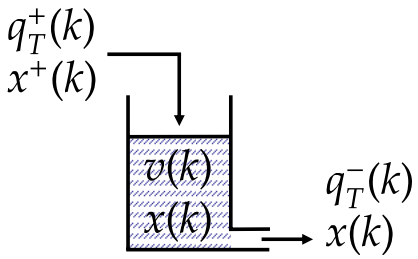


Figure 2.3: A simple tank schematic; the inflow, outflow, volume and concentrations are given in discrete time.

quality parameters in the output vector $y(k)$ follow the reference signal vector $r(k)$ specified for safe operation, computed using the $f_u(\cdot)$ controller function.

In practice, due to the complex interdependencies between hydraulic and quality dynamics, the design of the hydraulic controller is typically considered independently from the quality controller [127, 182]. A schematic of the controllers is depicted in Fig. 2.4; the water distribution system is driven by unknown consumer demands which exhibit certain periodicity, and uncertainty; at certain locations within the distribution network, on-line flow and pressure sensors are installed and monitored through SCADA, as well as some quality sensors measuring various chemical parameters. In addition, manual sampling for laboratory examination is performed at certain locations. A hydraulic and a quality controller take the measurement outputs into consideration, as well as the desired hydraulic and quality specifications and any constraints, and compute an input signal which regulates the flows and pressures at valves and pumps, as well as concentration at the disinfectant boosters.

2.5 Hydraulic Faults

Hydraulic faults may correspond to leakages within the water distribution network or at tanks, to pipe bursts, to blocked pipes or to malfunctioning pumps and valves. In addition, we may consider as a hydraulic fault the unauthorized back-flow in the network using a pump for injecting contaminants.

Water loss may be due to a variety of reasons, such as leaks, theft or unauthorized use and faulty water meters. The largest portion of the water lost is due to leaks or breaks [178]. A break is an abrupt fault needing immediate action, and it is usually easy to isolate the location of the problem. On the other hand, leaks due to cracks at pipes, tanks or loose fittings can remain unnoticeable, are difficult to isolate and may cause significant water losses and escaping revenues. Some of these problems are prompted by the deterioration of the infrastructure,

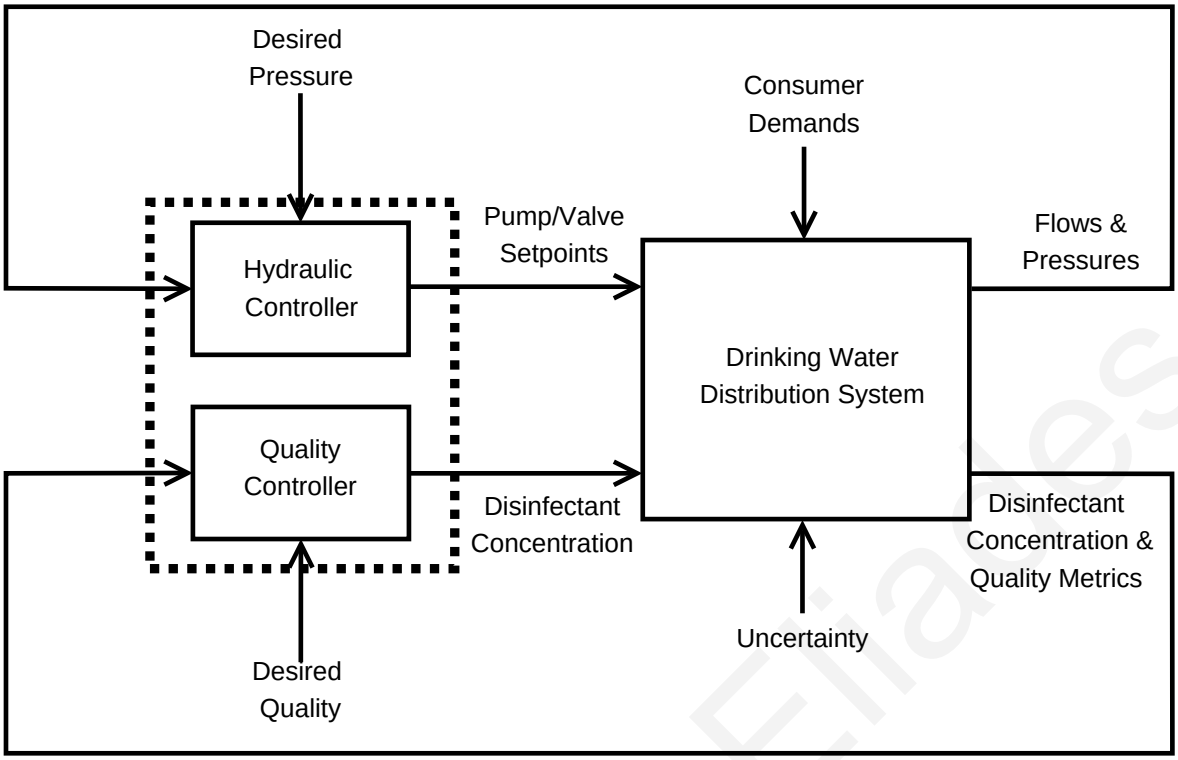


Figure 2.4: Controller for water distribution networks. Hydraulic and quality control are decoupled, but may exchange some information. The control objective is to regulate pressures and water quality so that they are within the desired bounds specified by the regulations.

mainly due to age, or by high pressures. Water loss imposes an economic burden on the water utilities while reducing water supplies; furthermore it may cause quality faults. For example a crack in a pipe under certain circumstances can be the entry location for contaminants (e.g. organic matter).

Mathematical models which describe the leakage flow with respect to the pressure at the leakage location have been proposed in various empirical studies [74,75,177]. Let $\phi_h(k;h(k))$ be the hydraulic leakage fault, i.e. the flow due to leakage measured in $\frac{m^3}{hr}$, occurring at a node with head $h(k)$, at time k. Hydraulic faults can be modelled mathematically as

$$\phi_h(k;h(k)) = a_D[f_l(h(k))]^{a_E}, \tag{2.5}$$

where $a_D > 0$ is a discharge coefficient, $a_E \in [0.5, 2.5]$ is an exponent term which depends on the leakage type and $f_l : \mathbb{R} \mapsto \mathbb{R}$ is an unknown function which maps the measured hydraulic head to the pressure at the leakage location (the pressure at the location where the leakage has occurred is usually not measured). In this model, both the discharge coefficient and the exponent term are unknown. However, empirical studies has shown that the exponent for small background leaks is $a_E \approx 1.5$, for larger leaks in plastic pipes is $a_E \geq 1.5$ and for larger leaks in metal pipes is $a_E \approx 0.5$ [98]. In this work we consider the leakage fault model

in simulating realistic leakage faults.

When a demand-driven hydraulic model is used, the leakage can be assumed as an additional time-varying demand, proportional to the corresponding nodal pressure/head; for the j-th node, the demand with a leakage fault is modelled as

$$d_j(k) = d_j^*(k) + \phi_h(k; h_j(k)),$$

where $d_j^*(k)$ is the real consumption demand, and $h_j(k)$ the nodal head. For modelling purposes, leakages which occur within a pipe are assigned as outflow from one of the adjusted nodes, which may be selected randomly.

2.6 Quality Faults

Quality faults may occur due to the contamination of water by certain substances, usually chemical, biological or radioactive, which travel along the flow, and they may exhibit decay or growth dynamics. A contaminant substance can be injected into a network at any point by connecting a pump and forcing the outflow direction to reverse. The contaminant travels within the network, following the path of the carrier. Digestion of the contaminated water by consumers may affect the health of the served population; in addition, use of contaminated water in industrial production may cause economic losses.

Consider a water distribution network composed of pipes, junctions and water storage tanks. The topology of this network can be represented as a graph with edges corresponding to pipes, and N_m nodes corresponding to junctions and water storage tanks. For modeling purposes, each pipe in the network is *a priori* virtually segmented into a number of finite volume cells. Let N_n be the total number of all nodes and finite volume cells considered in the network. Let $x_i(k)$ denote the average concentration of a certain contaminant at discrete time k, for $i \in \{1, ..., N_m, ..., N_n\}$. The vector $x(k) = [x_1(k), ..., x_{N_m}(k), ..., x_{N_n}(k)]^\top$ is the state of the contaminant concentration dynamics. Let $\mathcal{V}$ be the set of all node indices, such that $\mathcal{V} = \{1, 2, ..., N_m\}$.

The advection-reaction equations [101] describing the propagation of a contaminant in a water distribution network can be expressed in a state-space formulation:

$$\begin{aligned} x(k+1) &= A(k)x(k) + f_R(x(k)) + F\phi(k), \\ y_c(k) &= Cx(k) \end{aligned} \tag{2.6}$$

where $A(k)$ is an $N_n \times N_n$ matrix which characterizes the advection dynamics and captures the network topology, and $f_R : \mathbb{R}^{N_n} \mapsto \mathbb{R}^{N_n}$ is a function which describes the reaction dynamics

of the contaminant. For N_m possible injection locations (i.e., at the nodes), let F be an $N_n \times N_m$ matrix describing the locations of the injected contaminant. The function $\phi : \mathbb{N}^+ \mapsto \mathbb{R}^{N_m}$ describes the change in the contaminant concentration due to a substance injection. The output vector $y_c(k) \in \mathbb{R}^{M_s}$ corresponds to the state measurements, which are monitored using M_s online sensors. C is a binary output matrix, $C \in \{0,1\}^{M_s \times N_n}$, such that the element (i,j) is $C_{(i,j)} = 1$ when the i-th quality sensor measures the j-th state, and $C_{(i,j)} = 0$ when there is no quality sensor installed.

We define a finite set $\mathcal{E}$ of fault-location matrices E^j, $j = \{1, ..., 2^{N_m}\}$, given by

$$\mathcal{E} = \left\{ \begin{bmatrix} E^j_{(1,1)} & & 0 \\ & \ddots & \\ 0 & & E^j_{(N_m,N_m)} \end{bmatrix} \mid E^j_{(i,i)} \in \{0,1\}, j = \{1, ..., 2^{N_m}\} \right\}, \tag{2.7}$$

where $E_{(i,i)} = 1$ corresponds to the case when a contaminant is injected at the i-th node $i \in \mathcal{V}$. For the i-th fault-location matrix $E^i \in \mathcal{E}$, we define the injected contaminant location matrix F given in (2.6) as $F = [E^i \mid \mathbf{0}]^\top$, where $\mathbf{0}$ is an $(N_n - N_m) \times N_m$ zero matrix; F is of dimension $N_n \times N_m$.

The function $\phi(k) = [\phi_1(k), ..., \phi_{N_m}(k)]^\top$ corresponds to the signals of the injected contaminant concentrations. These have a certain start time and duration, and are non-negative. The function $\phi(k)$ can be represented through N_z linearly parameterized basis functions $\zeta(k) = [\zeta_1(k), ..., \zeta_{N_z}(k)]^\top$, such as pulses or radial-basis functions. Therefore, $\phi(k)$ is expressed as

$$\phi(k) = \Theta\zeta(k), \tag{2.8}$$

where $\Theta \in \mathbb{R}^{N_m \times N_z}$. The (i,j) element of Θ, denoted as $\Theta_{(i,j)}$, represents the amplitude of the basis function $\zeta_j(k)$ which is added to the state $x_i(k)$. Hydraulic dynamics are considered as approximately periodic (e.g. with a daily or weekly period) due to the periodic nature of consumer water demands. The basis functions are used to break up one hydraulic period into N_z time segments with possible overlaps, as in the case of radial basis functions. The motivation behind the use of a linearly parameterized form of the fault function, is that it simplifies the process of computing a finite set of fault parameter matrices, either through grid sampling or otherwise. This will be useful during the solution methodology for sensor placement (see Chapter 3).

From a practical viewpoint, the contaminant injection is measured in terms of the injected contaminant mass per unit time ($\frac{mg}{min}$), while the state-space formulation is described in terms of contaminant concentration ($\frac{mg}{L}$). The fault function $\phi_i(k)$ affecting the i-th node can be expressed as a fraction of a contaminant mass injection rate ($\frac{mg}{min}$) over the nodal inflows ($\frac{L}{min}$).

2.7 Simulation Examples

2.7.1 Example of Advection Dynamics

To illustrate the construction of the set of discrete dynamic equations describing the advection dynamics, we consider the simple network in Fig. 2.5.

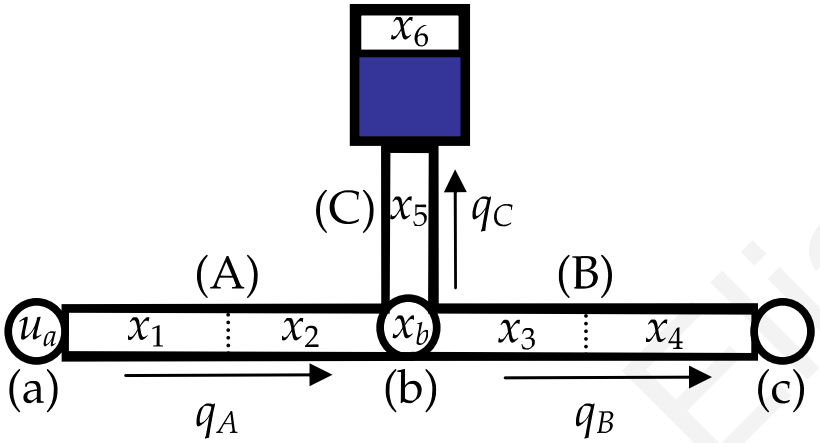


Figure 2.5: A simple water distribution network with three pipes, three junctions and one tank. Concentration is controlled at junction (a).

Junction node (a) delivers water to pipe (A); the water arrives at junction node (b), where it is delivered to pipes (B) and (C). Water from pipe (C) is used to fill the water tank, and water from pipe (B) is consumed at node (c). At node (a), the concentration signal $u_a(k)$ is controlled. Let $q_A(k)$, $q_B(k)$ and $q_C(k)$ correspond to the flows in pipes (A), (B) and (C) respectively, and let $x_b(k)$ correspond to the substance concentration at junction (b).

We segment offline the network into finite volumes, such that pipes (A) and (B) are comprised of two volume cells, while pipe (C) with one. Let $x_i(k)$ be the substance concentration within the i-th finite volume cell, at discrete time k, for $i \in \{1, ..., 5\}$; $x_6(k)$ is the concentration at the tank.

By using the *leapfrog* scheme, the following set of equations is constructed:

$$
\begin{aligned}
x_1(k+1) &= x_1(k) - \lambda_A(k)(x_2(k) - u_a(k)) \\
x_2(k+1) &= x_2(k) - \lambda_A(k)(x_2^+(k) - x_1(k)) \\
x_3(k+1) &= x_3(k) - \lambda_B(k)(x_4(k) - x_b(k)) \\
x_4(k+1) &= x_4(k) - \lambda_B(k)(x_4^+(k) - x_3(k)) \\
x_5(k+1) &= x_5(k) - \lambda_C(k)(x_5^+(k) - x_b(k));
\end{aligned}
$$

where $\lambda_A(k) = \frac{q_A(k)}{a_P}$, $\lambda_B(k) = \frac{q_B(k)}{a_P}$ and $\lambda_C(k) = \frac{q_C(k)}{a_P}$, and a_P is the cross-sectional area of pipes (A), (B) and (C). The concentration at the ghost cells is given by

$$x_2^+(k) = x_2(k)$$
$$x_4^+(k) = x_4(k)$$
$$x_5^+(k) = x_5(k)$$

and the concentration at junction (b) is $x_b(k) = x_2(k)$.

The concentration in the tank is given by

$$\begin{aligned} x_6(k+1) &= \frac{q_C(k)\Delta t}{v(k) + q_C(k)\Delta t} x_5(k) + \frac{v(k)}{v(k) + q_C(k)\Delta t} x_6(k) \\ &= a_1(k)x_5(k) + a_2(k)x_6(k), \end{aligned}$$

where $a_1(k) = \frac{q_C(k)\Delta t}{v(k)+q_C(k)\Delta t}$ and $a_2(k) = \frac{v(k)}{v(k)+q_C(k)\Delta t}$. The set of equations can be written in a state-space formulation

$$x(k+1) = A(k)x(k) + B(k)u(k)$$

where $x(k) = [x_1(k), ..., x_6(k)]^\top$ is the state vector of finite volume concentrations, the state matrix $A(k)$ is given by

$$A(k) = \begin{bmatrix} 1 & -\lambda_A(k) & 0 & 0 & 0 & 0 & 0 \\ \lambda_A(k) & 1-\lambda_A(k) & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_B(k) & 1 & -\lambda_B(k) & 0 & 0 & 0 \\ 0 & 0 & \lambda_B(k) & 1-\lambda_B(k) & 0 & 0 & 0 \\ 0 & \lambda_C(k) & 0 & 0 & 1-\lambda_C(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_1(k) & a_2(k) \end{bmatrix},$$

the input vector is $u(k) = u_a(k)$, and the input matrix is $B(k) = [\lambda_A(k), 0, ..., 0]^\top$.

2.7.2 Example of Quality Modeling

To illustrate the quality equations, consider the network depicted in Fig. 2.6; the node elevations are $\{800, 710, 700, 700, 695\}$ ft for the reservoir and the 4 junctions; the pipe parameters are given in Table 2.1. By computing the velocity in each pipe, we are able to compute the travel time and the pipe segmentations, with respect to $\Delta t = 11$ min (the shortest travel time), as given in Table 2.1. Therefore, the lengths of the finite volumes are $\Delta z = \{1000, 500, 1000, 333.33, 142.86\}$ ft, and $\lambda = \{1.00, 0.89, 0.56, 0.93, 094\}$. In total, 23 states

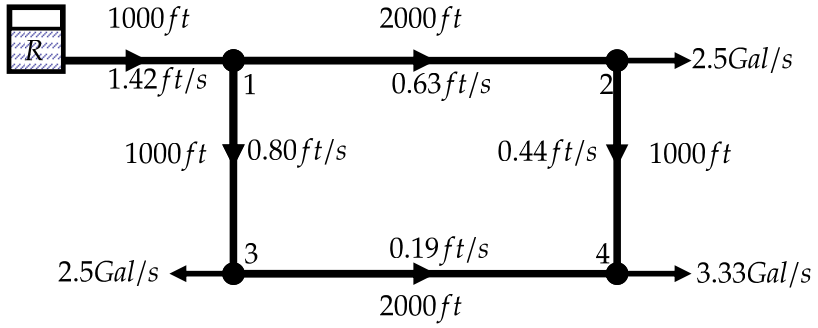


Figure 2.6: A simple water distribution network of one reservoir and four nodes.

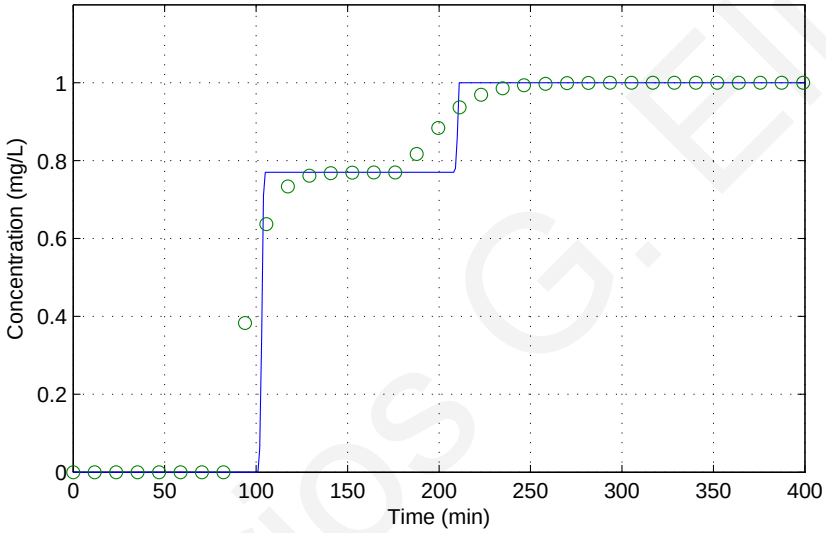


Figure 2.7: Comparison of exact solution (blue line) and numerical approximation (circles) using the forward Euler scheme, at each discrete time step.

Edge	Length (ft)	Diameter (ft)	Roughness	Velocity (ft/s)	Travel Time (min)	Finite Volumes
(R,1)	1000	12	100	1.42	11.74	1
(1,2)	2000	14	100	0.63	52.91	4
(1,3)	1000	10	100	0.8	20.83	1
(2,4)	1000	12	100	0.44	37.88	3
(3,4)	2000	10	100	0.19	175.44	14

Table 2.1: Pipe Characteristics

are constructed. At reservoir, a substance with constant concentration of $1\ \frac{mg}{L}$ is injected into the network. By using the simple forward Euler discretization scheme, we construct 23 state-space equations; for the i-th state in the j-th pipe, $x_i(k+1) = x_i(k) - \lambda_j(x_i(k) - x_{i-1}(k))$, and in the case of boundary pipe locations, the junction concentrations are computed using (2.3). Figure 2.7 illustrates the actual and the approximated concentration at node 4, by solving the state-space equations. The approximation could be improved by using a numerical approximation with higher accuracy, or by considering a smaller time-step.

2.8 Sensor and Communication Faults

In addition to hydraulic and quality faults, sensor and communication faults may occur in a water distribution system. Faults affecting sensors e.g. due to low batteries or calibration errors, may provide measurements with substantial errors to the utility operator as well as to the monitoring algorithms, which might cause errors in the calculation of various system parameters and in extend to miscalculate the control actions. Moreover, communication faults caused e.g. by the interruption of service by the telecommunication provider, may cause the loss or delay in receiving measurements from sensors, as well as controlling the actuators, thus interfering with the control process.

In Section 3.2 we propose a maximum redundancy quality sensor placement scheme, for increasing fault tolerance due to a possible quality sensor failure.

2.9 Impact Dynamics

When a contamination fault occurs, the contaminant is propagated in the network and eventually it may reach the customers who will consume the outflow water. Let $\mathcal{W} \subseteq \mathcal{V}$ be the set of $N_w = |\mathcal{W}|$ node indices corresponding to locations which outflow water to consumers based on demand requests. For each *demand node* $w_i \in \mathcal{W}$, $i = 1, ..., N_w$, an impact value can be computed at each time step. The impact due to a contamination fault can be expressed using epidemiological terms, e.g. how many people are affected; or using economic terms, such as the cost of productivity loss. Another impact measure which can be considered is the consumed volume of contaminated water which exceeds a certain concentration threshold [123].

In general, the impact of a fault depends on the volume and contaminant concentration of the contaminated water consumed. This can be described by a dynamic equation, as shown

below. Let $\xi(k) \in \mathbb{R}^{N_w}$ be the impact state vector which describes the "damage" caused at each demand node, at discrete time k, after a contaminant has been injected somewhere in the network. For $w_i \in \mathcal{W}$, a state-space representation of the impact dynamics is given by

$$\xi_i(k+1) = \xi_i(k) + f_\xi(x_{w_i}(k), d_{w_i}(k)), \tag{2.9}$$

$$\psi(k) = f_\psi(\xi(k)) \tag{2.10}$$

where $d_{w_i}(k)$ is the outflow demand (in $\frac{m^3}{s}$) at demand node w_i, and $f_\xi : \mathbb{R}\times\mathbb{R} \mapsto \mathbb{R}^+$ is a non-negative function, which characterizes the impact increase at each time step. Furthermore, let $f_\psi : \mathbb{R}^{N_w} \mapsto \mathbb{R}$ be a metric of the overall impact, $f_\psi(\xi(k))$, which characterizes the total impact, or "damage", which has been caused by a certain contamination fault. It is sometimes useful to compute the impact dynamics of N_ξ metrics, in order to optimize sensor placement with respect to various "damage" types.

To give some insight into the type of impact metrics which can be taken into consideration, we formulate in a state-space representation two metrics, the quantity of polluted water consumed, and the number of people affected [123]. It is worth noting that the proposed problem formulation can be easily extended to consider other type of impact metrics.

2.9.1 Contaminated Water Consumption Volume

Let $\xi_i(k)$ be the volume of consumed contaminated water at demand node $w_i \in \mathcal{W}$ until time k. This can be computed at each time-step by using the discrete state-space equation (2.10), where $f_\xi : \mathbb{R}\times\mathbb{R} \mapsto \mathbb{R}$ computes the volume of polluted water which is consumed. In this example,

$$f_\xi(x_{w_i}(k), d_{w_i}(k)) = \begin{cases} d_{w_i}(k)\Delta t & \text{if } x_{w_i}(k) > \overline{C}_p \\ 0 & \text{if } x_{w_i}(k) \leq \overline{C}_p \end{cases} \tag{2.11}$$

where $\overline{C}_p$ is the concentration threshold above which water is considered as polluted. The overall impact function f_ψ is given by the aggregated volume of consumed contaminated water, i.e.,

$$f_\psi(\xi(k)) = \sum_{i=1}^{N_w} \xi_i(k). \tag{2.12}$$

2.9.2 Population Infected

More advanced impact metrics use epidemiological models. Let $\xi_i(k)$ be the mass of contaminant ingested by the consumers of demand node $w_i \in \mathcal{W}$ until time k. Following the

formulation in [120], the increase of the contaminant ingested mass $\xi_i(k)$ at time-step k, is given by

$$f_\xi(x_{w_i}(k), d_{w_i}(k)) = a_\gamma \; x_{w_i}(k) \frac{d_{w_i}(k)}{\bar{d}_{w_i}} \Delta t, \tag{2.13}$$

where a_γ is a constant characterizing the volumetric rate of water consumption per person, and $\bar{d}_{w_i}$ is the daily average demand outflow at demand node w_i.

For the calculation of the overall impact $f_\psi(\cdot)$ (in this case, the total number of people infected), a log-normal dose-response model can be considered [39]. Let a_Ψ be the *median infective dose*, i.e. the average dosage (in $\frac{mg}{kg}$) for which there is 50% probability that a person will become infected or symptomatic. In addition, let a_W be the average body mass (in kg). Moreover the number of consumers served by demand node w_i is given by $\frac{\bar{d}_{w_i}}{a_\mu}$, where a_μ is the average daily consumption per person. From the empirical studies mentioned above, the number of people affected at a demand node depends on the number of consumers served, multiplied with the probability of one person ingesting enough contaminant mass so that to become infected or symptomatic. For $\Phi(\cdot) \in [0, 1]$ the standard normal cumulative distribution function and a_S the probit (probability unit) slope parameter, the overall impact is given by

$$f_\psi(\xi(k)) = \sum_{i=1}^{N_w} \frac{\bar{d}_{w_i}}{a_\mu} \Phi\left(a_S \log_{10}\left(\frac{\xi_i(k)}{a_W a_\Psi}\right)\right). \tag{2.14}$$

2.10 Concluding Remarks

In this chapter, background information was presented, regarding the hydraulic and quality models which are typically used to describe water distribution networks. The advection dynamics were discussed, as well as some numerical approximation schemes to solve the partial differential equations describing the advection dynamics. Various reaction dynamic models in water systems were presented as well as the chlorine reaction dynamics, which is injected into the distribution network for disinfection purposes. The coupled advection and reaction dynamics were then formulated, for pipes and water tanks, in continuous and discrete time. Next, the problem of hydraulic and quality control was introduced. Hydraulic, quality and sensor/communication faults were discussed. Regarding the quality faults, a state-space representation was formulated, to capture the advection-reaction dynamics under the influence of a contamination fault, which can be analyzed using a linearly parameterized model. Finally,

the general contamination impact dynamics were discussed and a state-space representation was formulated.

The contribution of this work is formulation of the advection-reaction and quality fault dynamics, coupled with the impact dynamics, into a state-space representation, which is used as a mathematical framework for the optimization and control problems addressed in this thesis. Part of the work in this chapter has been published and presented in peer-reviewed journals [58, 61, 66] and conferences [59, 60, 63, 64].

Chapter 3

Quality Fault Monitoring

In this chapter we provide background information on the quality sensor placement and manual sampling scheduling problem in water distribution systems. The problem of quality sensor placement for maximum redundancy, when the steady-state hydraulic dynamics can be expressed as a graph, is formulated and solved as an integer quadratic optimization program. This work is extended when the system dynamics are available, for the sensor placement and manual sampling scheduling problem, as well as for the problem for identifying the most neuralgic locations in a water distribution network. We formulate an optimization problem in which one or more risk-objective functions are minimized to compute the locations where to install quality sensors, or, in the case of the manual sampling problem, compute the locations and time where and when to perform manual sampling. The methodology is demonstrated on an illustrative network, on a benchmark network, as well as on part of a water distribution network in Cyprus.

3.1 Background

The problem of where to place facilities[1] or sensors, in order to keep certain objectives and constraints satisfied within a network, has been examined in various research disciplines such as operational research [162] and automatic control [8].

The "Set Covering" method was one of the first mathematical formulations of the problem and it has since been applied in various fields, such as facility location [162]. According to this approach, an integer optimization program is formulated in order to determine a set of nodes from a topological graph to install facilities, so that all the remaining nodes are next

[1] In this context, *facility* can be a public service such as the police or a fire station.

to at least one facility. A related approach is the "Maximal Covering" formulation described in [40] for the calculation of a set of nodes which maximize the population served in an area within a certain distance. A similar formulation was considered in [100], which selected the locations to install water quality sensors in drinking water distribution systems, so that the largest volume of water consumed was examined. The authors proposed a scenario-based approach which segmented a day into time-periods, corresponding to different flow patterns, and the optimization was solved for all scenarios simultaneously. Following this formulation, aiming at solving bigger networks, other researchers utilized heuristics [91] and genetic algorithms [2], a mixed-integer problem formulation was presented in [16]. A multi-objective weighted-sum extension of [100] was presented in [82] which considered certain physical network characteristics and time-delays.

A mathematical formulation suitable for the security issues related to the location selection, is the "p-median" [81], with the objective of minimizing the "maximum distance" of a facility. A similar formulation was examined in [11, 15] for water distribution systems. By considering a number of contamination scenarios and their impacts, the authors formulated a mathematical program to minimize the average "contamination impact". The formulation was extended to take into consideration imperfect sensors [14]. The "p-median" formulation was further examined in a stochastic framework [152]. A multi-objective extension was examined in [186]; however for solving the mixed-integer optimization program, significant computational power was required. In [19] the authors proposed a modification of this formulation, to determine locations for monitoring disinfection byproducts.

Within the water resources management community, the design competition of the *Battle of the Water Sensor Networks* (BWSN), in 2006, instigated significant research interest and discussion on security issues of water distribution systems. The task was to find sets of locations to install quality sensors using two real benchmark networks, so that a number of objectives are optimized under various fault scenarios, and most of the participant research groups formulated a multi-objective integer optimization program [90, 123, 133]. According to the BWSN instructions, only the average impact of observed faults was considered; furthermore, different fault scenarios could have different solutions. A methodology was proposed in [187] to measure the contamination risk, considering contamination detection failures and its consequences. Some single and multiple objective methodologies for sensor placement have recently been proposed [6, 52], and in [83], some of the issues related to sensor placement strategies were reviewed.

When online quality sensors are not available, or do not cover part of the distribution

network, the standard approach is for the water utility to perform manual water sampling for quality analysis. Water quality may be checked at a few nodes within a network, and for a few times during the day. Sampling locations may be selected by water utility personnel in an arbitrary fashion, based on their own experience, which could be subjective; it may be otherwise chosen using certain regulatory requirements, depending on consumer distributions and historical data. In practice, due to the large-scale nature of water distribution networks, as well as the partial knowledge of the time-varying, hydraulic and quality dynamics, it is difficult to optimally identify the best locations and times to conduct manual sampling, or install on-line quality sensors. In addition, each node in the network has certain characteristics, such as the outflow water pattern and the number of customers, which makes the selection problem non-trivial.

The problem of manual sampling was discussed in [12], where the authors examined the problem of scheduling manual sampling for contaminant detection. They proposed a mixed integer program for the calculation of the sampling route; i.e. the location and time to take samples, while considering certain real conditions such as utility working hours, the time required for sampling and the traveling time between nodes. Computational difficulties in solving large-scale problems by using this formulation were expressed in [11,12] for the sensor placement and manual sampling scheduling problem respectively.

3.2 Quality Sensor Placement for Maximum Redundancy

As a first step towards security monitoring, we address the problem of selecting locations in the network for on-line quality monitoring. Although water providers are obliged to monitor the quality in the distribution network which they operate, there exist no guidelines for deciding where to measure water quality within the network. In [100], a "Maximal Covering" method was proposed. Water arriving at a node has originated from another node, unless the node is a water source; the idea is to compute "how much" each node contributes to the other nodes and to neglect those nodes with small contribution. Their work proposed a framework for water network-based approaches, which solves the problem of maximizing the monitored volume of water when the number of on-line sensors is *a priori* known.

We examine a variation of the problem posed by [100]: find the minimum number of on-line monitoring sensors required for covering the distribution network, and the locations at which to install them, while maximizing the coverage redundancy. Redundancy is important in the case of a sensor failure, so that the water quality is still monitored with as much

information as possible.

As a solution to this problem, we formulate an integer quadratic optimization problem, which uses the binary *coverage matrix*, as presented in [100] and a theoretical upper bound of the maximum redundancy possible for that coverage matrix. The formulation and solution methodology presented are generic, thus they may as well be applied to other application domains, such as in power systems for placement of Phasor Measurement Units (PMU) [33, 34].

3.2.1 Problem Formulation

The general problem is defined as to find the minimum number of sensors required to monitor a certain network, and from all the possible sensor placements, find those solutions which have the biggest redundancy.

Let $H \in \{0,1\}^{N_h \times N_h}$ be the binary coverage matrix of N_h nodes, such that $H_{(i,j)} = 1$ if a certain minimum percentage of the water volume arriving at the j-th node was supplied through the i-th node; otherwise, $H_{(i,j)} = 0$. Let R_s and R_c be diagonal matrices whose non-zero elements correspond to the significance metric and installation cost of each node, respectively.

Let $\bar{r}$ be the vector of the maximum redundancy possible, i.e. the maximum number of node junctions which are able to measure a sufficient water quantity originating from a certain node; this can be computed through $\bar{r} = H \cdot \mathbf{1}$, where $\mathbf{1} = [1, 1, ..., 1]^\top$.

The optimization objective function $J(\chi)$ is formulated as

$$\begin{aligned} J(\chi) &= a_R(\bar{r} - H\chi)^\top R_s(\bar{r} - H\chi) + \chi^\top R_c \chi \\ &= \frac{1}{2}\chi^\top(2a_R H^\top R_s H + 2R_c)\chi + (-2a_R \bar{r}^\top R_s H)\chi + a_R \bar{r}^\top R_s \bar{r}, \end{aligned}$$

where $\chi \in \{0,1\}^{N_h}$ is a binary vector, indicating the sensor locations, such that $\chi_i = 1$ if a sensor is placed and $\chi_i = 0$ if not, for $i = 1, ..., N_h$; in addition, $a_R = (\bar{r}^\top R_s \bar{r})^{-1}$ is a normalizing factor. In addition to maximizing the coverage redundancy, the formulation takes into consideration weights corresponding to installation costs, as well as some significance metric (such as the number of customers), at a certain node. When all nodes are equally significant and have the same installation costs, R_s and R_c are equivalent to the identity matrix I. The optimization problem is to find a binary vector χ, such that it minimizes the objective function $J(\chi)$.

The intuition behind the objective function is that a minimization algorithm will minimize the term $\chi^\top R_c \chi$, which corresponds to the integer number of sensors used (or the cost), and

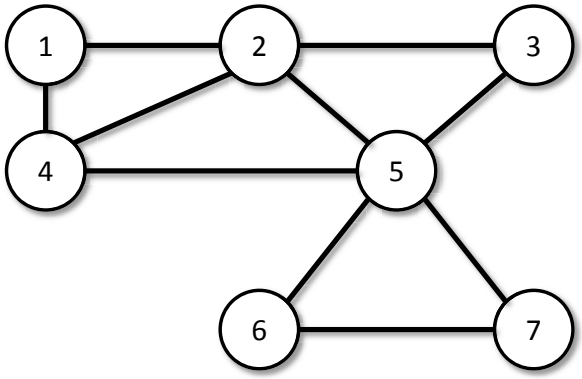


Figure 3.1: Coverage graph from a water distribution network example with seven nodes.

then it will minimize the term $a_R(\bar{r} - H\chi)^\top R_s(\bar{r} - H\chi)$, a normalized distance metric of the redundancy given by $H\chi$ to the maximum redundancy possible $\bar{r}$ (when a sensor is installed at each node).

The constant term in the optimization function can be neglected, and the optimization problem can be formulated as an integer quadratic program, such that

$$\min_{\chi \in \{0,1\}^{N_h}} \quad \tfrac{1}{2}\chi^\top(2a_R H^\top R_s H + 2R_c)\chi - 2a_R \bar{r}^\top R_s H\chi \tag{3.1}$$

$$\text{subject to} \quad H\chi \geq \underline{r}, \tag{3.2}$$

where $\underline{r} \in \mathbb{N}^{N_h}$ is the minimum redundancy requirement vector. The optimization can be solved using standard Mixed Integer Quadratic Programming algorithms. Overall, this formulation will give priority to redundant coverage to nodes with the greater significance.

3.2.2 Example

Figure 3.1 is a undirected graph of a network with seven nodes. A sensor at each node can monitor its neighboring node, i.e. '1' can monitor '1', '4' and '2'. The coverage matrix H of this graph can be derived from this graph, where each row is the node where a sensor can be placed and each column corresponds to the nodes monitored by the sensor; H is therefore given by

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Therefore the maximum redundancy vector is $\bar{r} = [3, 5, 3, 4, 6, 3, 3]^\top$. We consider that the significance and installation-cost matrices, R_s, R_c are equivalent to the identity matrix I. The redundancy normalization factor is $a_R = \frac{1}{113}$, and $\underline{r} = [1, ..., 1]^\top$. By using a mixed integer optimization algorithm, the computed solution is

$$\chi = [0, 1, 0, 0, 1, 0, 0]^\top,$$

which corresponds to installing quality sensors at nodes '2' and '5'. With this solution, all nodes are monitored at least by one sensor, while at the same time four nodes are monitored by both sensors. Note that other solutions are feasible, but with lower redundancies (e.g. three redundant measurements when installing sensors at nodes '4' and '5' or two redundant measurements when installing sensors at nodes '1' and '5').

Remark: The limitation of the maximum redundancy algorithm is that it takes into consideration the average values of the hydraulic dynamics; in practice, hydraulic and quality dynamics are time-varying, with long time delays, and in addition, contamination faults have impact dynamics; these are not considered in the formulation presented in this section. In the rest of the chapter, we formulate and solve the sensor placement and manual sampling location selection problem, in a security-oriented framework.

3.3 Sensor Placement Design Methodology

Given a water network topology with certain known hydraulic dynamics (subject to uncertainties), let $\mathcal{V}_s \subseteq \mathcal{V}$ be the set of N_s candidate "sensing nodes", i.e. indices of feasible locations for placing sensors, such that $\mathcal{V}_s = \{c_1, ..., c_{N_s}\}$. The sensor placement problem is described as finding the set of location indices $\mathcal{Y} \subset \mathcal{V}_s$ (where $|\mathcal{Y}| = M_s$ and $M_s < N_s$ represents the number of available sensors), such that one or more objectives are optimized.

An approach for solving the sensor placement problem is to consider a number of representative contamination scenarios within a specific time-period, while the water distribution

network is assumed to be operating under normal conditions. A scenario is comprised of two parts, the parameter-fault matrix which characterizes the injection, and the time-delays in shutting down the system. The simulated impact data for each scenario can then be used to determine the effectiveness of various sensor placement schemes within an optimization algorithm.

As defined in (2.7), $\mathcal{E}$ represents the set of fault-location matrices which describe the locations of contaminant injection. In the case of fault identification or fault isolation, the objective is to find the true fault-location matrix among the set $\mathcal{E}$. However, in the case of the sensor placement problem, where all locations are possible injection nodes of contaminants, we can assume that $\mathcal{E}$ is equivalent to the identity matrix I. This allows a fault to occur at any of the states corresponding to the index set $\mathcal{V}$.

The fault function $\phi(k)$ as defined in (2.8), is characterized by the parameter matrix $\Theta \in \mathbb{R}^{N_m \times N_z}$, and the N_z-dimensional basis function vector $\zeta(k)$. In practice, the parameters in matrix Θ which describe the fault, are bounded due to various physical constraints. Let $\boldsymbol{Q}^*$ be the set of parameter matrices Θ defined as

$$\boldsymbol{Q}^* = \left\{ \Theta \in \mathbb{R}^{N_m \times N_z} \mid |\Theta_{(i,j)}| \leq \overline{\Theta}_{(i,j)} \right\}, \tag{3.3}$$

where $\overline{\Theta}_{(i,j)}$ are pre-specified bounds on each element of Θ.

By sampling within $\boldsymbol{Q}^*$, we can construct a finite set $\boldsymbol{Q} \subset \boldsymbol{Q}^*$ of $N_q = |\boldsymbol{Q}|$ parameter matrices, such that $\boldsymbol{Q} = \{\Theta^{(1)}, ..., \Theta^{(N_q)}\}$ and $\Theta^{(i)} \in \boldsymbol{Q}$. The finite set $\boldsymbol{Q}$ can be obtained through grid sampling; however, this approach does not guarantee that the selected samples will be the most representative faults. More advanced optimization techniques can be utilized in order to choose a finite number of representative fault scenarios, which can take into consideration the network topology and other *a priori* information about potential faults.

As shown in the previous chapter, the state-space equation (2.6) describes mathematically the advection-reaction dynamics of a water distribution network, and (2.10) couples those dynamics with a fault's impact dynamics. Thus, for a certain contamination fault, we can compute at each time-step its propagation in the network, as well as its impact at each node and overall in the network.

In order to evaluate the impact of a certain fault, we first need to compute the time it takes for the fault to propagate up to a certain sensing node. Let $x_{c_j}(k)$ be the contaminant concentration at a candidate sensing node $c_j \in \mathcal{V}_s$. Let $T(\epsilon_c)$ be the propagation-time matrix of size $N_q \times N_s$ computed for a concentration detection threshold ϵ_c; its (i, j) element, $T_{(i,j)}(\epsilon_c)$, corresponds to the time taken for a certain fault with parameter matrix $\Theta^{(i)} \in \boldsymbol{Q}$ to be first

detected at node $c_j \in \mathcal{V}_s$, when it is above a specific concentration $\epsilon_c \in \mathbb{R}^+$. From a practical perspective, the detection time $T_{(i,j)}(\epsilon_c)$ is bounded by a finite constant, because even if it is not detected using sensor technologies, it will eventually be detected, after some time instance T_{max}, by other methods, such as water utility manual quality sampling, or in more severe cases, by customer complaints and hospitalizations.

Let $k_0^{(i)}$ be the discrete time when the i-th contamination fault starts, given by

$$k_0^{(i)} = \min_{k \in \mathbb{N}^+} \{k \mid \|\Theta^{(i)} \zeta(k)\| > 0\};$$

the starting-time vector is $k_0 = [k_0^{(1)}, ..., k_0^{(N_q)}]^\top$. In addition, the discrete time $k_d^{(i,j)}(\epsilon_c)$ when a fault with parameter matrix $\Theta^{(i)}$ is first detected at node $c_j \in \mathcal{V}_s$ with concentration threshold ϵ_c, is given by

$$k_d^{(i,j)}(\epsilon_c) = \min_{k \in [k_0^{(i)}, k_0^{(i)} + T_{max}]} \{k \mid x_{c_j}(k) > \epsilon_c\}.$$

From the above, we define the fault propagation-time as $T_{(i,j)}(\epsilon_c) = k_d^{(i,j)}(\epsilon_c) - k_0^{(i)}$. If the condition $x_{c_j}(k) > \epsilon_c$ is not satisfied for $k \leq k_0^{(i)} + T_{max}$, then $T_{(i,j)}(\epsilon_c) = T_{max}$.

After a fault is detected, some additional time may be required, in order to shut down the system. Time-delays can be different for each sensing node, i.e. due to distances from water utility headquarters. Moreover, various time-delay schemes can be examined.

We define $\mathcal{D} = \{\delta^{(1)}, ..., \delta^{(N_k)}\}$ as a set of the N_k discrete time delays considered, in order to stop the system after a contamination fault has been detected. From the set of N_q fault parameters and the set of N_k time-delays, we define the set $\mathcal{H}$ as the finite set of $N_v = |\mathcal{H}|$ fault scenarios, given by

$$\mathcal{H} = \{(i, j) \mid i \in \mathcal{Q}, j \in \mathcal{D}\},$$

i.e. for each scenario in set $\mathcal{H} = \{h^{(1)}, ..., h^{(N_v)}\}$ corresponds to a 2-tuple, comprised of a fault parameter matrix and a time-delay.

We define $K(\epsilon_c)$ as the stopping-time matrix of size $N_v \times N_s$, whose element (i, j) is the discrete time at which the system shuts down, after a contaminant detection (above the detection threshold ϵ_c) has occurred at node $c_j \in \mathcal{V}_s$ during fault scenario $h^{(i)} \in \mathcal{H}$. For $h^{(i)} = (\Theta^{(i_1)}, \delta^{(i_2)})$, when detection occurs at node $c_j \in \mathcal{V}_s$, then $K_{(i,j)}(\epsilon_c) = k_0^{(i_1)} + T_{(i_1,j)}(\epsilon_c) + \delta^{i_2}$.

Thus, the impact vector $\xi(K_{(i,j)}(\epsilon_c))$ corresponds to the "damage" caused by fault scenario $h^{(i)}$, until the water system is shut down, as a result of a contaminant detection at node c_j. The overall impact of a fault scenario on the network can be computed by measuring the output

signal at time $K_{(i,j)}(\epsilon_c)$, i.e.

$$\Omega_{(i,j)} = \psi(K_{(i,j)}(\epsilon_c)). \tag{3.4}$$

There can be more than one overall-impact matrix Ω, when several overall-impact metrics are considered. In general, for N_ξ different impact metrics considered, we compute N_ξ overall-impact matrices $\Omega^{(i)}$, $i = 1, ..., N_\xi$, which belong to the set $\mathcal{J} = \{\Omega^{(1)}, .., \Omega^{(N_\xi)}\}$.

3.3.1 Securing Neuralgic Locations

The first part of a security scheme in water distribution networks is to determine the locations in the network which could be considered as "high-risk" for contaminant injection, so that proper action is taken in order to secure them through physical means, e.g. with closed-circuit television (CCTV) systems.

We first need to decide on a representative impact metric; from its corresponding overall-impact matrix $\Omega \in \mathcal{J}$, we will calculate the maximum scenario impact $\overline{\Omega}$, given by

$$\overline{\Omega} = \max_{i=1,...,N_v} \max_{j=1,...,N_s} \Omega_{(i,j)},$$

where $\Omega_{(i,j)}$ is the (i, j)-th element of Ω. From this, we can compute the set of the worst-case scenarios; for example, the set of the top 5% fault indices is computed by

$$\mathcal{Y}_0 = \left\{ i \mid \max_{j=1,...,N_s} \Omega_{(i,j)} \geq 0.95\, \overline{\Omega},\ i \in \{1, ..., N_v\} \right\}.$$

In this equation, $\mathcal{Y}_0$ is the set of scenario indices with correspond to high-impact faults (i.e. with impact greater than 95% of the worst-case fault). Each fault scenario corresponds to a location in the network and a detection time-delay; therefore, the set of all fault scenario locations corresponding to the fault scenario indices in $\mathcal{Y}_0$, indicates one or more geographical regions in the network, where severe contaminations are possible.

These locations can be considered as neuralgic, and need to be physically secured. However, due to the nature of the water distribution networks, certain contamination faults may override physical security measures. Hence, it is imperative to use an extra layer of protection by installing a number of sensors at different locations in the network.

3.3.2 Solution Methodology

Consider the task of selecting M_s out of N_s node locations to install quality sensors. Let $\mathcal{L}_s = \{1, ..., N_s\}$ be the set of the candidate sensing node indices. The set of all solution com-

binations $\mathbb{X}$ is given by

$$\mathbb{X} = \left\{ \mathcal{X} \mid \mathcal{X} \in \mathcal{L}_s^{M_s} \right\}.$$

For the i-th risk objective function $F_i : \mathbb{X} \mapsto \mathbb{R}$, the single-objective optimization problem is formulated as

$$\mathcal{Y} = \underset{\mathcal{X} \in \mathbb{X}}{\operatorname{argmin}} \{F_i(\mathcal{X}; \Omega)\}, \tag{3.5}$$

where $\Omega \in \mathcal{J}$ is the overall-impact matrix defined in (3.4). In the single-objective case, the solution set $\mathcal{Y}$ is a set of indices which corresponds to certain locations where sensors can be installed.

3.3.3 Risk Objective Functions

In order to address the problem of security, it is vital to have an understanding on what "risk" is and how it can be quantified. Risk has been examined in many fields and especially in the financial and operational research literature. For instance, in finance, risk is defined as "...the possibility that an unpredictable future event will result in a financial loss, with the consequence that [...] the institution will not meet some specified financial criteria." [140].

In financial practice, the most commonly used risk-objective is the "Value-at-Risk" (VaR), which represents the maximum loss with a certain confidence level over a time period. This metric, however, ignores the worst scenarios, which may be crucial in the case of intentional water contamination. All in all, risk management provides useful tools and insights for the problem of security in critical infrastructure systems.

The proposed formulation can utilize various types of objective metrics; in this thesis we consider the following objectives: a) the Average impact, b) the Maximum impact and c) the Conditional Value-at-Risk. The optimization problem is to minimize one or more of these objectives.

In computing the risk objective functions, it is useful to define the scenario index set $C = \{1, ..., N_v\}$. In practice, certain scenarios could be neglected in the optimization, depending on each overall-impact matrix. Thus, for each overall-impact matrix $\Omega^{(i)}$ corresponds to a set $C_i \subseteq C$.

Average Impact: The average impact metric is suitable for the optimization of reliability, when contaminant injection can occur at any node with equal probability. This metric, however, has limitations when considering the security framework, since it fails to take into sufficient consideration rare faults with extreme consequences. For a certain overall-impact

matrix Ω and for a specific set of location indices $\mathcal{X} \in \mathbb{X}$, the overall impact across all faults is given by

$$F_1(\mathcal{X}) = \frac{1}{|C|} \sum_{i \in C} \min_{j \in \mathcal{X}} \Omega_{(i,j)}. \tag{3.6}$$

Maximum Impact: The maximum impact metric is used to reduce the effect of the most extreme fault, in terms of causing the most damage. This metric is useful from a security perspective; on the other hand, it fails to take into consideration the fault frequency distribution, and in specific, the frequency of extreme faults. The maximum impact metric is given by

$$F_2(\mathcal{X}) = \max_{i \in C} \min_{j \in \mathcal{X}} \Omega_{(i,j)}. \tag{3.7}$$

Conditional Value-at-Risk: The Conditional Value-at-Risk (CVaR) metric, which is frequently used in finance optimization applications, is defined in [167] as the average "loss" for the worst a_C% scenarios. In the present work, "loss" corresponds to the overall impact. This metric is quite suitable for the water security problem, since it can be used to minimize extreme contamination faults while at the same time taking into account the frequency of extreme faults. For this metric, a decision maker needs to specify the parameter $a_C \in (0, 100)$, so that only fault impacts above $(1 - \frac{a_C}{100})F_2(\mathcal{X})$ are considered. Let $C_0 \subset C$ be the set of extreme-fault indices, given by

$$C_0 = \{i \mid \min_{j \in \mathcal{X}} \Omega_{(i,j)} \geq (1 - \frac{a_C}{100})F_2(\mathcal{X}), i \in C\}.$$

Therefore, the average tail-impact metric is given by

$$F_3(\mathcal{X}) = \frac{1}{|C_0|} \sum_{i \in C_0} \min_{j \in \mathcal{X}} \Omega_{(i,j)}. \tag{3.8}$$

3.3.4 Multiple Objectives Optimization

The different objectives presented in the previous subsection will in general yield different solutions. Often, it is desirable to compute a set of "good" solutions which satisfy an N_f number of objectives instead of a single one. Minimizing one objective function may result in maximizing others; it is thus not possible to find one optimal solution which satisfies all objectives at the same time. It is possible, however, to find a set of solutions, laying on a Pareto front, where each solution is no worse than the other. The multi-objective optimization problem can be formulated as follows:

$$\mathcal{Y} = \operatorname*{argmin}_{\mathcal{X} \in \mathbb{X}} \left\{F^{(1)}(\mathcal{X}), ..., F^{(N_f)}(\mathcal{X})\right\}, \tag{3.9}$$

where $F^{(l)}(\mathcal{X}) = F_i(\mathcal{X}; \Omega^{(j)})$, for $l \in \{1, ..., N_f\}$, $j \in \{1, ..., N_\xi\}$ and $i \in \{1, 2, 3\}$. The solution set $\mathcal{Y}$ is comprised of the sensor location solutions on the Pareto front.

A feasible solution $\mathcal{X}$ is called Pareto optimal, if for a set of objectives $\{F^{(1)}, ..., F^{(N_f)}\}$, there exists no other feasible solution $\mathcal{X}^*$ such that $F^{(i)}(\mathcal{X}^*) \leq F^{(i)}(\mathcal{X})$ with $F^{(j)}(\mathcal{X}^*) < F^{(j)}(\mathcal{X})$ for at least one j. Therefore, a solution is Pareto optimal, if there is no other feasible solution which would reduce some objective function, without simultaneously causing an increase in at least one other objective function [139, p.779].

One of the most popular solution methodologies for multiobjective optimization problems is to assign a scalar weight for each cost function and calculate their weighted sum, so that the problem is reduced into a single-objective optimization. However, the computed solution might not belong to the set of Pareto front solutions; in addition, weight assignment is susceptible to biases by the decision maker.

3.3.5 Proposed Optimization Methods

A trivial solution to the sensor placement problem would be to compute the objective functions for all possible node combinations and then find the Pareto front. Although this method is complete and optimal (i.e. it guarantees to find all the desired solutions), it is unsuitable for large solution spaces, as it has large memory requirements ; therefore, it is impossible to evaluate all possible combinations. For instance, in the problem of where to install five sensors in a network with 12 500 nodes, there are 2.5×10^{18} combinations. It is therefore inefficient to search for solutions without using any extra information.

In the artificial intelligence literature, searching can be represented as a tree-graph, with its root being the initial state, on which connected are the next choices (nodes), which themselves have the next choices connected on, and so on. The decision of which state to expand next is determined by the search strategy. Searching for a solution can be made more efficient by the use of problem specific knowledge beyond the definition of the problem [146]. The use of heuristics can reduce the cost of searching.

Members of the algorithmic family that use heuristics are, among others, best-first search and evolutionary/genetic algorithms [77]. In best-first search, the node that expands is the one that appears to be closer to the goal according to the evaluation function. This is not an optimal method, though it usually produces sufficiently good results in reasonable time and with small memory requirements. Genetic algorithms randomly create an initial population; their "best" individuals (with respect to the optimization function) "reproduce" and "mutate" for

a number of epochs, until they reach a solution consensus which satisfies certain constraints.

Search Optimization

Instead of searching to find the optimal Pareto solutions, a solution is to reduce the problem by searching for "good enough" solutions. By referring to "good enough" solutions, we accept the compromise that there may be better solutions than the ones found, in exchange for a significant reduction in computational time and space complexity.

A key goal is to develop a design methodology that does not search blindly in the solution space, but is instead guided iteratively towards the good sets of solutions, and prunes those that are likely of no interest. In this section we present one such approach, the *Iterative Deepening of Pareto Solutions Search* algorithm.

This algorithm was presented in the *Battle of the Sensor Networks* competition [123], where it managed to compute 11 nondominated solutions, which corresponded to the 7-th place out of the 14 teams which participated.

Consider the single-sensor problem, which can be trivially solved in a complete and optimal manner, in the sense that all the points in the Pareto front are identified. If decision makers were to place only one sensor, they would place it in one of those locations. If another sensor was to be placed, then we assume that the first sensor would not be removed from its installed location, therefore another location would be found, which in combination with the first, is on the "local" Pareto front. The locality of the solution originates from the fact that the location of the first sensor defines to a certain extent, the Pareto front for the two sensors optimization problem; if the first sensor is placed elsewhere, the optimization would yield a different Pareto front.

This procedure can be improved by searching the Pareto-front solutions in parallel. By this, in the first iteration, all the Pareto front solutions are stored in a list and instead of choosing one of those solutions, all of them are available and expanded to their next combinations. This means that all the one-sensor solutions are combined with all the rest of network nodes. The Pareto solutions are then calculated and the procedure is repeated until all sensors have been placed. The final solution of the algorithm is a set of Pareto front points, which may not necessarily be on the global Pareto front.

Another method for reducing the search space is by removing nodes that are obviously not useful, specifically nodes that have very small network and demand coverage. Moreover, nodes that are dead-ends, meaning that they have no consumption and the pipes that are connected

with them have no flows, may be removed, so that we avoid unnecessary computations. This can be attained with a threshold value in terms of network coverage, below which nodes are not considered.

Evolutionary Multi-objective Optimization

We consider utilizing a multi-objective evolutionary algorithm in which each objective is optimized in such a way that the algorithm computes solutions which are non-dominant to each other [50].

The procedure for solving the multi-objective placement of M_s sensors by using the algorithm described in [50] is as follows: Randomly select N_x solutions from the set of all feasible solutions $\mathbb{X}$, where $N_x \ll |\mathbb{X}|$, and construct a solution "parent" set $\mathcal{Y}_p \subset \mathbb{X}$. The values of the objective functions for each solution in the parent set $\mathcal{Y}_p$ are then computed. Next, the solutions are sorted according to their non-dominance and are assigned in Pareto ranks, based on which Pareto front they reside. A subset of the parent set is selected and an "offspring" population set is computed, by using the genetic operators of mutation and crossover, suitably modified so that only feasible solutions are generated. From the combined set of parent and offspring solutions, the elements are sorted according to non-dominance. A "crowding distance" metric is computed for each solution, expressing how close it is with its neighboring solutions. An N_x number of solutions is selected from the parent-offspring set, which will comprise the new parent set; this is achieved by selecting solutions with the highest Pareto rank, as well as by considering the crowding distance for better dispersion of the solutions on the Pareto front. The algorithm iterates for a certain number of epochs. The set of sensor placement solutions on the Pareto front computed within the last epoch are the solutions of the problem. The decision on which is the most suitable sensor placement solution among the computed solutions, can be made by a human operator.

3.3.6 Summary of Methodology

The proposed methodology for sensor placement by taking multiple objectives into consideration, can be summarized in the following steps:

I. Construct a model of the real water-distribution network, e.g. by using state-space equations or specialized software.

II. Choose suitable basis functions for the calculation of contamination fault functions, e.g. pulses or radial-basis.

III. Select one or more impact metrics, and the overall impact functions for the formulation of the state-space equations of the impact dynamics, e.g. number of people infected.

IV. Decide on the different types of faults to be considered.

V. From the set of all faults parameters, construct a subset of representative faults.

VI. Compute the propagation-time matrix T.

VII. For different time-delays, construct the different fault scenarios and compute the overall-impact matrix Ω.

VII. Specify the objectives considered for each impact metric, e.g. average or maximum of some impact metric.

IX. Formulate the optimization problem and solve it by using a suitable nonlinear integer optimization algorithm, e.g. using heuristic or multi-objective evolutionary algorithms.

This information is used to construct and solve the Sensor Placement problem, by constructing a mathematical program of some objective metrics (e.g. cost, population killed) with respect to some objective functions (e.g. mean, worst), some response times and with respect to some constraints. Finally, regret analysis is considered to compute the best sensor placement with respect to a large number of possible contamination scenarios.

In general, the methodology presented in this thesis is relevant to TEVA-SPOT, as it is based on a mathematical model of the water distribution network, contamination scenarios are considered and a single or multi-objective optimization problem is formulated. However, in the work examined in this thesis, we investigate the problem from the system theoretic point-of-view, and to provide a mathematical framework which is suitable for analysis.

3.3.7 Simulation Examples

In this section, we present two simulation examples to illustrate the formulation and the solution methodology for sensor placement in drinking water distribution networks. The first example is a rather simple network, which is used to derive valuable intuition on the proposed methodology, while the second example is a real-world drinking water distribution network with 129 nodes and time-varying hydraulic dynamics.

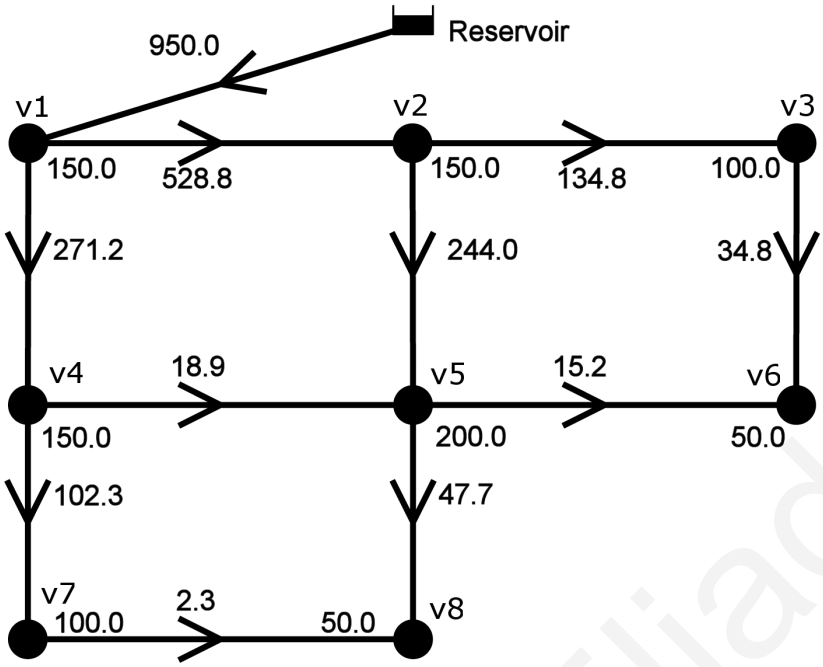


Figure 3.2: Water distribution network with one reservoir and eight demand nodes. Weights on nodes and arcs correspond to flows measured in $\frac{gal}{min}$; for example, for node v_1, $d_1(k) = 150\ \frac{gal}{min}$.

Illustrative Network

The objective in this example is to find the optimal locations in a simple water distribution network, depicted in Fig. 3.2, for the installation of M_s water quality sensors, so that the average and maximum risk-objectives are minimized over a set of representative faults.

As shown in Fig. 3.2, a reservoir supplies water to eight demand nodes, such that $\mathcal{V} = \{v_1, ..., v_8\}$; we consider that $\mathcal{V}_s \equiv \mathcal{V}$ and $\mathcal{W} \equiv \mathcal{V}$. The nodal weights correspond to the consumption outflow at a junction, and the arc weights correspond to the flow of water in a pipe connecting two nodes; flows are measured in US-gallons per minute ($\frac{gal}{min}$). Demand flows $d_i(k)$, $i \in \mathcal{W}$, are known from historical data, and pipe flows are computed using the EPANET hydraulic solver [142]. The structural characteristics such as pipe lengths and diameters, junction elevations and hydraulic coefficients are considered known.

It is assumed that: a) a contamination fault with a non-reactive contaminant can occur at any demand node; b) quality sensors measuring the contaminant concentration can be installed at any demand node; c) there is no time-delay for shutting down the network after a contaminant with concentration above $\epsilon_c = 0.1\ \frac{mg}{L}$ has been detected (i.e., $\mathcal{D} = \{0\ hr\}$); d) the hydraulic time-step is $\Delta t = 20\ min$.

We consider that the network is segmented into N_n finite volumes; the linear state-space

model of the contaminant propagation dynamics is

$$x(k+1) = Ax(k) + F\Theta^{(i)}\zeta(k),$$

where state $x(k)$ is the contaminant concentration (in $\frac{mg}{L}$), $x(0) = 0$; matrix A describes the advection dynamics, and $F = [I \mid \mathbf{0}]^\top$, where I is the 8×8 identity matrix and $\mathbf{0}$ a zero-matrix of size $(n-8) \times 8$. State $x_i(k)$ is the contaminant concentration for demand node v_i, for $i = \{1, ..., 8\}$.

The contamination fault $\Theta^{(i)}\zeta(k)$ is characterized by the i-th fault parameter matrix $\Theta^{(i)}$ and, in this example, the unit step function $\zeta(k)$ for which $\zeta(k) = 1$ for $k \geq 0$.

As impact metric, we consider the contaminated-water consumption volume, which can be expressed in a state-space formulation for the i-th demand node, as

$$\xi_i(k+1) \;=\; \xi_i(k) + f_\xi(x_i(k), d_i),$$

where $\xi(0) = 0$ and, for $\overline{C}_p = 0.5\ \frac{mg}{L}$, $f_\xi(x_i(k), d_i) = d_i \Delta t$ if $x_i(k) > 0.5\ \frac{mg}{L}$, otherwise $f_\xi(x_i(k), d_i) = 0$. The overall impact is computed through

$$f_\psi(\xi(k)) = \xi_1(k) + \xi_2(k) ... + \xi_8(k).$$

Representative faults are considered the ones describing a contaminant injection with concentration $1\ \frac{mg}{L}$ at a demand node. The finite set $\boldsymbol{Q}$ of the fault parameters considered is comprised of eight matrices, such that $\boldsymbol{Q} = \{\Theta^{(1)}, ..., \Theta^{(8)}\}$, which have been computed so that they simulate a contaminant injection with the desired concentration. Note that all faults are initiated at time $k = 0$, i.e. fault starting time is $k_0 = [0, ..., 0]^\top$.

A fault propagation-time $T_{(i,j)}(\epsilon_c)$ is the minimum time necessary for a fault that has occurred at the i-th node, to be detected by a sensor (i.e. above concentration ϵ_c) at the j-th node. An 8×8 propagation-time matrix $T(\epsilon_c)$ is constructed, by simulating each fault. For example, when a fault occurs at node v_4, detection at v_7 with concentration above $\epsilon_c = 0.1\ \frac{mg}{L}$ occurs after $T_{(4,7)}(0.1) = 140\ min$. We can verify this by considering that the pipe connecting v_4 with v_7 has $14\ 330\ gal$ of water, and this water flows with $102.3\ \frac{gal}{min}$. Therefore, a contaminant injected at node v_4 will reach node v_7 after $140\ min$.

Since there are no stopping time-delays in the system, the fault propagation-time matrix $T(\epsilon_c)$ and the stopping-time matrix $K(\epsilon_c)$ are equivalent, $K(\epsilon_c) = T(\epsilon_c)$. Therefore, the overall impact of fault i until the system stops due to detection at node j, is computed using $\Omega_{(i,j)} = \xi_1(K_{(i,j)}(\epsilon_c)) + \xi_2(K_{(i,j)}(\epsilon_c)) + ... + \xi_8(K_{(i,j)}(\epsilon_c))$. For example, consider a fault at node v_1 that is detected by node v_7 at $K_{(1,7)}(0.1) = 220\ min$. Before this time, the fault contaminates

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
average	110	73	77	91	61	142	91	86
max	307	171	171	307	166	435	307	201

Table 3.1: Average and maximum volume of contaminated water until it is detected at each node (in 10^3 *gal*).

nodes v_2 and v_4 at $K_{(1,2)}(0.1) = K_{(1,4)}(0.1) = 100$ *min*. The overall impact $\Omega_{(1,7)}$ is computed as

$$\begin{aligned}\Omega_{(1,7)} &= d_1 K_{(1,7)}(0.1) + d_2(K_{(1,7)}(0.1) - K_{(1,2)(0.1)}) + d_4(K_{(1,7)}(0.1) - K_{(1,4)}(0.1)) \\ &= 150 \cdot 220 + 2 \cdot 150 \cdot (220 - 100) \\ &= 69\ 000\ gal.\end{aligned}$$

Finally, by computing the overall impacts for each fault when detected by each possible sensor location, the overall-impact matrix Ω is constructed.

In this example we consider two different risk-objectives: the "average" impact F_1 and the "maximum" impact F_2. The problem is to find where to install M_s quality sensors from the set of all possible location combinations $\mathbb{X}$, so that one objective is minimized. The optimization program is formulated as in (3.5).

For a single sensor placement ($M_s = 1$) the problem can be solved by inspection, using Table 3.1; this table describes the average and maximum impacts (measured in 10^3 *gal*) of each column of the overall-impact matrix Ω. Thus, from the set of $|\mathbb{X}| = 8$ solutions, the optimal location for the placement of the sensor is at the fifth node, v_5, since at that location both the average impact ($61\ 000$ *gal*) and maximum impact ($166\ 000$ *gal*) of the eight faults considered are minimized. Due to the topology of the network, and to the small number of nodes, the optimal node location for both objectives is the same. In practice, for different solutions may correspond to different objectives.

For more sensors ($M_s > 1$) the optimization program can be solved using commercially available integer non-linear optimization software. We conclude the illustrative example with the solution for $M_s = 3$, for each objective: amongst $|\mathbb{X}| = \binom{8}{3} = 56$ solutions, $\{v_3, v_5, v_7\}$ corresponds to the optimal locations for both the average impact ($29\ 000$ *gal*) and the maximum impact ($69\ 000$ *gal*).

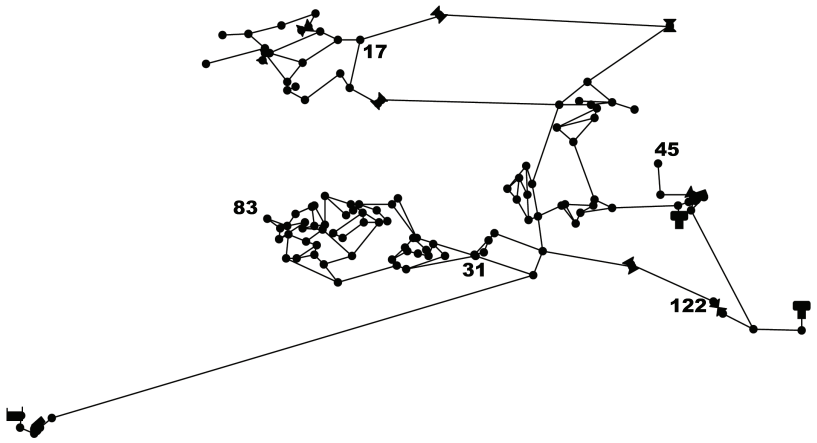


Figure 3.3: A real water distribution system with 129 nodes

Real Water Distribution Network Example

In this example, we demonstrate the solution methodology on a small-scale real water distribution network. The tasks are a) to examine the improvement in minimizing the maximum fault-impact by using different numbers of sensors $M_s \in \{1, ..., 10\}$, and b) to find solutions by using different numbers of sensors $M_s \in \{5, 10\}$ when two of the objective metrics are optimized simultaneously, the average contaminated water consumption volume objective and the CVaR of the population infected objective.

Figure 3.3 depicts one of the benchmark networks in the "Battle of the Water Sensor Networks" design competition [123]. This network is composed of 178 pipes connected to $N_m = 129$ nodes (126 junctions, two tanks and one reservoir). The structural characteristics are considered known. Each junction node is assigned with a daily average consumption volume as well as a discrete signal describing the rate of water consumption within 48 hours, with a 30-minute time step. These are assumed to describe the normal operation. The hydraulic dynamics are computed using the EPANET software [142]. For simplicity, no reactions were considered in this example.

The model of the real water distribution network can be expressed in a discrete state-space formulation as in (2.6), with a $\Delta t = 5$ *min* time step. Let the set of node indices $\mathcal{V}$ be given by $\mathcal{V} = \{1, ..., N_m\}$.

The fault function is formulated using N_z linear basis functions; in this example we consider these functions to have a rectangular shape and a 5-minute length. Since hydraulic dynamics are approximately periodic with period of $T_H = 24$ *hr*, we consider contamination faults which are initiated within one period, i.e. one day. We further consider faults with maximum duration of 24-hours; by using $N_z = 576$ basis function we can represent faults which

start within one period, and can finish at the end of the next period (i.e. the next day). The maximum contaminant propagation time is set to have a duration of $T_{max} = 1152$ time steps (i.e. 4 days), after which the fault is assumed to be detected by other means. The fault function is formulated using N_z linear basis functions; in this example we consider these functions to have a rectangular shape and a 5-minute length. Since hydraulic dynamics are approximately periodic with period of $T_H = 24\ hr$, we consider contamination faults which are initiated within one period, i.e. one day. We further consider faults with maximum duration of 24-hours; by using $N_z = 576$ basis function we can represent faults which start within one period, and can finish at the end of the next period (i.e. the next day). The maximum contaminant propagation time is set to have a duration of $T_{max} = 1152$ time steps (i.e. 4 days), after which the fault is assumed to be detected by other means.

Two impact metrics are considered: the population infected and the contaminated water consumption volume. For the impact state equation (2.13) we set the daily average amount of water consumed per person as $a_\gamma = 2\ \frac{L}{day}$; for the overall impact (2.14) we assume that the probit slope is $a_S = 0.34$, the average body mass is $a_W = 70\ kg$ and the contaminant dose is $a_\Psi = 41\ \frac{mg}{Kg}$ for a 50% probability of infection. The average per-person daily water consumption is $a_\mu = 300\ \frac{L}{day}$. Regarding the contaminated water consumption volume (2.11), we assume that the contamination threshold, above which impact is measured, is $\overline{C}_p = 0.3\ \frac{mg}{L}$.

Three different types of faults are considered: a) one-node contamination faults of 2-hour duration, b) one-node contamination faults of 10-hour duration, c) contamination faults of 2-hour duration which occur independently at two different nodes. These faults can be modeled using the rectangular basis functions.

According to the solution methodology, we need to construct a finite subset $\mathcal{Q}^* \subset \mathcal{Q}$ of fault-parameter matrices, such that $N_q = |\mathcal{Q}^*|$. For the one-node contaminations, a systematic sampling approach was applied, so that all possible starting times and injection locations were included; for the two-node contaminations, random sampling was used for the selection of the starting times and injection locations. In addition to constructing the set of representative faults, we construct the set $\mathcal{V}_s \subseteq \mathcal{V}$ of the feasible sensing locations. In the real network considered, seven "leaf nodes" from the set $\mathcal{V}$ with zero consumer demands could be considered as unfeasible sensing locations. By removing those nodes, the feasible set $\mathcal{V}_s$ of sensing locations has $N_s = |\mathcal{V}_s| = 122$ elements.

To compute the propagation-time matrix, we consider the following fault scenarios: all types of fault, as described above, with zero time-delay in shutting down the system, $\delta^{(1)} = 0\ hr$, and a 3-hour time-delay for the one-node contamination faults of 2-hour duration,

$\delta^{(2)} = 3\ hr$, thus $\mathcal{D} = \{0\ hr, 3\ hr\}$. In total, $N_v = 130700$ scenarios are constructed.

For these scenarios and the N_s candidate sensing locations, the propagation-time matrix $T(\epsilon_c)$ of size $N_q \times N_s$ is computed by simulating the hydraulic and advection dynamics for each fault separately, and finding the minimum detection time. The sensing threshold is $\epsilon_c = 0\ \frac{mg}{L}$, so that any non-zero contaminant concentration triggers a fault detection.

By using the impact dynamics describing the population infected and the contaminated water consumption volume, and by considering the time-delays according to each scenario, two overall-impact matrices $\mathcal{J} = \{\Omega^{(1)}, \Omega^{(2)}\}$ are computed, where $\Omega^{(1)}$ corresponds to the population infected and $\Omega^{(2)}$ to the contaminated water consumption volume.

Selecting the Number of Sensors

Water quality sensors are quite expensive and require frequent maintenance. As a result, a water utility would consider installing only a small number of them for monitoring the water distribution network. It is useful to measure the reduction of a certain impact metric, for different number of sensor schemes.

We formulate the single objective optimization problem as in (3.5) with the maximum impact objective (3.7), in order to minimize the maximum number of people infected, using $\Omega^{(1)}$. The problem is solved for 10 different cases, for $M_s \in \{1, ..., 10\}$ sensors. The corresponding solution space, $\mathbb{X}$, can be extremely large; e.g. for $N_s = 122$ feasible sensing nodes and $M_s = 10$ sensors, there can be $|\mathbb{X}| = \binom{122}{10} \approx 1.3782 \cdot 10^{14}$ solutions. Thus, evaluating the objective function for all the possible solutions in $\mathbb{X}$ has high computational cost.

We use a single-objective genetic algorithm to compute feasible solutions for this problem. The normalized maximum impact of the computed solutions for $M_s \in \{1, ..., 10\}$ sensors is depicted in Fig. 3.4. From these results we observe that four sensors manage to reduce the maximum impact by approximately 70%; using more sensors causes a small reduction on the maximum impact.

Multi-Objective Sensor Placement

Before solving the multi-objective sensor placement problem, which can be computationally expensive, we perform a pre-processing step, aiming at discarding the trivial fault-scenarios. From each overall-impact matrix $\Omega^{(1)}$ and $\Omega^{(2)}$, we find the maximum (worst-case) impact of each fault, and decide whether it is considered trivial or not. In this example, faults with maximum impact larger than 10% of the maximum overall impact are considered as

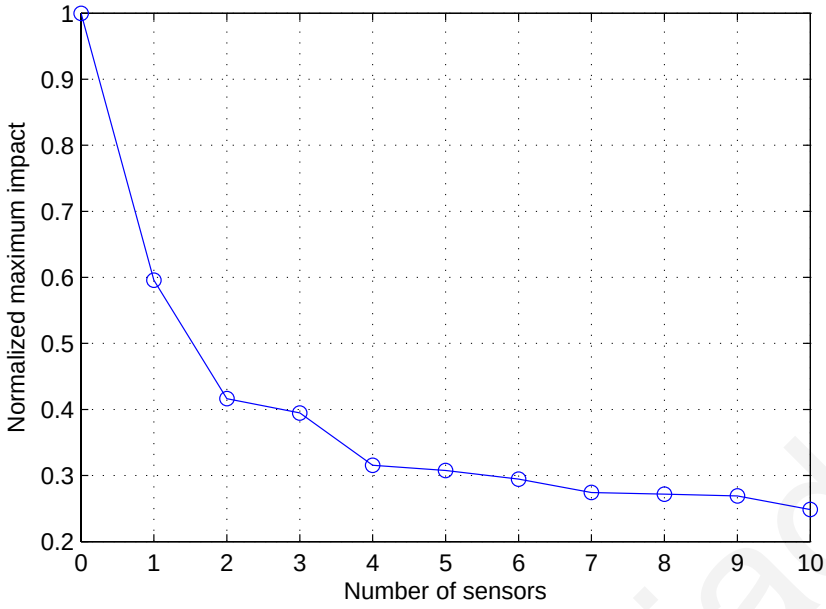


Figure 3.4: Normalized maximum impacts of optimal solutions with $M_s \in \{1, ..., 10\}$ sensors installed

non-trivial. As a result, we construct the two scenario index sets, C_1 and C_2, which correspond to the non-trivial scenarios. The histogram in Fig. 3.5a depicts the normalized number of people affected when no sensors are considered, for all scenarios which belong to the set C_1.

We consider the following two objective functions for the multi-objective case: the average contaminated water consumption volume, and the CVaR of the population infected (with $\alpha_C = 25$). The multi-objective optimization problem is given by

$$\mathcal{Y} = \underset{X \in \mathbb{X}}{\operatorname{argmin}} \{F_1(X; \Omega^{(2)}), F_3(X; \Omega^{(1)})\}.$$

The problem is solved for $M_s = 5$ and $M_s = 10$ sensors. The solution set $\mathcal{Y}$ is comprised of a number of solutions laying on the Pareto front computed.

In order to solve the optimization problem, we use the multi-objective evolutionary algorithm NSGA-II [50] which computes a set of solutions, which are well spread and lay near the real Pareto front. The algorithm was modified to accept nodal indices as inputs, and to return valid feasible solutions. For the evolutionary algorithm, $1\ 000$ epochs are iterated for, and for each epoch a population of $2\ 000$ solutions is built. 90% of these solutions "crossover" by randomly exchanging node indices, and 10% are "mutated" by randomly changing the index of a node in the solution.

Figure 3.6 depicts the computed 5-sensor and 10-sensor Pareto fronts. Each point in the graph corresponds to a solution, i.e. a set of node indices. From these Pareto sets, a decision

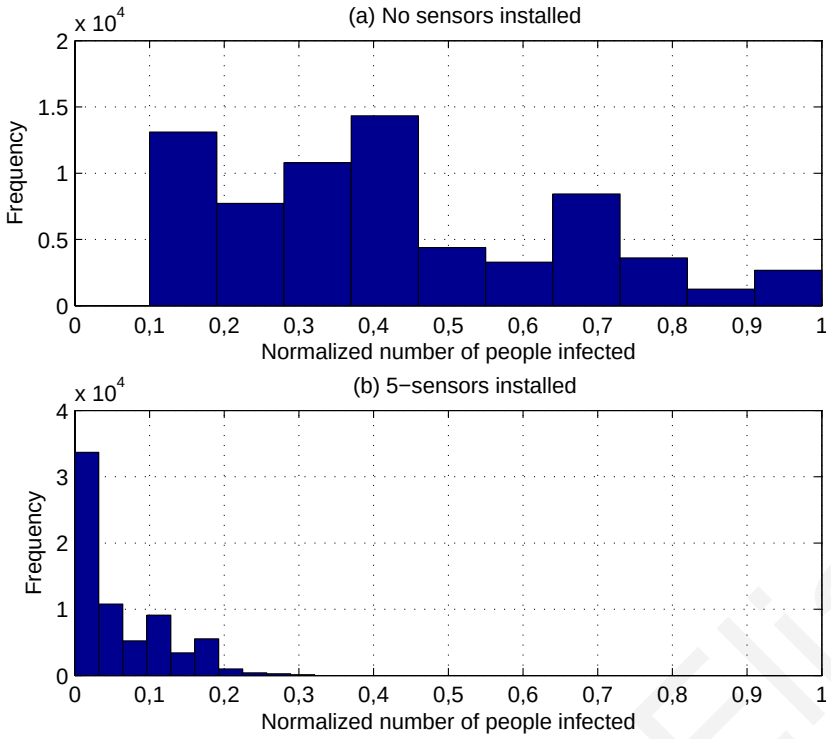


Figure 3.5: Histogram of the normalized number of people affected, for all scenarios in set C_1, (a) when no sensors are considered, and (b) when a Pareto solution with five sensors is considered.

maker should examine the optimal solutions, and through higher-level reasoning, choose the most suitable solution. To demonstrate the impact reduction, we arbitrary choose one solution from the 5-sensor set of Pareto solution that is relatively good on both objectives. The histogram in Fig. 3.5b illustrates the frequency of the scenarios considered with respect to the normalized impact. As expected, the impacts have been reduced significantly.

3.4 Manual Sampling Scheduling

The proposed formulation can be extended to take into consideration the problem of manual sampling scheduling. This is different from the problem of finding locations to install permanent sensors, because of the extra dimension of sampling time which needs to be taken into consideration. In this section we present a problem formulation and a solution methodology for selecting M_m out of N_b possible sampling locations and times, in order to minimize certain objectives.

We define $\mathcal{T}_m$ as the set of N_t discrete time instances when manual sampling can be performed within one day (e.g. during working hours). In addition $\mathcal{V}_m \subseteq \mathcal{V}$ is the set of indices corresponding to locations where manual sampling is feasible. We define $\mathcal{B}_m$ as the set of

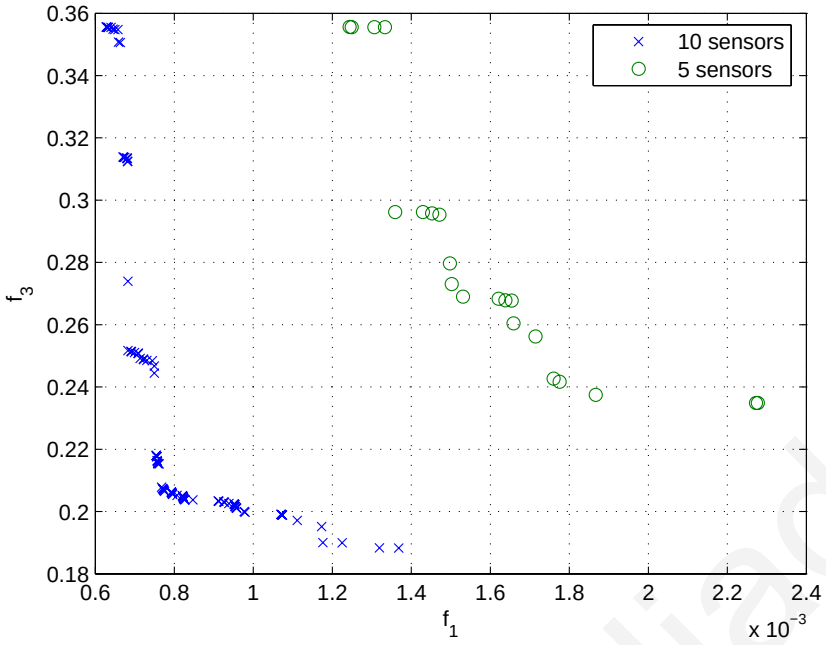


Figure 3.6: Two Pareto fronts computed for the 10-sensor and 5-sensor schemes. F_1 is the average contaminant water consumption volume and F_3 is the CVaR metric of the population infected.

2-tuples, corresponding to candidate sampling node indices and sampling times, such that

$$\mathcal{B}_m = \{(i, j) \mid i \in \mathcal{V}_m, j \in \mathcal{T}_m\}.$$

The set $\mathcal{B}_m = \{b^{(1)}, ..., b^{(N_b)}\}$ is comprised of N_b sampling 2-tuples. To compute the overall-impact matrix Ω we need to modify the algorithm presented in the previous section.

We define a discrete maximum time T_{max}, so that for any fault scenario, its impact will not increase after that time. We assume that $T_{max} = a_\lambda T_H$, where $a_\lambda \in \mathbb{N}^+$ is a period index and T_H is the discrete time duration of a one-day period. It is useful to define $\mathcal{T} = \{1, 2, ..., T_{max}\}$ as the set of time steps for which the system dynamics are simulated, as well as the period index set $\mathcal{T}_p = \{0, 1, ..., a_\lambda\}$. The discrete time $K_{(i,j)}(\epsilon_c)$ corresponds to the time when a contamination scenario i is detected and stopped due to detection at the j-th manual sampling or on-line sensor, at concentration above ϵ_c. For the i-th scenario with $h^{(i)} = (\Theta, \delta)$, then

$$K_{(i,j)}(\epsilon_c) = \begin{cases} \min\{\mathcal{G}_1\} + \delta & \text{if } \mathcal{G}_1 \notin \emptyset \text{ (Online Sensor)} \\ \min\{\mathcal{G}_2\} + \delta & \text{if } \mathcal{G}_2 \notin \emptyset \text{ (Manual Sampling)} \\ \min\{\mathcal{G}_3\} + T_{max} + \delta & \text{if } \mathcal{G}_1 \in \emptyset \text{ or } \mathcal{G}_2 \in \emptyset \end{cases}$$

where

$$\mathcal{G}_1 = \{k \mid x_{c_j}(k;\Theta) > \epsilon_c, k \in \mathcal{T}\}$$
$$\mathcal{G}_2 = \{lT_H + b_2^{(j)} \mid x_{c_j}(lT_H + b_2^{(j)};\Theta) > \epsilon_c, l \in \mathcal{T}_p\}$$
$$\mathcal{G}_3 = \{k \mid \|\Theta\zeta(k)\| > 0, k \in \mathcal{T}\}.$$

for which c_j is the j-th online quality sensor and $b_2^{(j)}$ is the second element in the sampling 2-tuple $b^{(j)}$, corresponding to the sampling time. Finally, the overall impact matrix can be computed as in (3.4).

3.4.1 Manual Sampling Scheduling for a Limassol DMA

In this section we illustrate the formulation and the solution methodology for the manual sampling scheduling problem in a real drinking water distribution network, depicted in Fig. 3.7, operating under realistic conditions. The network is comprised of 321 pipes connected to $N_m = 198$ junctions, 100 of which are used for water consumption. The main source of disinfected water is a storage tank which is located in the lower part of the figure. The structural characteristics are assumed to be known, i.e. pipe length, diameters and pipe roughness coefficients, node elevations and daily average consumption volume at each node. In addition, historical flow-data are provided, measured at the supply node, and are assumed to describe the normal operation over all nodes. We consider the set of nodes $\mathcal{V} = \{1, ..., N_m\}$ and a $\Delta t = 5$ minute time step for computing the hydraulic and quality dynamics; each 24-hour period is therefore comprised of $T_H = 288$ time steps.

There are $N_w = 100$ nodes in the network which outflow water to consumers; the set of demand nodes is $W = \{w_1, ..., w_{N_w}\}$, and $\mathcal{W} \subset \mathcal{V}$. We further assume that a contaminant substance can be detected at any concentration, i.e. $\epsilon_c = 0$. In this example, we consider that the set of candidate sampling locations is the same as the demand nodes, i.e. $\mathcal{V}_s \equiv \mathcal{W}$. Regarding the finite set $\mathcal{T}_m$ comprised of time steps at which sampling can be conducted, we assume that sampling can be performed every 30 minutes; thus, there are maximum $N_t = 48$ time instances for manual sampling within one day. From these, the set $\mathcal{B}_m$ of all feasible sampling locations and times has size $N_b = 4\ 800$. Therefore, a contamination fault may occur anywhere and at anytime in the distribution network; this is a common assumption when considering malicious contamination attacks.

For the construction of the fault parameter set $\mathcal{Q}$, we consider faults which start the contamination at a certain location in the network, and continue to contaminate the water at that

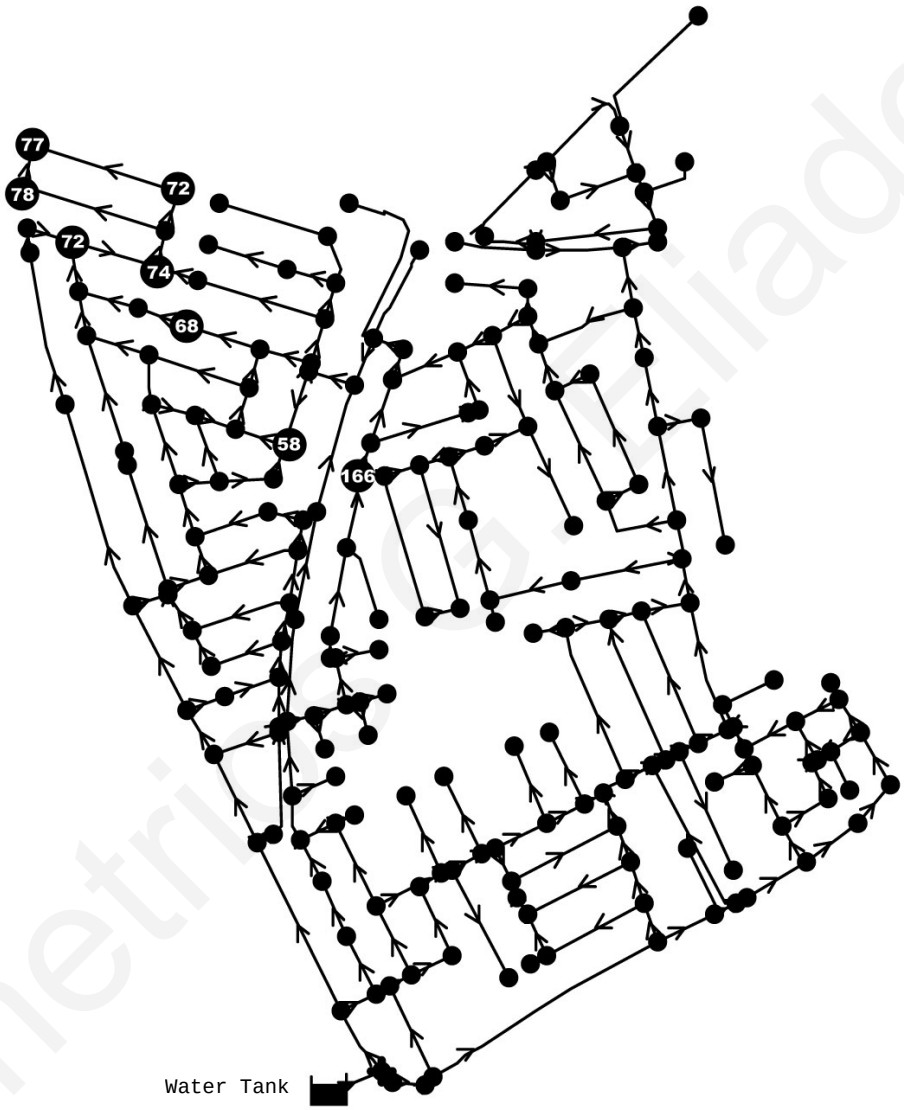


Figure 3.7: Real water distribution system in Limassol, with 321 pipes and 198 junctions. The numbered nodes correspond to the location indices computed, as in Table 3.2.

Table 3.2: Solutions for various sampling schemes and objectives. The smaller the objective function the smaller the possible impact due to a contamination fault.

Samples	Objective	Period	Solution (node,time)	% Objective Function
1	mean	all day	(74, 9A.M.)	7,2%
2	mean	all day	(74, 6A.M.), (74, 2:30P.M.)	4%
3	mean	all day	(74, 1A.M.), (74, 10P.M.), (74, 4:30P.M.)	2,78%
2	mean	8A.M.-5P.M.	(74, 8A.M.), (74, 4:30A.M.)	4,13%
3	mean	8A.M.-5P.M.	(74, 8A.M.), (77, 12A.M.), (74, 5P.M.)	3,21%
1	max	all day	(58, 9:30A.M.)	41,91%
2	max	all day	(58, 9A.M.), (68, 6P.M.)	20,85%
3	max	all day	(76, 5A.M.), (72, 11:30A.M.), (78, 6P.M.)	15,78%

location until detection. In specific, a fault with a parameter $\Theta \in \boldsymbol{Q}$ is modeled as a step-function of $1\ mg/L$, which starts at a certain time within a certain period $k_0 \in \{1, ..., 288\}$ and is terminated when the contamination is detected through sampling, or after a certain number of time-steps has passed, in specific $T_{max} = 288$. We construct the set $\boldsymbol{Q}$ of fault parameters characterizing faults which occur at one demand node every half hour within one day; 48 faults are therefore considered for each node; $\boldsymbol{Q}$ has size $N_q = 4\ 800$.

For the construction of the overall-impact matrix Ω, we repetitively simulate the operation of the system corresponding to duration of 48 hours, applying one fault $\Theta^{(i)}$ at a time, for $i \in \{1, ..., 4\ 800\}$. Zero time delays are considered. Within one simulation, we measure the impacts based on the various sampling schemes $b^{(j)}$, for $j \in \{1, ..., 4\ 800\}$. Finally, the overall-impact matrix Ω of size $4\ 800 \times 4\ 800$ is constructed.

We examine the single-objective optimization problem, for computing a solution with $M_m \in \{1, 2, 3\}$ sampling events within one period, i.e. a day, in order to minimize the average (and maximum) volume of polluted water consumed. The optimization problem is solved using an integer evolutionary optimization algorithm.

In each problem examined, the evolutionary algorithm iterated for 50 epochs, and for each epoch a population of 10 000 individual solutions was built. The following scenarios were considered: to optimize the average/maximum consumption of polluted water for various numbers of samples, under various sampling periods constraints (e.g. during work hours).

The results are summarized in Table 3.2. The smaller the objective function the smaller the

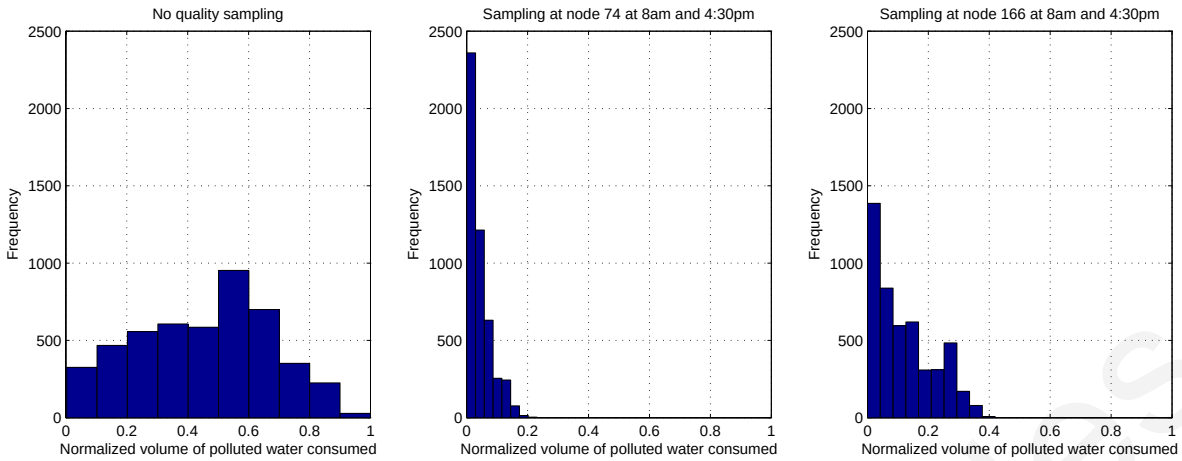


Figure 3.8: Histogram of the normalized impact for all the faults considered.

possible impact due to a contamination fault. Each element in the solution corresponds to a node and a time when sampling should be performed. For instance, for the first row, if only one sampling is to be conducted within a day, by sampling at node "74" at 9A.M., the average consumption of polluted water consumed is minimized. Furthermore, for this solution, the objective function, i.e. the average consumption of polluted water, is 7.20% (with respect to the maximum impact of the worst contamination fault). When an all-day sampling period is considered, some solutions may correspond to non-working hours; this however can be adjusted by imposing constraints to the optimization algorithm. As expected, the constrained solutions are equal or worse to the all-day sampling solutions. For the network examined, it appears that node "74" is quite important from an average impact objective. By minimizing the maximum polluted water consumption, however, the solutions may be different; for instance, for two-samplings per day, examining node "58" at 9A.M. and node "68" at 6P.M. can reduce the maximum impact by 20.85% (with respect to the maximum impact of the worst contamination fault).

The three histograms in Fig. 3.8 illustrate the effectiveness of the proposed algorithm for the selection of manual sampling locations and times, in the volume reduction of polluted water consumed for each contamination fault-scenario, over the case when no manual sampling is conducted. The first histogram is computed by simulating each contamination fault scenario from the set of all contamination fault scenarios Q, under a nominal consumption model, and by measuring the volume of polluted water consumed, if no detection ever occurs.

The second histogram depicts the fault-scenario distribution with respect to their maximum impact when an optimal solution is used from Table 3.2; in specific, the solution methodology was used to compute two samples per day (at node '74'), within working hours (at 8A.M.

and 4:30P.M.) and by minimizing the average impact, i.e. the volume of polluted water consumed. It is assumed that when a contamination is detected, the system stops and no damage is further caused. What we can observe by comparing the first two histograms is that, for the 4 800 fault-scenarios considered, the impact, i.e. the volume of contaminated water consumed, is significantly reduced. In fact, the worst-case fault scenarios in the first histogram can cause less than 40% of their maximum impact, if manual sampling at node '74' at times 8A.M. and 4:30A.M. is conducted.

To illustrate the effectiveness of the proposed methodology over other sampling schemes, the third histogram depicts the frequency distribution of manual sampling at the node '166', which has been arbitrarily selected based on the fact that it is located in a central location in the distribution location, as seen in Fig. 3.7. Sampling times have been arbitrarily selected to be the same as in the computed solution described in the second histogram. By comparing the second and the third histogram, i.e. the optimally computed and the arbitrarily selected solution, we confirm that, although both manage to reduce the impact for all fault-scenarios, the solution computed by using the proposed methodology is significantly better than the arbitrary one.

In general, the larger the number of nodes in the network and the larger the number of fault-scenarios considered, the larger the space and time requirements in computing the overall-impact matrix. By applying expert knowledge, however, it may be possible to sufficiently reduce the size of this matrix, in accordance to the available computational space.

3.5 Concluding Remarks

In this Chapter we have examined the problem of finding the best locations in the network for the installation of sensors which measure contaminant concentrations, as well as the problem of manual sampling scheduling. In specific, a methodology was proposed for the formulation of the problem into a single or a multiple objective optimization problem, utilizing risk-objectives such as the average and maximum impact, as well as the Conditional Value-at-Risk. Simulation examples were presented, which demonstrated the solution methodology on one illustrative network and one real drinking water distribution network. The optimization problems were solved using single and multi-objective evolutionary algorithms.

Difficulties in computing solutions for the sensor placement problem may arise due to various factors, such as the number of nodes, fault scenarios, objectives and sensors, as well as the optimization program parameters. In practice, as the problem is solved off-line, it is

possible to reduce the size of the problem so that it is tractable with respect to the computational power available. In specific, the distribution network can be further "skeletonized" by grouping adjacent nodes with similar hydraulic characteristics and by aggregating their demand flows; higher level reasoning can be applied for the selection of a finite number of the most representative fault scenarios. In addition, the optimization program parameters (such as the solution population number and the number of iterations) can be chosen so that near-optimal solutions are computed within a desired time.

Furthermore, in this chapter we addressed certain issues with regard to the selection of locations and times for manual sampling, in order to examine the quality of water, i.e. whether sufficient quantity of chlorine residual exists, or whether the concentration of a substance exceeds the pre-specified limits. The selection of location and time for manual sampling in this work was driven by security-oriented criteria, aiming at minimizing the possible damage caused due to a contamination event, which may occur accidentally, or even intentionally. The mathematical formulation which describes the dynamics involved was presented, and a solution methodology was proposed. The problem was formulated into a single objective optimization problem, utilizing certain risk-objectives such as the average and maximum impact. A simulation example was presented on a real drinking water distribution network. The optimization problem was solved using a single-objective evolutionary algorithm.

The contributions of this work are the presentation of a rigorous mathematical formulation and the solution of the maximum redundancy sensor placement problem, as well as the security-oriented sensor placement problem considering risk-objective functions; the formulation is extended for the security-oriented manual sampling scheduling problem. The work presented in this chapter has been published and presented in peer-reviewed journals [58,61,66] and conferences [55,59,60,62–64], as well as by collaborators in [123]. In addition, the maximum redundancy formulation has been applied to power transmission networks for Phasor Measurement Units location selection; part of the results have been published in a peer-reviewed journal and a conference [33,34] by the collaborative team.

Chapter 4

Quality Regulation

In this chapter we address the issue of designing algorithms for the regulation of the spatial-temporal distribution of chlorine residual in drinking water distribution networks, which are influenced by unknown time-varying water demands. A solution methodology is presented using model predictive control principles combined with on-line adaptive forecasting of the system hydraulics as they are driven by consumer demands. The periodic nature of water demands allows the use of Fourier series with coefficients which change adaptively, for forecasting future flows at various nodes in the distribution network. The objective is to regulate chlorine residuals, which act as a disinfectant for enhanced water quality, within certain lower and upper concentration bounds. Simulation results on a distribution network are used to illustrate the performance of the proposed chlorine regulation algorithm under unknown water demands and the trade-offs involved in the chlorine regulation methodology.

4.1 Background

Water can serve as the transportation means for various pathogens, chemical agents, metals and bacteria, which may affect human and animal health. In the past, certain outbursts of cholera or typhus in urban areas were caused by water infected with pathogens. It is therefore important to deliver safe water to consumers, by disinfecting it through the use of suitable reacting agents. The most common disinfectant is chlorine, which neutralizes micro-organisms and other substances so that they are below safety concentrations, as set by various organizations. Chlorine is usually added to water in a gas form, through boosting stations located at treatment points at the beginning or within the network. Under normal conditions, at concentration $0.03 - 0.06\ \frac{mg}{L}$ chlorine can inactivate bacteria within 20 minutes [168].

Quality specifications have been set by various organizations, such as the European Com-

mission [67] and the US Environmental Protection Agency [169]. For example, to avoid odour or taste complaints chlorine concentration in drinking water should be less than $0.6 - 1.0\ \frac{mg}{L}$. An upper limit of $4\ \frac{mg}{L}$ was set by the EPA for safety reasons, and a lower limit of chlorine concentration is usually considered at $0.2\ \frac{mg}{L}$, which needs to be detectable throughout the network to guarantee disinfection. Unfortunately, high chlorine concentrations have negative effects on human health; research has suggested that there is a positive association with bladder and rectal cancer in humans consuming water with chlorine by-products, such as trihalomethanes (THM) [114], whose total concentration should be less than $0.08 - 0.1\ \frac{mg}{L}$. Reducing chlorine concentrations across the network corresponds to reduced concentrations of by-products.

Maintaining a detectable residual of chlorine throughout the network is a control problem of regulating the spatial-temporal distribution of disinfectant across the water distribution network at a certain concentration and within bounds. It is a multi-input multi-output problem, in which we consider as inputs the chlorine mass or concentration added by actuation at the booster stations, and as outputs the measurements from quality sensors installed at various locations in the network. Some of the challenges for this control problem are the time-varying consumer demands which affect the hydraulic dynamics and are usually unknown, as well as the complex chlorine reaction dynamics.

In practice, chlorine injection concentrations as well as the detectable chlorine residual in the network may vary significantly; in a study of water quality at treatment facilities serving 19 Massachusetts communities in the USA, it was found that injected chlorine concentrations ranged between $1 - 29.7\ \frac{mg}{L}$ and residual levels in the network ranged between $0.2 - 6.0\ \frac{mg}{L}$ [168].

From a control systems perspective, both hydraulic and the water quality dynamics in a water distribution network are influenced by disturbances due to unknown variations in water demands. Large transport time-delays and storage tanks affect the system with respect to quality control. Other factors such as variations in the temperature or faults may affect the system. Controlling the level of chlorine residuals has gained attention in recent years [1, 20, 24, 26, 27, 35–37, 43, 47, 53, 119, 122, 126, 127, 130, 135, 136, 147, 181–184]. A number of studies have examined the problem within an off-line framework for finding an input scheduling algorithm using optimization techniques such as linear programming [20, 43], least squares [135], genetic algorithms [47, 119, 130] and goal programming [1]. Optimization of water pumps operation has also been considered in relation to the quality control problem using non-linear optimization [147], and genetic algorithms [122].

A robust model predictive control scheme [26, 35–37] and an adaptive control framework considering periodic variation of demands [127, 182] have also been examined. Both approaches are based on control systems theory in addressing the on-line quality control problem. The Model Predictive Control (MPC) framework [30] is specifically suitable for systems with time-delays and hard constraints. The MPC approach is based on explicitly using a system's model to predict the outputs at a future time, and then computing the control sequence by minimizing an objective function. In [26], the uncertainties in the time-varying parameters, errors in model structure, actuator and measurements errors have been modelled within a set-bounded approach. Safety zones are applied that narrow the constraints in order to deal with uncertainties. From a different approach, the adaptive control framework presented in [127, 182] uses Fourier series to capture the periodic nature of the system's dynamics. By using this information, the time varying coefficients of the identification model are replaced with a number of sinusoid terms with constant unknown coefficients that are to be estimated on-line.

For regulating quality, it is useful to model water consumption, which is important in computing the hydraulic dynamics. Hydraulic dynamics are driven by consumer demands and as a result they inherit their periodic property. Various approaches have been examined in previous research for modelling demands and predicting a number of future steps, using auto-regressive models, expert systems and others [5, 22, 29, 89, 151, 153, 157, 158].

4.2 Outline of the Quality Regulation Problem

Consider a water distribution network, with a number of chlorine booster actuators installed at certain locations so that the chlorine concentration at those locations is controllable, and a number of chlorine concentration sensors installed at certain locations which capture the overall water quality. In addition, we consider that the flows and pressures at certain "representative" locations in the network are monitored. We consider that there exists a model whose parameters have been calibrated with respect the measurements and information from the actual system. In practice, the model is a software algorithm which takes into consideration the various discrete and continuous parameters in a water distribution system; in addition, water providers often utilize and calibrate a software model of their distribution networks, to better monitor its behaviour during normal and faulty operation.

In this chapter, we present a quality regulation methodology based on model predictive control, the underlying idea of which is to compute a sequence of future control inputs, so that

an objective function is minimized over a prediction horizon [30]. For solving the optimization problem, we need a forecast of the hydraulic dynamics; for this, we utilize an adaptive approximation methodology to learn the time-varying demand dynamics.

Let $u(k) \equiv [u_c(k), u_h(k)]^\top$ be the control input vector, where $u_c(k)$ and $u_h(k)$ is the quality and hydraulic control input vector respectively. In addition, let $y(k) = [y_c(k), y_h(k)]^\top$ be the measurable output vector, where $y_c(k)$ and $y_h(k)$ is the measurable quality and hydraulic output vector respectively. In addition, let $r(k) \equiv [r_c(k), r_h(k)]^\top$ be the reference signal vector comprised of the reference quality $r_c(k)$ and the reference hydraulics $r_h(k)$ signal vectors. Finally, let $f_{u_c}(\cdot)$ correspond to the quality controller and $f_{u_h}(\cdot)$ to the hydraulic controller. The control inputs are given by

$$u_c(k) = f_{u_c}(y(k), r_c(k))$$
$$u_h(k) = f_{u_h}(y_h(k), r_h(k)),$$

assuming that the hydraulic controller is decoupled from the quality controller. In this work we consider that a hydraulic controller $f_{u_h}(\cdot)$ has been designed separetly and operates independently from the quality controller $f_{u_c}(\cdot)$.

In the following we present an overview of the quality regulation algorithm. We assume that reaction dynamics are linear, and that the principle of superposition holds [20, 127, 149, 184, 197], such that for the single input single output case, the hydraulic output is given by

$$y_h(k) = \sum_{i=1}^{N_y} \Gamma_{(1,i)}(k) u_c(k-i)$$

where N_y is the time window, i.e. a number of time steps before which inputs do not have significant impact on the output $y_h(k)$ at time k, and $\Gamma(k)$ is an impulse response coefficient matrix. In a similar way we can formulate the input/output relation in the multi-input multi-output case, where M_y nodes are utilized for monitoring chlorine concentration, and M_c nodes for regulating chlorine concentration.

At each time k, the hydraulic measurements $y_h(k)$ are used by a hydraulic calibration algorithm to estimate the normalized demand patterns, using an adaptive algorithm for learning the Fourier series approximation. The Fourier series approximation is used for the prediction of a certain number of future demands use those predictions in computing the system's hydraulic dynamics; for this, a model of the hydraulic controller is utilized within the quality regulation algorithm.

To construct the time-varying impulse responce matrix $\Gamma(k)$ we compute the impulse responses using the calibrated software model, the forecasted demands and the hydraulic con-

trol model. We further compute $Y(k)$ as the future responses vector, which is due to the past inputs, i.e. those that have already been injected into the system and which will appear in the future.

The control algorithm is in fact an optimization program, in which $J_c(\cdot)$ is the objective function; this function corresponds to the square error between the prediction of chlorine concentration with a reference signal $r_c(k)$, at some locations over a time frame. Constraints can be included in the optimization program, so that the input and output signal are bounded within some safety regions. The output of the optimization is a vector of future inputs $U(k)$. The quality control inputs for the next time step are applied to the system, and the algorithm iterates. The chlorine regulation algorithm is designed in the following subsections.

4.3 Estimating Hydraulic Dynamics

The driving force behind the hydraulic dynamics is water usage [179], mainly due to consumer demands as well as due to water losses. Consumer demands may be affected by various parameters, such as time of day, weekends, season, temperature, rainfall and others; in general consumer demands are considered to be approximately periodic signals.

Demands can only be estimated implicitly, since there is no practical way to measure the water that exits the network in real time. Instead, flows in certain pipes and water head in certain junctions are measured. Some information is also gathered through utility billing. Specifically, by considering historical data of water meters at buildings and their billing records, utilities can determine the average demand for a specific meter over a certain period of time (e.g. of two months).

In general, the volume of water entering the distribution network should be equal to the measured volume that is consumed; however, this is typically not the case due to water losses from leaks which occur at unknown locations of the network and other reasons.

Let flow $d_{w_i}(k)$ be the demand at node $w_i \in \mathcal{W}$, at discrete time k. We define $p_{w_i}(k)$ as the normalized demand pattern at node w_i, which is given by

$$p_{w_i}(k) = \frac{d_{w_i}(k)}{\bar{d}_{w_i}}, \tag{4.1}$$

where $\bar{d}_{w_i}$ is the *base demand*, i.e. the average water demand at node w_i, which is calculated using historical data gathered for billing purposes, over a period of a few months. In practice, however, water demands are not monitored at every node in the distribution network; instead,

only for certain consumers (e.g. at factories), water utilities may measure water demands directly using flow meters. Thus, the actual water demand $d_{w_i}(k)$ can be computed at discrete time k, by multiplying the demand pattern $p_{w_i}(k)$ with the base demand $\bar{d}_{w_i}$ [142]. Since the normalized demand pattern is in general not known, certain assumptions must be taken, such that consumers of similar type has similar normalized demand patterns, as in schools, office complexes etc.

Studies have shown that temporal variations of water consumption usually follow a diurnal demand pattern with a 24-hour period [153, 157, 179]. Water usage is usually low during the night in populated areas, fluctuates during the day and repeats in a similar pattern in the next days. In practice, there is not a common demand pattern for all nodes, since demands are heavily based on social characteristics and the types of business that operate within a certain region.

Since most demand patterns are not explicitly measured they can only be estimated by using on-line flow and pressure measurements at various locations in the network. In water research, the calibration of water distribution models describes the process of modifying nodal demands and other parameters so that the difference between the on-line measurements and the simulated ones is minimized [38]. The general hydraulic calibration model has received significant attention in the previous decades and it remains an open research area [134, 148].

Flows and pressures in certain pipes and nodes are often measured using sensors and the Supervisory Control And Data Acquisition (SCADA) system. Estimating water demands for all nodes in the network, based on these measurements alone, is not a trivial task. Some issues and a solution methodology for this problem using predictor-corrector methodology and extended Kalman filtering, are presented in [151].

As a basic scenario, consider the case where flow meters are placed only at the pipes which supply water to the distribution system. According to the mass conservation principle, inflows should be equal to outflows; in this case outflows may correspond to consumer demands, water loss due to leakages or water exiting to some other part of the network, e.g. a tank. In current practice, water utilities segment their distribution networks into District Metered Areas (DMA), whose inflows and outflows are monitored. Each DMA may have a different overall demand pattern, depending on the consumer types it serves.

For the l-th DMA in a distribution network, we define $\mathcal{W}_l$ as the set of demand nodes which are included in that DMA, and define $q_{in}^{(l)}(k)$ and $q_{out}^{(l)}(k)$ as the sum of its inflows and outflows respectively. Let $f_p : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ be a calibration function, which is selected off-line and is used to compute the total nodal demands at the l-th DMA, at discrete time k. The

estimated normalized demand pattern, $\hat{p}_{w_i}(k)$, for the i-th demand node in a DMA, is given by

$$\hat{p}_{w_i}(k) = \frac{f_p(q_{in}^{(l)}(k), q_{out}^{(l)}(k))}{\sum_{w_i \in \mathcal{W}_l} \overline{d}_{w_i}}. \tag{4.2}$$

In general, the estimated normalized demand pattern $\hat{p}(k)$ may not be equal to the actual normalized demand pattern $p(k)$, due to uncertainties; however, we assume that the selection of the calibration function $f_p(\cdot)$ is such that the modelling uncertainty is within certain bounds.

4.4 Adaptive Forecasting of Periodic Time Series

Forecasting the hydraulic dynamics at various nodes in the distribution network is important in the design of the quality regulation algorithm, since the forecasted demands are used in the software model to compute future state of the system under certain control inputs. In the case of water distribution networks hydraulic dynamics are driven by unknown water demands; these demands however are approximately periodic.

Various methods exist in the literature in forecasting time-series, such as autoregressive moving average models, regression models, Kalman filtering, neural networks and others [23, 93]. In this work we examine the use of Fourier series as the approximation structure in an adaptive framework. The motivation behind the use of Fourier series, is that they are used to model periodic signals, as in the case of the hydraulic dynamics, by formulating a sum of suinosoidal functions. An added benefit in using Fourier series is that they are linear in the parameters, and can be represented as the multiplications of two vectors, a parameter vector θ and a regressor vector $\zeta(k)$.

Let $\hat{q}_d(k) \equiv f_p(q_{in}^{(l)}(k), q_{out}^{(l)}(k))$ be the water demands at a certain DMA, as computed using the calibration function. The periodic variation in $\hat{q}_d(k)$ can be represented in terms of a Fourier series, which is given by

$$\hat{q}_d(k) = \theta^\top \zeta(k) + \epsilon_d, \tag{4.3}$$

where $\theta \in \mathbb{R}^{2N_z+1}$ is a vector of unknown parameters representing the first N_z Fourier coefficients. The regressor vector $\zeta(k) \in \mathbb{R}^{2N_z+1}$ is given by

$$\zeta(k) = [1, \sin(\omega k), ..., \sin(N_z \omega k), \cos(\omega k), ..., \cos(N_z \omega k)]^\top, \tag{4.4}$$

where $\omega = \frac{2\pi}{T_H}$ and ϵ_d is the residual error of the Fourier series approximation.

The unknown parameter vector θ is estimated on-line by computing the estimation error and updating the adaptive law. Let $\hat{\theta}(k)$ be the parameter estimate of the unknown θ at time k, and let

$$e_q(k) = \hat{q}_d(k) - \hat{\theta}^\top(k)\zeta(k) \tag{4.5}$$

be the output estimation error. Using the standard gradient approach [180], the adaptive law for $\hat{\theta}(k)$ is given by

$$\hat{\theta}(k+1) = \hat{\theta}(k) + A_G \frac{\zeta(k)}{a_F + \zeta(k)^\top \zeta(k)} e_q(k) \tag{4.6}$$

where a_F is a design parameter and A_G a gain diagonal matrix, which are selected offline. To guarantee the learning law stability, $a_F \geq 0$ and $0 < A_G^i < 2$, for the i-th diagonal element of the gain matrix [7, p.54].

Once we obtain the coefficients of a Fourier series which approximate the actual periodic total demands, we can compute for a certain horizon, forecasts for the estimated demand pattern at each demand node.

4.5 Model Predictive Controller

In this section a basic multivariable and constrained model predictive control formulation of the water quality control problem will be presented. The underlying idea is to compute a sequence of future control inputs, so that an objective function over a predicted horizon is minimized. The controller's constituting parts are analyzed in the following sections.

4.5.1 Constraints

Drinking water distribution systems are subject to various constraints because of regulations for safe operation and good water quality. Actuators injecting chlorine cannot remove excess chlorine from water, and in addition, there may be a physical limit on the maximum quantity of disinfectant that can be injected. Therefore, the chlorine concentration control input at the i-th booster station $u_c^{(i)}(k)$ is bounded within $u_c^{(i)}(k) \in [\underline{u}, \overline{u}]$, where $\underline{u} = 0\ \frac{mg}{L}$.

Let $y_c(k)$ be the output vector of chlorine concentrations, monitored by on-line sensors. Depending to the regulations, this output concentration must be constrained to a lower and upper bound, $\underline{y}$ and $\overline{y}$ respectively. For example, in some cases, the lower limit is specified by regulations and is a hard limit (e.g. $\underline{y} = 0.2\ \frac{mg}{L}$), whereas the upper limit $\overline{y}$ can take any value below the highest concentration allowable set by regulations (e.g. $\underline{y} = 4.0\ \frac{mg}{L}$). It is important to note that due to the dangerous chlorine by-products, such as trihalomethanes,

chlorine concentrations must be kept as low as possible, without compromising in the same time water safety.

4.5.2 Process Model

The process model used by the chlorine regulation algorithm must be able to capture the quality dynamics of the water distribution system and be used for computing quality output predictions. We follow a soft-computing approach by utilizing the hydraulic and quality simulation system, EPANET, which, when calibrated, can be used in computing discrete impulse responses of multiple quality inputs measured at multiple output, utilizing the forecasted estimated normalized demand patterns.

To compute the impulse response matrix $\Gamma(k)$, we use an algorithm $f_E(\cdot)$, such that $\Gamma(k) = f_E(\hat{p}(k))$. The impulse response matrix is constructed for each input/output with this algorithm, through iterative simulations of discrete impulses corresponding to contaminant injections at one discrete time instance.

Let M_y be the number of output nodes where chlorine concentration is monitored and is to be regulated, and M_c the number of input nodes where chlorine is injected. In addition, let H_p and H_c be the prediction and control time horizon respectively. The dynamic matrix $\Gamma(k)$ is given by

$$\Gamma(k) = \begin{bmatrix} \Gamma_{(1,1)}(k) & \dots & \Gamma_{(1,M_c)}(k) \\ \vdots & \ddots & \vdots \\ \Gamma_{(M_y,1)}(k) & \dots & \Gamma_{(M_y,M_c)}(k) \end{bmatrix},$$

where $\Gamma_{(i,j)}(k) \in \mathbb{R}^{H_p \times H_c}$ for $i = \{1, ..., M_y\}$ and $j = \{1, ..., M_c\}$, is an impulse response matrix of one the j-th input with respect to i-th output. The matrix $\Gamma_{(i,j)}(k)$ given by

$$\Gamma_{(i,j)}(k) = \begin{bmatrix} \gamma^{(i,j)}(k,k) & \dots & \gamma^{(i,j)}(k,k+H_c-1) \\ \vdots & \ddots & \vdots \\ \gamma_{(i,j)}(k+H_p-1,k) & \dots & \gamma_{(i,j)}(k+H_p-1,k+H_c-1) \end{bmatrix},$$

where $\gamma^{(i,j)}(k_1,k_2)$ is the impulse coefficient, which is computed when a discrete impulse is applied at the j-th input at time k_1, and its impact is measured at the i-th output at time k_2.

4.5.3 Predicted Outputs

Let $\hat{Y}(k)$ be the predicted output vector for H_p steps into the future based on information at time k, $Y(k)$ is the future responses vector which are caused by the past inputs, and $U(k)$

the computed future inputs vector, such that

$$\hat{Y}(k) = \Gamma(k)U(k) + Y(k), \tag{4.7}$$

for which $U(k) \in \mathbb{R}^{M_c H_c}$ and

$$\begin{aligned}
\hat{Y}(k) &= [\hat{y}_c^{(1)}(k+1), ..., \hat{y}_c^{(1)}(k+1+H_p), ..., \hat{y}_c^{(M_y)}(k+1), ..., \hat{y}_c^{(M_y)}(k+1+H_p)]^\top \\
U(k) &= [u_c^{(1)}(k), ..., u_c^{(1)}(k+H_c), ..., u_c^{(M_c)}(k), ..., u_c^{(M_c)}(k+H_c)]^\top \\
Y(k) &= [Y^{(1)}(k+1), ..., Y^{(1)}(k+1+H_p), ..., Y^{(M_y)}(k+1), ..., Y^{(M_y)}(k+1+H_p)]^\top.
\end{aligned}$$

4.5.4 Objective Function

The objective is to compute vector of inputs $U(k)$ which minimizes the difference of $\hat{Y}$ with the reference signal $r_c(k)$ as well as minimize the overall magnitude of the input signal $U(k)$.

Let $J_c(k)$ be the quadratic objective function which we will use to compute the control input signal, such that

$$J_c(k) \equiv \sum_{j=1}^{H_p} \left[\hat{y}_c(k+j) - r_c(k)\right]^2 + \sum_{j=1}^{H_c} \eta_c \left[u_c(k+j)\right]^2, \tag{4.8}$$

where η_c is a penalizing factor for the input variable variations, and H_p and H_c are the prediction and control horizon time steps. Using (4.7), the objective function given by (4.8) can be rewritten in a vector form as

$$\begin{aligned}
J_c(k) &= (\hat{Y} - r_c)^\top (\hat{Y}(k) - r_c) + \eta_c U(k)^\top U(k) && (4.9)\\
&= (\Gamma(k)U(k) + Y(k) - r_c)^\top (\Gamma(k)U(k) + Y(k) - r_c) + \eta_c U(k)^\top U(k). && (4.10)
\end{aligned}$$

By rearranging the terms, the objective function can be written as

$$J_c(k) = \frac{1}{2} U(k)^\top P_1(k) U(k) + P_2(k)^\top U(k) + P_3(k), \tag{4.11}$$

where $P_1(k) = 2(\Gamma(k)^\top \Gamma(k) + \eta_c I)$, $P_2(k)^\top = 2(Y(k) - r_c)^\top \Gamma(k)$ and $P_3(k) = (Y(k) - r_c)^\top (Y(k) - r_c)$.

With this formulation, optimization can be performed using standard off-the-shelf constrained optimization algorithms using quadratic programming or least squares, to find the vector $U(k)$ which minimizes the objective function.

The optimization problem is therefore formulated as

$$U(k) = \underset{U \in \mathbb{R}^{M_c H_c}}{\operatorname{argmin}} \frac{1}{2} U^\top P_1(k) U + P_2(k)^\top U, \tag{4.12}$$

subject to

$$\begin{bmatrix} I \\ -I \\ \Gamma(k) \\ -\Gamma(k) \end{bmatrix} U \leq \begin{bmatrix} \mathbf{1}\overline{u} \\ -\mathbf{1}\underline{u} \\ \mathbf{1}\overline{y} - \Upsilon(k) \\ -\mathbf{1}\overline{y} + \Upsilon(k) \end{bmatrix},$$

where $\mathbf{1} = [1, 1, ..., 1]^\top$.

4.6 Simulation Examples

To illustrate the adaptive forecasting and the model predictive controller, we present a number of simulation studies using a network topology of a realistic drinking water distribution system, as provided in the EPANET package, namely 'Network 3', which is depicted in Fig. 4.1. The network is comprised of two water supplies as well as three tanks and two pumps working on a daily schedule. Chlorine booster stations as well as hydraulic and quality dynamic sensors are installed in the network at the two water sources. We consider that the network parameters, such as roughness coefficients and pipe lengths, are known.

4.6.1 Forecasting hydraulic dynamics

The adaptive approximation must learn the parameters of an approximation structure, which describes the demand patterns. The closer the demand forecasts are to the future demands, the better the quality regulation. The ability of the adaptive approximation to learn the demand pattern is shown in the following example, by computing the unknown coefficients of a Fourier series, and for comparison, of a radial basis function.

At each time step, $\hat{q}_d(k)$ is computed based on the on-line flow measurements, and the adaptive law updates the parameters so that the approximation structure learns the unknown demands.

The adaptive approximation updated for demand data corresponding to ten days, with samples every 5 minutes. The weights are adjusted in each time step in order to capture the hydraulic dynamics. For the Fourier series, the first $N_z = 50$ terms are considered with daily period; for the radial basis function, we selected $N_z = 120$ centers are spread evenly between [0, 24] hours, with 0.2 standard variation.

Various scenarios were examined, where the demands at in each node which is not monitored, were affected by random disturbances. The baseline scenario is when the estimated

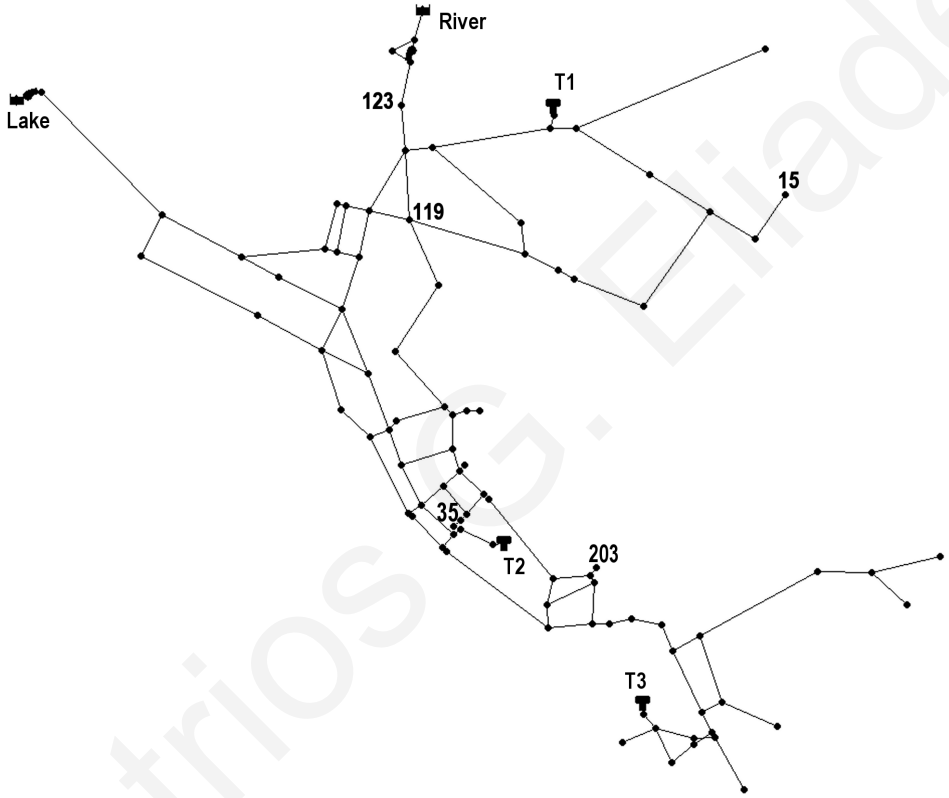


Figure 4.1: A drinking water distribution network; flows from the water sources (Lake and River) and from the storage tanks (T1, T2, and T3) are monitored on-line, as well as the demand flows at certain nodes ('15','35','123','203'). On-line quality sensors measuring chlorine concentrations are installed at nodes '119' and '203'.

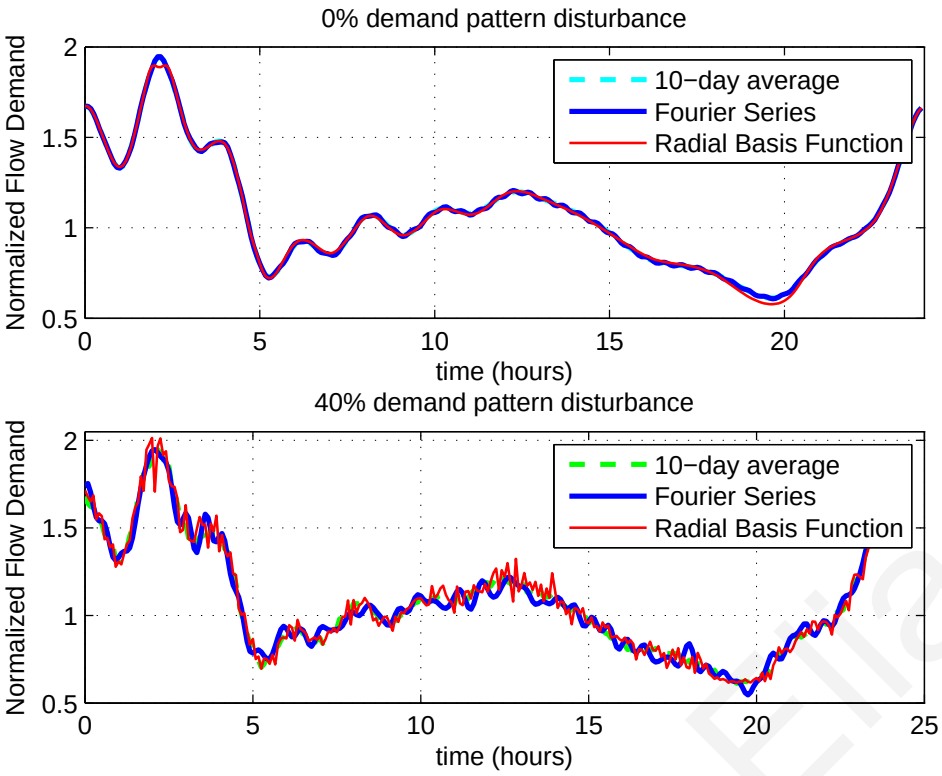


Figure 4.2: Approximated demand patterns using a Fourier series and a Radial Basis Function, after a 10-day period of learning with zero and 40% disturbances.

normalized demand pattern is same as the actual normalized demand patterns at each node. The scenario with disturbances uses white noise with zero mean to change the nominal demand flows.

The result of learning for 10 days are presented in Fig. 4.2. We observe that both approximation structures manage to learn the hydraulic characteristics, even when the nodes are influenced by some disturbances.

Another issue worth considering is that adaptive approximation does not terminate; therefore it is able to respond to changes in the network or consumption behaviour, to capture seasonal changes etc. Furthermore, by using extra information regarding the nature of consumers, it is possible to compute demand pattern models for groups of nodes that are similar to the actual patterns.

4.6.2 Multi-Input Multi-Output Quality regulation

In this experiment we demonstrate the ability of the proposed algorithm to regulate the chlorine concentrations at specific nodes in the network, by adjusting chlorine concentrations at the points of chlorine injection. We consider that the adaptive approximation was

performed with the Fourier series for a 10-day period. For the first 48 hours, the input at both locations where chlorine boosters are installed, is the step signal $u_c(k) = [0.4, 0.4]^\top mg/L$. The quality regulation algorithm is turned on at time 48, with prediction horizon $H_p = 48$, control horizon $H_c = 48$ and $\eta_c = 10$.

We consider the case where chlorine booster actuators are installed at the 'River' and the 'Lake', whereas on-line chlorine concentration sensors are installed at nodes '119' and '203'. Figure 4.3 shows the input and output signals for a period of 90 hours. No disturbances are considered in this example. The objective is to regulate the output signals to a setpoint $r_c = 0.25\ \frac{mg}{L}$, while the input and output signals must be bounded with $\underline{u} = 0\ \frac{mg}{L}$, $\underline{u} = 4\ \frac{mg}{L}$, $\underline{y} = 0.2\ \frac{mg}{L}$ and $\underline{y} = 0.5\ \frac{mg}{L}$.

As shown in the figure, for certain periods the lake has zero inflows because the hydraulic controller shuts down the pumps operating at that location. In addition, the high magnitude of concentrations at the lake needed for regulating the concentration at the two nodes, suggests that another chlorine booster should be placed somewhere, possibly somewhere in the center of the network.

Note that, because the regulation algorithm is not driven by quality feedback, the optimization could be performed with respect to various nodes in the network, thus achieving better spatial distribution of chlorine across the network, assuming the hydraulic model is well-calibrated.

In some situations it is impossible for the quadratic program to find a solution which does not violate the constraints. When this occurs, the controller can select a predetermined input signal, or use a different optimization method with relaxed constraints. The controller switches back to quadratic optimization program, when a feasible solution can be computed.

When the hydraulic dynamics are influenced by disturbances, the measured chlorine concentrations in the output locations are different that those estimated by the model. For example, Fig. 4.4 demonstrates the measured and estimated outputs in a node, for a period of time after the controller is enabled. The controller is under the constraints used in the previous example, which are relaxed in case of a not feasible solution. We consider for the simulation that the demand patterns are influenced by a $\pm 30\%$ disturbance, whereas the base demands by a $\pm 5\%$ disturbance. We observe that in certain time-steps, the measured chlorine concentration dropped below the $0.2\ \frac{mg}{L}$ limit, because disturbances significantly influenced hydraulic dynamics.

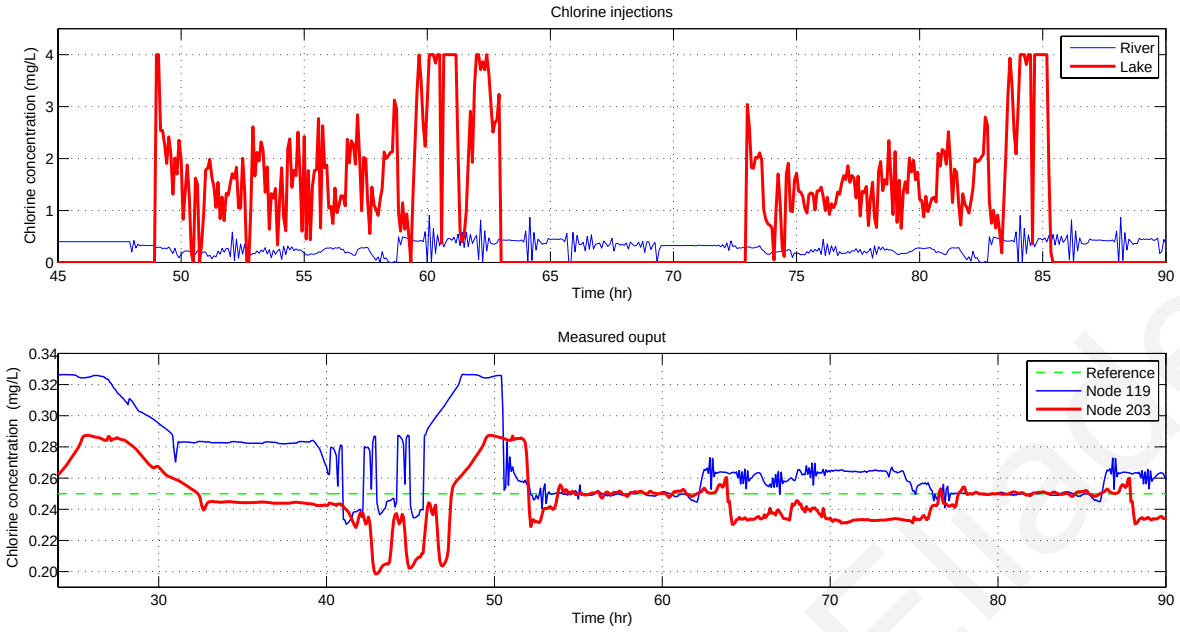


Figure 4.3: A step input is applied for 48 hours, and then the constrained quality regulation algorithm is activated with 48-hour control horizon. The effect of the control action appears after 13 hours, at time 51.

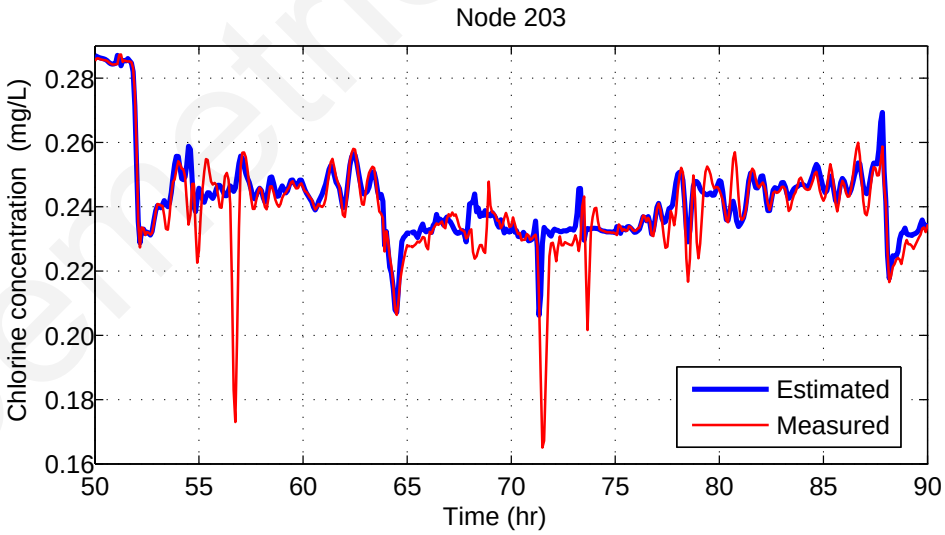


Figure 4.4: The estimated and measured chlorine concentrations at node '203', with ±30% uncertainty in the total demand and ±5% uncertainty at the base demands.

4.7 Concluding Remarks

In this chapter we have studied the problem of quality regulations for drinking water distribution networks, by exploiting the approximately periodic nature of hydraulic dynamics. The approach presented was based on an online predictive control framework. As the actual system is significantly time-varying, since it is driven by unknown demands, future states of the distribution system are unknown; therefore the future hydraulic dynamics, and by projection the water demands, need to be predicted.

The hydraulic dynamics, which are typically of periodic nature, are represented by a Fourier series or other similar approximation structures, with an online parameter estimation of the unknown coefficients. Various approximation structures are suitable for the calculation of daily demand patterns, as shown in the experiments, even under the influence of disturbances. The predicted demand patterns can be directly applied to simulation models which solve the hydraulic and quality equations numerically for the future time steps.

When the adaptive approximation has learned the hydraulics, the controller is activated, in order to regulate the chlorine concentrations across the drinking water distribution network. The objective function is in a quadratic form, while the optimization program minimizes both the difference between the future output concentration and a regulation concentration, and also the input concentrations.

Certain constraints may be applied to guarantee that the input and output are within those bounds. The optimization algorithms may consider regulating all nodes in the network, even though only some of them are actually monitored. Better hydraulic forecasting methods as well as better water demand estimators would provide more accurate models of the drinking water distribution system to be used with the regulation algorithm, which will assist in achieving better control of the system.

The contributions of this work is the design and implementation of a quality regulation algorithm which learns adaptively the unknown demands, while using existing water distribution system models to compute the input signals by using a regulation algorithm based on model predictive control. The work presented in this chapter has been published and presented in a peer-reviewed conference [54].

Chapter 5

Hydraulic Fault Detection

In this chapter, we formulate the problem of leakage detection in a systems engineering framework, and propose a solution methodology to detect leakages in a class of distribution systems. In specific, we consider the case when water utilities use flow sensors to monitor the water inflow in a District Metered Area (DMA). The goal is to design algorithms which analyze the discrete inflow signal and determine as early as possible whether a leakage has occurred in the system.

The DMA inflow signal is normalized to remove long-term trends and seasonal effects, and two different algorithms are presented for leakage detection. The leakage fault detection algorithm presented in this work is based on learning the unknown, time-varying, weekly periodic DMA inflow dynamics, with the use of an adaptive approximation methodology to update the coefficients of a Fourier series; as detection logic we utilize the CUSUM algorithm. For reference and comparison, we present a solution methodology based on night-flow analysis, using the normalized DMA inflow signal. To illustrate the solution methodology, we present results based on real hydraulic data measured at a DMA in Limassol, Cyprus.

5.1 Background

The International Water Association (IWA) Task Force on Water Losses has defined a set of metrics for water utilities to audit how provided water is consumed or lost. In general, water intended for consumption is segmented into "Authorized Consumption", which corresponds to the billed or unbilled authorized consumption, and to the "Water Losses", which corresponds to the "Apparent Losses" (due to unauthorized use, metering inaccuracies or calibration issues) and to the "Real Losses" (due to leakages, breaks etc.) [3]. Water losses impose a huge economic burden; hence the reduction of both is an important goal in most water

utilities.

Leakage is a type of hydraulic fault, which may be caused due to pipe breaks, loose joints and fittings, as well as to overflowing water storage tanks [69]. Some of these problems are prompted by the deterioration of the water delivery infrastructure, which is affected by age and high pressures.

Leakage faults which occur within the water distribution network may correspond to slowly developing incipient faults, as well as due to abrupt faults, which may require immediate attention. Leakages may cause consumer problems, health risks as well as financial losses [69], therefore their early detection and repair, if possible, is imperative. Leakages are classified by water utilities as "background" (small undetectable leaks for which no action to repair is taken, with single leakage outflow less than $0.25 - 0.5\ \frac{m^3}{hr}$ at $50\ m$ water head), "unreported" (moderate flow-rates which accumulate gradually and require eventual attention), and "reported" (high flow-rates which require immediate attention) [97, 159]. We should note that in practice, there may be a significant time delay between the time a leakage occurs, to the time the water utility detects the leakage occurrence and the time the leakage is located and repaired [159].

From the systems and control viewpoint, we may classify leakages as: a) slowly increasing incipient faults, to describe breaks which in the beginning are small but may deteriorate further while their size increases; b) stepwise abrupt faults, which appear at a certain time step and whose physical characteristics do not change [87, 195].

In 1980, the UK Water Authorities Association proposed the concept of District Metered Area (DMA) management methodology to monitor leakage in water distribution networks [185]. According to [69], the main benefits of using DMA management over standard centralized monitoring and control are reduced fault detection time, better leakage detection, leakage location isolation and low pressure regulation. This approach has been adopted by a large number of water utilities. Each DMA corresponds to a decoupled water distribution network, with one or more water supply pipes; at each supply pipe, real-time flow sensors, pressure sensors and pressure reducing valves are installed [69, 115].

A number of water utilities apply the *minimum night-flow analysis* at each DMA, to estimate the quantity of water loss due to leakages [159]. The utilities examine the inflow at a DMA during the minimum-consumption hours (typically between 2-4 A.M.), at which times flow variance is at the lowest and the leakage outflow is high, due to high pressures. Through observation of the minimum flows during the minimum-consumption hours and by comparing them to measurements from previous days or to certain benchmarks, the water util-

ity operators may be able to detect an unreported leakage fault which has occurred within a DMA [159]. However, detection by this method may not be straight-forward, due to unpredictable variations in consumer demands and measurement noise, as well as long-term trends and seasonal effects. In addition, it may be difficult to detect incipient leakage faults that may occur due to slowly increasing leakages.

In a number of studies, leakage detection has been examined as the inverse problem of computing leakages (magnitudes and locations) based on flow and pressure measurements, as well as other network parameters (such as pipe lengths), by formulating and solving an optimization problem. In [137], steady-flows are considered when solving the inverse problem. In [102], this formulation is extended to take into consideration transients, i.e. changes in pressures due to the fact that pressure waves travel through the distribution network with some velocity. The authors formulate an optimization problem to compute the pipe friction coefficients to calibrate the system model, as well as to compute the orifice area of the leakage models, based on the available data measurements.

For a review of the solution methodologies of the inverse problem using transient analysis, see [42]. Case studies on physical systems have also been published [44]. According to [148], however, transient analysis is not widely accepted by the practitioners due to various reasons, such as cost and lack of expertise.

Some methodologies consider the use of computational intelligence techniques for leakage detection, such as neural networks and support vector machines [31, 72, 73, 106, 116–118]. In [176,190,191], they consider the use of genetic algorithms, in order to calibrate the network model and estimate leakage parameters. Probabilistic and statistical methodologies have also been proposed in [28, 128, 138]. Recently, results from the Wireless Water Sentinel project in Singapore (WaterWiSe@SG) were presented [188], in which transient pressure waves are analyzed using wavelets to detect leakages [156].

In [112, 113], a solution methodology is proposed for the abrupt leakage fault detection problem within a DMA, when measuring inflows and transient waves; the Cumulative Sum chart algorithm "CUSUM" [9, 124] is applied to detect changes in the mean of the difference between filtered estimates of two consecutive time instances.

In practice, some water utilities may not have sensors measuring transient pressure waves, with high-frequency sampling, installed in their distribution networks, in order to use transient leakage detection algorithms. Furthermore, some water utilities may not have well-calibrated models of their water distribution system, as well as representative consumer demand models, to solve optimization problems for leakage detection and estimation. On the other hand,

a number of water utilities measure inflows and pressures at DMA, and these measurements are used for calculating leakages, typically performed using alarm thresholds, or simply by operator observation.

In this work, we formulate the problem of leakage detection in a fault diagnosis framework, and propose a solution methodology for leakages detection in a class of distribution systems. In specific, we consider the case when water utilities use flow sensors to monitor the water inflow in a District Metered Area (DMA). The goal is to design algorithms which analyze the discrete inflow signal and determine as early as possible whether a leakage has occurred in the system. The inflow signal is normalized to remove long-term trends and seasonal effects, and two different algorithms are presented for leakage detection. We propose a leakage fault detection algorithm which is based on learning the unknown, weekly periodic DMA inflow dynamics, using an adaptive approximation methodology to update the coefficients of a Fourier series; as detection logic, we utilize the CUSUM algorithm, which detects changes in the average of a signal. The proposed algorithm is able to learn changes in the weekly periodic dynamics, and detect a leakage fault by monitoring changes in the average value of the off-set (DC) term of the Fourier series. For reference and comparison, we demonstrate a baseline algorithm based on night-flow analysis, and the leakage detection logic depends on a leakage detection threshold. We show that although this algorithm is able to detect leakage faults if the detection threshold is violated, the algorithm has a short memory and as a result, if the threshold has not been violated, after a few days it becomes more difficult to detect a leakage, since it considers the increased flow measurements as normal.

This chapter is organized as follows: a mathematical model of the DMA inflows and the leakages is presented, along with key concepts and definitions; next, a solution methodology based on adaptive DMA inflow approximation is presented and compared to a leakage fault detection algorithm using night-flow analysis. Finally, simulation results are presented based on actual data from a water distribution system in Limassol, Cyprus.

5.2 Mathematical Model of DMA Inflows and Leakages

Consider a District Metered Area (DMA) which is connected to the main water supply network through a single pipe. To reduce the possibility of high pressure at the DMA entrance, a pressure-reduction valve controller is utilized to regulate the pressure at the point of entry within certain bounds. Following the pressure-reduction valve, the hydraulic characteristics of the water are monitored by on-line sensors, with sampling time Δt. At each discrete time

instance $k \in \{0, 1, ...\}$, the water flow $q(k)$ and hydraulic head $h(k)$ measurements are recorded and stored using a SCADA system.

According to statistical theory, time series can be decomposed into some linear and periodic signals, which is useful for forecasting [23, 45]. These techniques have been applied to hydraulic signals [192, 196]. In our work, we consider that variations in flows $q(k)$ can be decomposed into two main components: (i) long-term trends and yearly seasonal changes, and (ii) weekly periodic changes. The long-term trend is typically a monotonic function which describes the increase of water consumption e.g. due to population increase. The seasonal changes describe the variation in water consumption as a result of seasonality within a year. The weekly periodic component describes the fluctuation of the signal throughout one week, which depends on the various social and economic characteristics of the consumers. In some studies, the weekly periodic component is simplified to a daily, 24-hour periodic component. In addition, we may also consider high frequency variations due to unpredictable consumer demands.

In hydraulic dynamics, the uncertainty variance is typically proportional to the weekly periodic component, and in addition, the magnitude of the weekly periodic component is proportional to the seasonal and trend component. Therefore, the multiplicative time-series model is a suitable option for the description of mathematically the hydraulic signals. In the case of DMA inflow $q(k)$, we define $r_t(k)$ as the function which describes both the monotonic trend and the yearly seasonal component of the flow, and $s(k)$ as the function which describes weekly periodic water demand signal; $n(k)$ corresponds to the multiplicative uncertainty component with zero average and variance σ_n^2, such that $n(k) \sim \mathcal{N}(0, \sigma_n^2)$. Therefore, the mathematical model of the flow signal $q(k)$ is given by

$$q(k) = r_t(k)s(k)(1 + n(k)). \tag{5.1}$$

Leakage faults increase the flow and decrease the pressure measurements at the DMA entrance. Mathematical models which describe the leakage flow with respect to the pressure at the leakage location have been proposed in various empirical studies [74, 75, 177]. For example, the leakage flow $\phi_h(k)$ can be modeled mathematically as

$$\phi_h(k) = a_D[f_l(h(k))]^{a_E}, \tag{5.2}$$

where $h(k)$ is the hydraulic head measured at the DMA entrance, $a_D > 0$ is a discharge coefficient, $a_E \in [0.5, 2.5]$ is an exponent term which depends on the leakage type and $f_l : \mathbb{R} \mapsto \mathbb{R}$ is an unknown function which maps the measured head at the DMA entrance to the pressure

at the leakage location (the pressure at the location where the leakage has occurred is usually not available for measurement). In practice, both the discharge coefficient and the exponent term are unknown; however, empirical studies have demonstrated that the exponent for small background leaks is $a_E \approx 1.5$, for larger leaks in plastic pipes $a_E \geq 1.5$ and for larger leaks in metal pipes $a_E \approx 0.5$ [98].

Suppose that a leakage fault starts at some unknown time k_0. In the presence of a leakage fault, the flow model described by (5.1) needs to be modified to account for the leakage. Therefore, the term $\beta(k-k_0)\phi_h(k)$ is added to the multiplicative model in (5.1), where $\beta(k-k_0)$ is the time profile of the leakage fault and $\phi_h(k)$ the flow due to the leakage. A suitable function which describes the time-profile of an abrupt fault is given by

$$\beta(k-k_0) = \begin{cases} 0 & k < k_0 \\ 1 & k \geq k_0 \end{cases}, \tag{5.3}$$

whereas in the case of a slowly evolving incipient leakage fault, a suitable function is

$$\beta(k-k_0) = \begin{cases} 0 & k < k_0 \\ 1 - a_B^{-(k-k_0)} & k \geq k_0 \end{cases}, \tag{5.4}$$

where $a_B > 1$ is an unknown fault evolution rate.

The mathematical model of DMA inflow with leakage is given by

$$q(k) = r_t(k)s(k)(1+n(k)) + \beta(k-k_0)a_D[f_l(h(k))]^{a_E}. \tag{5.5}$$

An important consideration to take into account is the fact that the uncertainty term $r_t(k)s(k)n(k)$ in the flow model (5.5) may have a significantly larger variance than the variance corresponding to the leakage fault. On the other hand, the uncertainty term $r_t(k)s(k)n(k)$ has a zero average value, while at the same time, the leakage fault has a positive (non-zero) average value. This is a key characteristic which we exploit in the leakage fault detection algorithms presented in this work.

The long-term trend and seasonal function $r_t(k)$ is in general unknown. To compute the trend and seasonal signal approximation, we consider a dataset of fault-free historical measurements. Let $\mathcal{K}_h$ be the set of historical time instances considered. Let $\hat{r}_t(k;\theta_r)$ be the estimate of the unknown $r_t(k)$, which is computed off-line, where $\hat{r}_t(\cdot)$ is a selected structure for the estimator and θ_r is a set of parameters that can be adjusted to improve the approximation. For example, $\hat{r}_t(\cdot)$ can be expressed as a polynomial or a Fourier series with fundamental period of a year, to capture the population increase and consumption changes due to seasons. In this case, the parameter θ_r is the coefficients of the polynomial of the Fourier series.

Given sufficient recorded hydraulic measurements, we can compute an estimation $\hat{\theta}_r \in \Theta_r$ of the unknown parameter vector θ_r, where Θ_r specifies the parameter bounds, by solving the least-squares optimization problem

$$\hat{\theta}_r = \underset{\theta_r \in \Theta_r}{\operatorname{argmin}} \sum_{i \in \mathcal{K}_h} \left(q(i) - \hat{r}_t(i; \theta_r)\right)^2 . \tag{5.6}$$

The approximation signal $\hat{r}_t(k; \hat{\theta}_r)$, which is computed off-line, is used in the following sections to normalize the flow signals used in the presented leakage fault detection algorithms. By making the assumption that $r_t(k) = \hat{r}_t(k; \hat{\theta}_r)$, the normalized DMA inflow is given by

$$q_r(k) = s(k)(1 + n(k)) + \beta(k - k_0)\frac{\phi_h(h(k), k)}{\hat{r}_t(k; \hat{\theta}_r)}, \tag{5.7}$$

where $q_r(k) = \frac{q(k)}{\hat{r}_t(k; \hat{\theta}_r)}$.

5.3 Leakage Fault Detection Using Adaptive Flow Approximation

In this section we formulate an on-line leakage fault detection methodology, based on the idea that we can learn adaptively the periodic signal corresponding to the normalized flow, and use that information to detect leakage faults.

The unknown weekly periodic demand signal $s(k)$ is approximated in the form $s(k) = \theta^*(k)^\top \zeta(k)$, where $\theta^*(k)$ is an unknown parameter, and $\zeta(k)$ represents the approximation structure (regressor). In this work, $\zeta(k)$ is a vector of Fourier functions, while $\theta^*(k)$ are the ideal coefficients. In general, $\theta^*(k)$ is time-varying, representing slowly changing consumption characteristics due to variations in people's behaviour, as well as due to various external events such as the daylight savings time changes, festivities, weather conditions etc. We let $\hat{\theta}(k)$ be the estimate of $\theta^*(k)$, which is updated at each time step k based on the available hydraulic measurements.

At discrete time k_s the leakage fault detection algorithm is activated; the algorithm is comprised of two parts: a) the on-line flow estimator and b) the leakage fault detection logic component.

Since the weekly periodic signal $s(k) = \theta^*(k)^\top \zeta(k)$ is time-varying, it is useful to compute and update in time the parameter vector $\hat{\theta}(k)$, so that $\hat{s}(k) = \hat{\theta}(k)^\top \zeta(k)$ is an estimation of the unknown $s(k)$.

As approximation structure, we consider a linearly parameterized Fourier series with N_z terms, where $\hat{\theta}(k)$ is the parameter estimation vector of size $2N_z + 1$, and, for $\omega_s = \frac{2\pi}{T_S}$, $\zeta(k)$, and T_S the number of samples within a weekly period, the regressor vector is given by

$$\zeta(k) = [1, \cos(k\omega_s), ..., \cos(N_z k\omega_s), \sin(k\omega_s), ..., \sin(N_z k\omega_s)]^\top. \tag{5.8}$$

The normalized flow estimation error $e_r(k)$ is given by $e_r(k) = q_r(k) - \hat{s}(k)$. The error $e_r(k)$ is useful in computing the parameter vector estimate for the next time instance, such as

$$\hat{\theta}(k+1) = \hat{\theta}(k) + A_G \frac{\zeta(k)}{a_F + \zeta(k)^\top \zeta(k)} e_r(k) \tag{5.9}$$

where a_F is a design parameter and A_G a gain diagonal matrix, which are selected offline. To guarantee the learning law stability, $a_F \geq 0$ and $0 < A_G^i < 2$, for the i-th diagonal element of the gain matrix [7, p.54].

In practice, the parameters in $\hat{\theta}(k)$ may drift towards infinity. This effect may be alleviated by using the projection algorithm, which bounds the parameter updates within a pre-defined region $\overline{\Theta}$ [70].

The leakage fault detection algorithm is activated at time k_s, and the initial conditions of the estimated parameter vector are $\hat{\theta}(k_s) = \theta^0$. The initial conditions vector can be computed off-line, by taking into consideration a set of historical measurements. Let $\mathcal{K}_p$ be the set of the most-recent time instances set prior to the time the leakage fault begins; by solving a least-squares optimization problem we compute the initial condition vector estimation

$$\hat{\theta}(k_s) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{k \in \mathcal{K}_p} \left(q_r(k) - \theta^\top \zeta(k)\right)^2, \tag{5.10}$$

where $q_r(k) = \frac{q(k)}{\hat{r}_t(k;\hat{\theta}_r)}$ is the normalized flow. Note that since the flow is normalized with respect to the estimated trend signal, the average value of the Fourier series approaches the unit value.

The update law (5.9) can be used to learn the changes in the consumption dynamics. However, in addition to the consumption, the update law can learn the unknown leakage fault dynamics which begin at discrete time k_0 the leakage fault begins In specific, the leakage fault acts as a positive offset to the flow signal and may exhibit periodicity in accordance to the periodicity of the pressure signal; the update law will change the elements of parameter vector estimator, to approximate the new flow characteristics and reduce the estimation error, which corresponds to learning the leakage faults.

Note that when certain known exogenous factors affect the periodic consumption, the update law can be suitably modified to capture the prior knowledge. For example, in some

countries the daylight savings time change occurs in spring and ends in autumn, by advancing the clock time by 60 minutes; as a result, all the social and economic patterns are shifted by 60 minutes. Unless this is taken directly into consideration, time change may appear as a large error and may trigger the leakage fault detection algorithm. The update law (5.9) and the estimator $\hat{s}(k)$ can be modified; for example, when the daylight savings time change occurs, the regressor $\zeta(k + \frac{60}{\Delta t})$ is used instead of $\zeta(k)$.

5.3.1 Leakage Fault Detection Logic

Detection of a leakage fault is performed by monitoring for change in the mean value of the parameter $\hat{\theta}_0(k)$, which corresponds to the offset DC term of the Fourier series. A suitable sequential analysis algorithm for the detection of changes in the mean value of a signal is the Cumulative Sum control chart (CUSUM) [10, 18]. The assumption behind using CUSUM is that the monitored signal has a constant average value during normal operation and that the magnitude of the change is considered known [9].

Let $S(k)$ be the cumulative sum for the k-th time instance, where $S(0) = 0$; for $k > 1$ and for a positive change in the mean due to leakage faults, the one-sided CUSUM algorithm is given by

$$S(k) = \max\{0, S(k-1) + \hat{\theta}_0(k) - \overline{\theta}_0 - \gamma_0\}, \tag{5.11}$$

where $\overline{\theta}_0 = 1$ corresponds to the theoretical average of the offset DC term and $\gamma_0 > 0$ is a parameter defined by the designer corresponding to the expected change. A detection alarm is triggered at the time instance k_d at which the metric $S(k)$ is greater than the threshold $\overline{h}_s$, such that

$$k_d = \min\{k \mid S(k) \geq \overline{h}_s\}. \tag{5.12}$$

An estimation of the leakage fault start time $\hat{k}_0$ can be computed as the latest time instance before detection when the metric $S(k)$ was less than a small positive number ϵ_k, such as

$$\hat{k}_0 = \max\{k \mid S(k) <,\ k < k_d\}. \tag{5.13}$$

Part of leakage fault diagnosis is to identify the magnitude of the leakage. The large uncertainties in the measurements, however, impose difficulties in the task. To compute an estimate of the average leakage outflow $\overline{\phi}(k)$, when a leakage fault is detected at time k_d and the leakage start time is estimated at $\hat{k}_0$, we use

$$\overline{\phi}(k) = \frac{1}{k - \hat{k}_0 + 1} \sum_{i=\hat{k}_0}^{k} \hat{\theta}_0(i), \tag{5.14}$$

which is computed iteratively at each discrete time k.

5.4 Leakage Fault Detection Using Night Flow Analysis

In practice, utility operators use night-flow monitoring to determine the presence of leakage faults. This is usually done by comparing the average of the minimum night flow with those of the previous days. However, the trend and seasonal signal are not taken into account, and the decision may be subjective. Below, we formulate a leakage fault detection algorithm based on the average night flow normalized with respect to the trend and seasonal signal. This will also serve as a baseline comparison to the adaptive approximation algorithm developed in the previous section.

In this section we formulate a leakage fault detection methodology based on the night-flow analysis methodology, which takes into consideration the DMA water inflow during low consumption hours. The intuition behind using the night-flows is that since the flows during the night have smaller variations than during the day, and since the leakage losses will be larger because of the higher pressures in the system, it will be easier to detect leakage faults.

Let $w(l)$ be the average night-flow measured for the l-th period, which corresponds to 24-hours. Let Δt be the sampling time (in minutes) for measuring the flow and let $T_H = \frac{24 \cdot 60}{\Delta t}$ be the number of samples in one day. Let t_a and t_b be the discrete times at which the night-flow begins and ends. Considering that the first discrete time $k = 0$ corresponds to midnight of period $l = 0$, we define $\mathcal{N}_f(l) = \{t_a + lT_H, t_a + 1 + lT_H, ..., t_b + lT_H\}$ as the set of discrete times corresponding to night-flows of the l-th day. In general, t_a and t_b are constant; however, in some cases (e.g. due to the daylight savings time change) they may need to change.

At the l-th day and after the night-flow period has finished, we compute the normalized average night flow $w(l)$ as

$$w(l) = \frac{1}{|\mathcal{N}_f(l)|} \sum_{i \in \mathcal{N}_f(l)} \hat{q}_r(i), \tag{5.15}$$

where $\hat{q}_r(\cdot)$ is the normalized flow with respect to the estimated trend signal $\hat{r}_t(k; \hat{\theta}_r)$, such that $\hat{q}_r(k) = \frac{q(k)}{\hat{r}_t(k;\hat{\theta}_r)}$.

We further define $\delta_w(l)$ as the difference of the average night flow $w(l)$ with the minimum average night flow of the previous $M_w \geq 1$ days, such that

$$\delta_w(l) \quad = \quad w(l) - \min\{w(l - M_w), ..., w(l - 1)\}. \tag{5.16}$$

Let l_d be the day a leakage fault is detected; a leakage detection alarm is activated at the l-th day, such that $l_d = l$, when the difference $\delta_w(l)$ has a greater value than the detection threshold $\overline{h}_w$, which is selected off-line by using historical measurements. This approach gives rise to certain issues: setting $\overline{h}_w$ too low may cause a large number of false positive leakage fault alarms, while setting it too high may cause the detection algorithm to miss some leakage faults. In addition, due to the large uncertainties in the flow measurements, e.g. due to festivities or other events, this threshold could be exceeded even when no leakage fault has occurred in the system.

We can modify this detection method, to consider a window of difference measurements, so that detection occurs when, for at least M_d days, the difference $\delta_w(\cdot)$ computed for each day within that period of time is greater than a certain threshold $\overline{h}_w$. We further consider that the estimated leakage start day $\hat{l}_0$ is given by $\hat{l}_0 = l_d - M_d$.

The detection threshold $\overline{h}_w$ can be selected off-line using historical data. We select a minimum detection window M_0, such that $1 \leq M_0 < M_d$. From a set of historical measurements, we compute the differences $\delta_w(\cdot)$ for each day in the dataset using the M_w window, and then, we compute a value for $\overline{h}_w$ so that there are no more than M_0 consecutive measurements, for which $\delta_w(\cdot)$ is greater than the selected $\overline{h}_w$.

Apparently, if a leakage fault has not been detected after M_w days from its day of occurrence, the algorithm will consider the previous average night flow measurements as normal, and may not be able to detect the leakage fault in the future. This is a major drawback in the proposed leakage fault detection methodology, which will be demonstrated using a simulation example.

5.5 Simulation Examples

In this section we present simulation results by applying the leakage fault detection solution methodology on historical hydraulic data taken from a DMA in Limassol, Cyprus, corresponding to the period of 426 days between 1/11/2006 and 31/12/2007. The hydraulic data were collected with a five-minute sampling time $\Delta t = 5$; a daily period is comprised of $T_H = 288$ samples. A subset of the historical data were considered to compute the trend and seasonal signal estimation $\hat{r}_t(k; \hat{\theta}_r)$, by solving the least-squares optimization problem (5.6) for finding the coefficients $\hat{\theta}_r$ of a two-term Fourier series. In specific, for $\mathcal{K}_h$ the set of discrete times corresponding to 241 days, estimated trend and seasonal signal is given by a Fourier series whose coefficients are estimated as $\hat{r}_t(k; \hat{\theta}_r) = 32.12 - 0.79\cos(\omega k) + 0.06\cos(2\omega k) -$

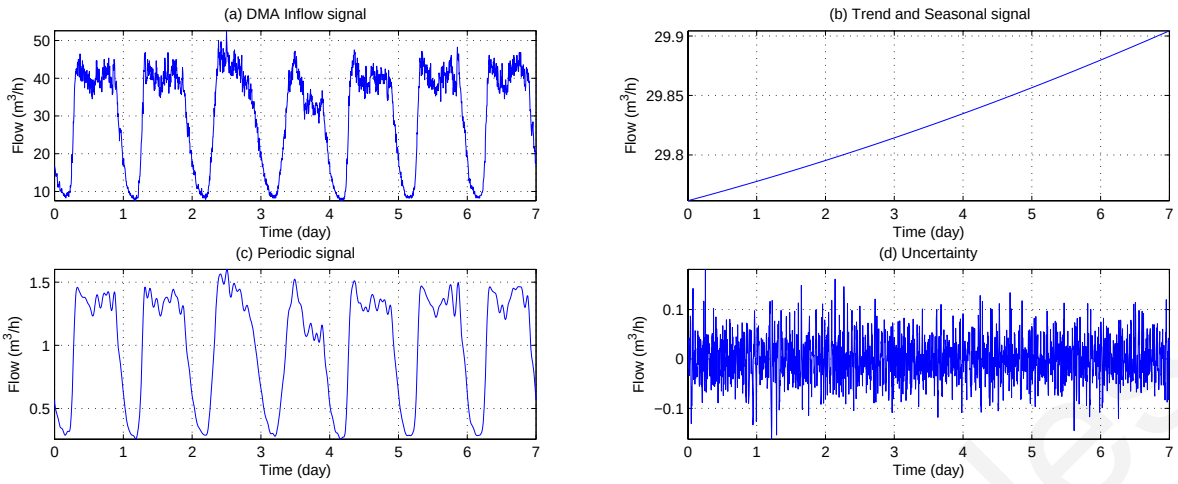


Figure 5.1: Decomposition of the flow signal for the time span between 1st March 2007 and 7th March 2007.

$2.23\sin(\omega k) + 0.87\sin(2\omega k)$, where $\omega = \frac{2\pi\Delta t}{365\cdot 24\cdot 60} = 5.9772 \times 10^{-5}$.

5.5.1 Nominal Flow Signal Decomposition Example

To illustrate the DMA inflow signal decomposition, we consider the historical hydraulic data of one week without leakage faults, taken from a DMA in Limassol, Cyprus, between 1/3/2007 and 7/3/2007; for $\Delta t = 5$ this period corresponds to 2016 samples. Let $\mathcal{K}_w$ be the set of discrete times corresponding to that week. The original signal is depicted in Fig. 5.1a and the estimated trend and seasonal signal for that week is depicted in Fig. 5.1b.

The weekly periodic function does not change significantly within the period we examine; therefore, we compute the coefficient vector $\hat{\theta}$ of the weekly periodic approximation function $\theta^\top \zeta(k)$ by solving the following optimization problem

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{i\in\mathcal{K}_w} \left(\frac{q(i)}{\hat{r}_t(i;\theta_r)} - \theta^\top \zeta(k) \right)^2. \tag{5.17}$$

The approximation signal which corresponds to a Fourier series with 100 terms is depicted in Fig. 5.1c. Finally, we compute the uncertainty estimation, which is given by

$$\hat{n}(k) = \frac{q(k)}{\hat{r}_t(k;\theta_r)\hat{\theta}^\top \zeta(k)} - 1; \tag{5.18}$$

and is depicted in Fig. 5.1d. The estimated uncertainty follows a normal distribution (as verified by the Lilliefors normality test [103]) with zero mean and variance $\sigma_n^2 = 0.0019$.

5.5.2 Outlier Handling

Outliers in the data may affect the leakage fault detection procedure; for this reason, it is useful to apply an on-line low-pass filter to remove these measurements. At each discrete time k, we compute the DMA inflow median $q_\mu(k)$ of the last N_μ flow measurements, then we calculate the absolute difference between the median $q_\mu(k)$ and the measured DMA inflow $q(k)$ and compare it with an outlier detection threshold $\overline{\delta}(k)$. The measurement $q(k)$ is considered to be an outlier if the absolute difference $|q_\mu(k) - q(k)|$ is greater than a detection threshold $\overline{\delta}$, such that $|q_\mu(k) - q(k)| > \overline{\delta}$; otherwise, it is considered as a healthy measurement. In case $q(k)$ is considered an outlier measurement, it is replaced by the median value $q_\mu(k)$. As detection threshold $\overline{\delta}$, we can use a constant value which is selected offline using historical data, or we can use the media absolute deviation (MAD) metric (the median of the last N_μ absolute differences) multiplied by a constant positive value [111, 125].

5.5.3 Leakage Detection Using Night-Flow Analysis

We consider that the minimum night flow occurs between 2A.M. and 5A.M. during the standard time period, thus $t_a = 24$ and $t_a = 60$ are the discrete times the night-flow period begins. For the daylight savings time period, however, we consider that $t_a = 12$ and $t_a = 48$. The set of night-flow discrete times corresponding to the l-th day, during standard time, is given by $\mathcal{N}_f(l) = \{24 + 288l, 25 + 288l, ..., 60 + 288l\}$.

Figure 5.2a depicts the average night-flows computed using the normalized flows for the first 242 days of the dataset. Even though the signal is normalized over the trend and seasonal signal, the uncertainty and the extreme values in the signal pose difficulties in detecting leakages by using simple observation of the time-series from a human operator. Figure 5.2b illustrates this issue, which depicts the normalized flow with a leakage fault at day $l_0 = 115$, with average leakage outflow $0.5\ \frac{m^3}{hr}$.

Note that the extreme value $w(61) = 0.46$ appears to be due to Christmas day festivities. If a leakage fault occurred in the network at the l_0-th day, it would correspond to an increase in the average night-flow measurements; in this example, a leakage fault with outflow $0.5\ \frac{m^3}{hr}$ would correspond to an average increase of 0.015 in the signal $w(l)$ for $l \geq l_0$, which is relatively small with respect to the magnitude of the noise.

The difference-based leakage fault detection algorithm is demonstrated; each day we compare the normalized flow to the previous $M_w = 7$ normalized flow measurements. We consider that a leakage fault is detected if the difference has been greater than a certain threshold

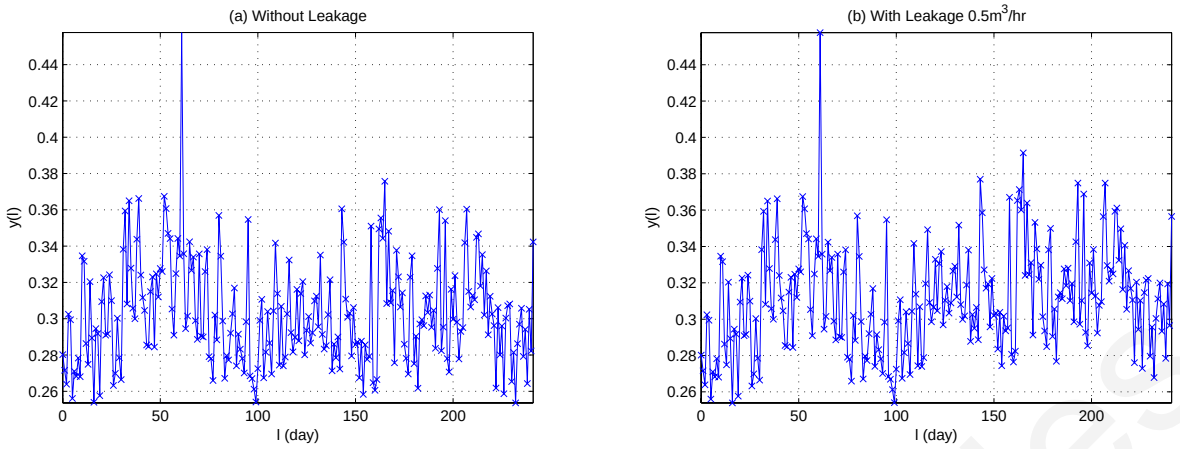


Figure 5.2: Average night-flow measurements for the first 242 days in the dataset, a) without leakage faults and b) with a leakage fault of $0.5\ \frac{m^3}{hr}$.

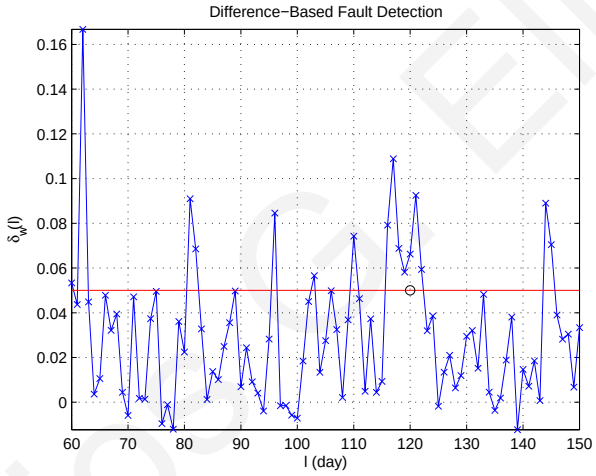


Figure 5.3: The leakage fault which occurs at $l_0 = 115$ is detected at time $l_d = 120$, when the difference signal $\delta_w(l)$ is greater than the threshold $\overline{h}_w = 0.05$ for more than $M_d = 5$ days.

for more than $M_d = 5$ days. We use the first 100 days of recorded data to compute the detection threshold; for $M_0 = 4$ the minimum detection window, we solve an optimization problem and compute the detection threshold as $\overline{h}_w = 0.05$. To demonstrate the leakage fault detection using night-flows, we simulate a leakage fault starting at day $l_0 = 115$ with constant outflow $1.5\ \frac{m^3}{hr}$. The computed differences $\delta_w(\cdot)$ for each day are shown in Fig. 5.3. In this example, the leakage fault is detected at day $l_d = 120$, and therefore the estimated leakage fault start day is $\hat{l}_0 = l_d - M_d = 115$, which coincides with the actual start day l_0.

Through simulations and for the parameters chosen, we compute that the smallest leakage fault detectable corresponds to the outflow $1.3\ \frac{m^3}{hr}$, which is a relatively large value when compared to other leakages with outflow greater than $0.5\ \frac{m^3}{hr}$. In addition, if the leakage detection algorithm fails to detect the leakage fault within the specified time window, it will eventually

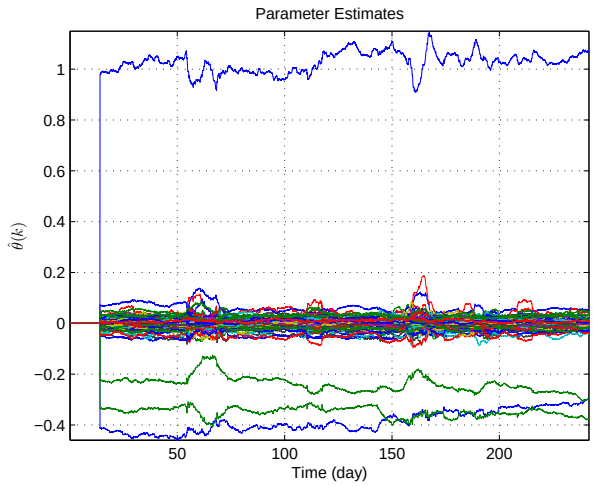


Figure 5.4: Updates of the Fourier parameters; a leakage fault with average outflow 1.5 $\frac{m^3}{hr}$ starts during the 115-th day, and the Fourier parameters are updated to learn the changes.

consider the flow measurement after the leakage as normal, i.e. below the detection threshold, as can be seen in Fig. 5.3 for times $l \geq 123$. Some of these limitations can be addressed with the leakage fault detection methodology demonstrated in the following paragraphs.

5.5.4 Leakage Detection using Adaptive Demand Flow Approximation

In this example we demonstrate the adaptive demand flow approximation methodology for leakage fault detection. We consider the use of a Fourier series with $N_z = 100$ terms and, to approximate the weekly periodic signal, $\omega_s = 0.0031$ for a weekly period comprised of $T_S = 2016$ samples. By using 4032 recorded hydraulic measurements (corresponding to the first two weeks of the dataset prior the fault detection algorithm start time k_s), the least-squares optimization problem (5.10) is solved to compute the initial conditions for the parameter vector $\hat{\theta}(k_s)$.

For the update law (5.9) we consider that $a_F = 0.01$ and $A_G = \text{diag}(0.1, ..., 0.1)$. Figure 5.4 depicts the evolution of the parameter vector estimate $\hat{\theta}(k)$ in time, when a leakage fault with average outflow 1.5 $\frac{m^3}{hr}$ occurs during the 115-th day ($k_0 = 33\ 119$). The Fourier coefficients are updated in time to approximate the changes in the consumption patterns. We consider that the signal is normalized with respect to the trend and seasonal signal, $\hat{r}_t(k; \hat{\theta}_r)$; therefore, we expect that for leakage faults with average outflows of $\{0.5,\ 1.0,\ 1.5\}$ $\frac{m^3}{hr}$, the expected change in the average value of offset DC term $\hat{\theta}_0$ from the Fourier coefficients will be approximately $\{0.015,\ 0.031,\ 0.047\}$ respectively. In the case of no leakage faults, the standard deviation of $\hat{\theta}_0$ is $\sigma_{\theta_0} = 0.031$, which is comparable to the magnitude of the leakage faults.

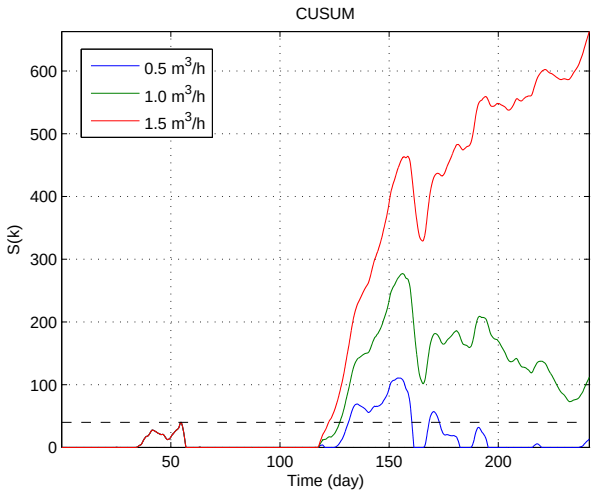


Figure 5.5: CUSUM metric for three leakage faults with average outflow $\{0.5, 1.0, 1.5\}$ $\frac{m^3}{hr}$ which occurs during the 115-th day. The dashed line corresponds to the detection threshold $\overline{h}_s = 40$.

The CUSUM fault detection logic is applied to detect the existence of a leakage fault. The selection of the parameter γ_0 depends on the expected change in the average value, e.g. $\gamma_0 \in [0.01, 0.05]$. We select the parameter $\gamma_0 = 0.03$ for the CUSUM metric. By considering the historical data from the real data and through simulation, we select the leakage fault detection threshold as $\overline{h}_s = 40$. The smallest the threshold, the more sensitive the detection algorithm is; however it may trigger false alarms. As a special case we consider the event of festivity periods, where this threshold may be violated due to increased consumption; we propose considering different detection thresholds for different time periods which may correspond to known increased water demands.

Figure 5.5 depicts the CUSUM metric for three leakage fault scenarios with average leakage outflow $\{0.5, 1.0, 1.5\}$ $\frac{m^3}{hr}$, which occurs from the 115-th day ($k_0 = 33\ 120$). Note that the positive CUSUM signal at times around the 50-th day might be explained due to festivities. For the leakage with 1.5 $\frac{m^3}{hr}$ outflow, detection occurs after 7 days, whereas for the leakages with 1.0 $\frac{m^3}{hr}$ and 0.5 $\frac{m^3}{hr}$ outflows, detection occurs after 12 and 16 days respectively. For all scenarios, the estimated leakage start time is during the 117-th day, for $\epsilon_k = 0$.

In Fig. 5.6 the estimates of the average outflow of the leakage faults are depicted, as computed for each time instance after the estimated start time $\hat{k}_0$ (the 117-th day). We observe that the estimation does not immediately converge to values near the actual average outflows, and is affected by an irregular event for the period within the days 125 and 165, which appears as increased leakage whereas it is actually increased consumption. In time, the estimations approach the actual average outflows; in general, this can only be used as a rough estimation

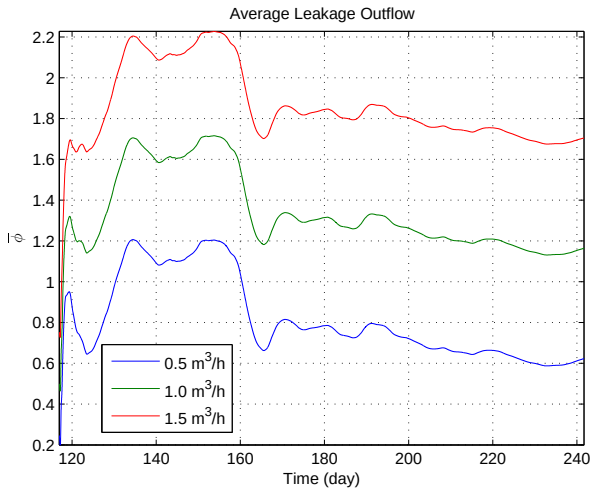


Figure 5.6: The average leakage outflow estimates for three leakage faults with average leakage outflow of $\{0.5,\ 1.0,\ 1.5\}\ \frac{m^3}{hr}$.

of the leakage magnitude.

5.6 Concluding Remarks

In this chapter we have addressed the problem of leakage detection in a DMA whose model is not known and for which the inflow flow and pressure are sampled every few minutes. A mathematical model of the DMA inflow and leakages is presented, which is based on several components: a linear trend, a yearly seasonal, a weekly periodic and an uncertainty component, along with the leakage fault signal. An optimization problem is formulated for computing the trend and seasonal signal over a set of historical flow measurements. To solve the leakage detection problem, we proposed an adaptive demand flow approximation methodology, in which an adaptive law is used to update the parameters of a Fourier series which learns the changing weekly periodic consumption dynamics, as well as the leakages.

The method proposed was able to detect leakages of various sizes; the time of detection, however, could vary depending on the leakage fault's magnitude. In addition to the proposed methodology, we discussed the use of night-flow analysis for leakage detection. Simulation results using the two methodologies are presented; the results indicate that the night-flow analysis is suitable for large leakage faults; however, its ability in leakage detection is reduced if the leakage fault has not been detected within a certain time.

The contribution of this work is the design and implementation of an automated leakage detection algorithm, which utilizes measurements which are already available in some DMA, within a adaptive approximation-based framework. A journal paper based on the results pre-

sented in this chapter is in preparation [56].

Chapter 6

Contamination Fault Isolation

In this chapter we propose a computational methodology to select a sequence of nodes to perform expanded sampling. The goal is to evaluate the water contamination impact, and isolate the source-area of the contamination, with as few samples as possible. We consider that the water utility has a number of fixed quality sensors installed in the distribution network, and that manual quality sampling can be conducted by a contamination response team at any feasible location in the distribution network. After the triggering of a contamination alarm by a quality sensor, and upon its verification as an actual contamination by the utility operator, a manual sampling scheduling scheme is computed. The scheduling scheme gives guidelines with regard to which nodes the contamination response team should sample, in order to isolate the source-area and to evaluate its possible impact, as quickly as possible. The proposed method is based on decision tree induction; the conditional terms of the decision tree indicate where expanded manual sampling should be conducted, with a certain order, aiming at evaluating the possible fault-impact and at isolating the source region. To illustrate the solution methodology, we present results based on a simplified network and a real water distribution system benchmark.

6.1 Background

Water security is a challenging task and has become a vital part of the drinking water distribution operation. To provide a framework for water security, the US Environmental Protection Agency (EPA) has published guidelines for water utilities describing a *contamination warning system* architecture comprised of the standard operational procedures and the *consequence management plan*. The guidelines provide information on the qualitative and quantitative parameters that need to be monitored, as well as on the response actions to be taken

from the moment a contamination fault alarm has been triggered until the system has been disinfected and returned to normal operation [170, 171]. Those actions include: a) monitoring and surveillance of the system; b) event detection and determination if the contamination fault is "Possible" (i.e. if there are no strong indications of a false alarm); c) determination if the contamination fault is "Credible" (i.e. evaluating field results from the area where the contamination fault has occurred), d) determination if the contamination fault is "Confirmed" (i.e. evaluating laboratory results from multiple samples); e) implementing remediation and system recovery [170].

Part of the confirmation operation plan is to develop and implement "Expanded Sampling", i.e. manual sampling at other parts of the distribution network, so as to determine the extent of the contamination, which is useful for the response and remediation task [171]. The EPA recommends the use of hydraulic models of the water distribution system to determine where to sample and to evaluate the spread of the contamination. The use of hydraulic models can reduce the time required to plan expanded sampling and assist in understanding the contaminant propagation path; in addition, operators can issue targeted restriction notices as necessary [171].

Choosing where to perform expended sampling can be a challenging task, due to the large-scale nature of the distribution network and the partially unknown hydraulic dynamics. In practice, water utilities may choose beforehand in an *ad hoc* manner certain locations in the network, where expanded sampling should be conducted in case of a contamination fault detection, e.g. at tanks, reservoirs and pumps. Sampling at these locations, however, may not provide adequate information regarding the possible impact due to a contamination fault, or the location in the distribution network where the contamination has originated from. In addition to these predetermined locations, a utility operator may select additional nodes in the network to conduct manual sampling. To reduce the bias due to subjective decisions when selecting manual sampling locations, especially in large-scale networks, it is desirable to utilize a computational method for finding an optimal sequence of nodes where manual sampling should be conducted, taking into account the actual quality and hydraulic measurements, as well as the physical network model of the system. Such a computational method could assist in reducing the reaction time required to take appropriate response measures, and to make more informed decisions, especially when there is a shortage of time and personnel.

The general problem of contamination source isolation has received some research attention in the last few years. Analytical problem formulations for the contamination source isolation problem have been presented in the framework of optimization [71, 79, 84, 94–96,

110, 163, 173]. A graph-based method was presented in [46], for the calculation of a set of possible node locations for a certain contamination fault.

The use of the *particle backtracking* algorithm [149] in isolating the location of a contamination source was examined [48, 49]. When a contamination is detected at a certain node by a quality sensor, the flows are tracked backwards in time, in order to compute all the possible input locations and injection time, which would in turn assist into computing an area of where the contamination fault may have originated from. In [121] a random contamination scenario set was considered, aiming at estimating the most probable-for-injection system nodes, the approximated injection starting time, their duration and the injected mass rate.

Computational and artificial intelligence techniques have been investigated, as well, such as hybrid model-trees [132] and genetic algorithms [131], which compute near-optimal solutions for the contamination source location, starting time, injection duration and injected mass rate. In general, various approaches have been considered, such as evolutionary algorithms [92, 104, 105, 155, 174, 193] and probabilistic methodologies [86].

Most of the methodologies presented in previous research require the use of fixed quality sensor measurements which would solve an optimization problem, and which would compute some possible contaminant injection locations or input time signals, given some known network structure and approximately known hydraulic dynamics. Due to the large scale of water distribution systems, in practice, many quality measurements are required to accurately isolate the fault. Some practical difficulties arise, such as the fact that water utilities may have installed only a limited number of fixed quality sensors, placed at specific locations after having solved the sensor-placement problem [123]. In general, the goal of the sensor-placement problem is to optimize certain impact-risk objectives; the solution corresponding to this problem, however, may not be optimal from the fault isolation perspective, since the locations of the installed sensors are usually such that the network coverage is maximized. Therefore, it is possible that the sensor measurements do not provide adequate information for fault isolation. Moreover, large time delays may exist between contamination detection at different nodes in the network, thus delaying the contamination source isolation procedure.

In the present work, we examine the problem from a different perspective. We consider that the water utility has a number of fixed quality sensors installed in the distribution network, after having solved the sensor placement problem for the minimization of the impact-risk. In addition, manual quality sampling can be conducted by a contamination response team at any feasible location in the distribution network. After a contamination alarm has been triggered by a quality sensor, and after its verification as an actual contamination by the

utility operator, a manual sampling scheduling scheme is computed. The scheduling scheme gives guidelines to the utility operator with regard to which nodes the contamination response team should deploy, in order to isolate the source-area and to evaluate its possible impact, as promptly as possible. As added value, this work is in accordance with the EPA guidelines, within the scope of conducting expanded sampling during the "Confirmed" determination stage of the contamination warning system [171]. The proposed method is based on decision tree induction; decision trees are hierarchical data structures which can be expressed using conditional statements, such as *if-then-else* rules. In this work, they are utilized to indicate where expanded manual sampling should be conducted, with a certain order, so as to evaluate the possible fault-impact and to isolate the source region.

This chapter is organized as follows: the problem formulation is described, key concepts and definitions are introduced and the solution methodology is presented. Simulation results are demonstrated on an illustrative and a real benchmark network to verify the proposed algorithm.

6.2 Problem Formulation

6.2.1 Network and Contamination Modeling

The network of water distribution system can be modeled as a directed graph $(\mathcal{V}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, ..., N_m\}$ is the set of N_m node indices which correspond to pipe junctions and water storage tanks, and $\mathcal{A}$ is the set of pipe edges which connect two nodes.

Typically, hydraulic signals in a water distribution system are measured in discrete time; in the following, we set k as the discrete time, and Δt the sampling time step. At discrete time k, $d_i(k)$ is the consumer demand outflow at the i-th node, and $b_j(k)$ is the water velocity in the j-th pipe. We consider that certain network parameters are available, such as pipe lengths, diameters and roughness coefficients, as well as node elevations, average daily base demands and, where available, consumer types.

An algorithm which solves the hydraulic equations of the distribution network may be used to compute the water velocity vector $b(k)$, considering the demands $d(k)$, the available network parameters, as well as the hydraulic control laws and the initial water levels in tanks. This algorithm may correspond to the hydraulic solver utilized in the EPANET software [142]. We should note that the EPANET solver in some cases can only compute a solution of the hydraulic dynamics by using a fraction of the time step Δt defined; this additional information,

when available, can be taken into consideration in the solution methodology.

We define $\mathcal{V}_s \subset \mathcal{V}$ to be the subset of node indices where fixed online quality sensors have been installed, after having solved the sensor placement problem [123], and let $\mathcal{V}_m \subseteq \mathcal{V}$ be the node indices subset where manual sampling can be performed. In general, some nodes may not be available for manual sampling, whereas some nodes with fixed sensors may require additional sampling.

We assume that contaminants can be injected at certain nodes, and propagate along with the water flows. Let $x_i(k)$ be the average contaminant concentration within the i-th node, at discrete time k, where $i \in \mathcal{V}$. In addition, we define the signal $\phi^i(k-k_0)$ which describes the increase in the contaminant concentration, when a contaminant is injected at the i-th node, starting from discrete time k_0. It is possible that multiple contaminations may be initiated, affecting multiple nodes; in this work we consider single-source contaminations. The time profile of the contaminant injection fault function $\phi^i(k-k_0)$, and in specific, the contaminant injection location, its start time, as well as the structure of signal, are considered, in general, unknown.

A contamination may occur at any node, at any time, with any rate and time profile. Let $\mathcal{S}^*$ be the set of all contamination fault functions which may affect a certain water distribution system, such that

$$\mathcal{S}^* = \left\{ \phi^i(k-j) \mid (i,j) \in \mathcal{J}^*,\ \phi \in \mathcal{F}^* \right\}, \tag{6.1}$$

where (i,j) corresponds to the i-th node from $\mathcal{V}$ and the j-th discrete time at which a contamination fault occurs, and $\mathcal{J}^*$ is the set of all pairs corresponding to node indices and start times. In addition, $\mathcal{F}^*$ is the set of all contaminant injection signal structures (such as unit step functions or rectangular functions).

For example, for $(i,k_0) \in \mathcal{J}^*$ the pair corresponding to the i-th node and to the injection start time k_0 , and for $\phi(\cdot) \in \mathcal{F}^*$ a rectangular pulse structure with magnitude a_ϕ and duration τ_d, the corresponding contamination fault function is given by $\phi^i(k-k_0) = a_\phi$ for $k_0 \leq k \leq k_0 + \tau_d$, and $\phi^i(k-k_0) = 0$ otherwise.

6.2.2 Contamination Impact Modeling

It is equally important to consider the impact dynamics, which characterize the amount of "damage" caused by a contamination fault function $s \in \mathcal{S}^*$ affecting a demand node at each time step, measured with a certain impact metric (e.g. number of people affected) [61]. Let $\xi_i(k;s)$ be a state describing the impact damage caused on the i-th node, at discrete time k due

to contamination fault function s, assuming zero initial conditions. The state-space equations describing these dynamics can be formulated, for $i \in \mathcal{V}$ as

$$\xi_i(k+1;s) = \xi_i(k;s) + f_\xi(x_i(k), d_i(k)) \tag{6.2}$$

$$\psi(k;s) = f_\psi(\xi(k;s)), \tag{6.3}$$

where $\xi(k;s)$ is the nodal impact vector for $s \in \mathcal{S}^*$, $f_\xi(\cdot)$ is the impact function which depends on the nodal water consumption $d_i(k)$ and contaminant concentration state $x_i(k)$ at the i-th node, at discrete time k. In addition, $\psi(k;s)$ is a measure of the overall impact at discrete time k for the contamination fault function $s \in \mathcal{S}^*$, and $f_\psi(\cdot)$ is the overall-impact function which depends on the nodal impact vector. For example, $\xi_i(k;s)$ may correspond to the number of people affected due to water consumption at the i-th node, and $\psi(k;s)$ to the total population affected.

6.2.3 Node Sampling Decision

When a contamination is detected, the general objective is to identify the severity of its impact, as well as to isolate some area in the network where the contamination may have originated from. We formulate the problem of iteratively selecting nodes which correspond to the sampling node set $\mathcal{V}_m$ to conduct manual sampling, and to examine whether or not contaminant traces exist at those nodes. In general, more than one response team may be available to be deployed for manual sampling at different locations. In this work we formulate the problem considering the availability of one response team; however, the formulation can be extended to the multiple response team case.

At the j-th manual sampling iteration, the problem is formulated as

$$\mu^j = f_\mu(\mu^{j-1}, \nu^{j-1}; \mathcal{S}^{j-1}) \tag{6.4}$$

where $\mu^j \in \mathcal{V}_m$ is a node where manual sampling should be conducted, and $\nu^j \in \{0,1\}$ is a contamination flag, such that $\nu^j = 1$ if contaminant traces have been detected at node μ^j, $\nu^j = 0$ otherwise. Function $f_\mu(\cdot)$ computes the node where manual sampling should be conducted in the next manual sampling iteration, and $\mathcal{S}^j \subset \mathcal{S}^*$, for $j \geq 0$, is the subset of the contamination fault functions which could have caused the contamination fault, according to the sensor measurements and manual sampling results. When a contaminant is first detected at the i-th node, for $j = 0$, $\mu^0 = i$, $\nu^0 = 1$; methods to compute the set $\mathcal{S}^0$ are discussed in the solution methodology section.

6.3 Solution Methodology

In the next paragraphs we present an algorithm for $f_\mu(\cdot)$ in (6.4), for the calculation of nodes at which to perform manual sampling, according to a set of contamination faults which could have caused the detected contamination fault, and a logic variable whether or not contaminant traces have been detected at certain nodes. The solution methodology presented in this work is based on *Decision Trees*, a decision support tool, typically used in classification problems, by using sequential measurements of some variables to assign a class on these data. The classification problem in this work can be expressed as the problem of determining the risk-level of a contamination event detected at a certain node, e.g. whether it belongs to a high-risk class (with severe consequences) or to a low-risk class (with minor or no consequences).

In our work, the selection of the best location for manual sampling is performed with respect to two objectives: a) the number of contamination fault functions which could have caused the detected contamination fault, and b) the calculated contamination impact. These objectives may be conflicting, and it is possible that no single solution is optimal for both objectives; instead, a Pareto front of solutions may be computed, and from that, a single solution is selected.

6.3.1 Possible Contamination Fault Functions

It is not necessary to compute the complete set $\mathcal{S}^*$ of all possible contamination fault functions, as it is extremely large; instead, we may compute a set of possible contamination fault functions, $\mathcal{S}^0 \subset \mathcal{S}^*$, which is given by

$$\mathcal{S}^0 = \left\{ \phi^i(k-j) \mid (i,j) \in \mathcal{J}^0,\ \phi \in \mathcal{F} \right\}, \tag{6.5}$$

where $\mathcal{F} \subset \mathcal{F}^*$ is a finite set of fault structures determined by the utility operators, and $\mathcal{J}^0 \subset \mathcal{J}^*$ is a finite set of pairs of node indices and start time where and when contamination could occur. For instance, rectangular pulse functions of various durations and unit-step functions may be considered as fault structures in the set $\mathcal{F}$; the set $\mathcal{J}^0$ may be computed using a suitable algorithm, discussed in the following paragraphs.

We should note that in the methodology proposed in this work, we consider that there are algorithms available for computing a set of possible network nodes and times where the contamination may have occurred. In this work, we use such results to construct and solve the manual sampling scheduling problem.

One example of an algorithm which computes the possible contaminant injection nodes

and times, is the *particle backtracking* algorithm [149]. In summary, for some concentration measured at a certain sensor node and time, the algorithm computes the reverse propagation path of the contaminant, so as to identify the network nodes where the contaminant had previously appeared, and with what concentration. This method was implemented by [49] for contamination detection. For the solution methodology discussed in this work, a modified *particle backtracking algorithm* can be used to compute the previous nodes and arrival times at them.

A crucial assumption taken by the particle backtracking algorithm is that the flows in all pipes, as well as tank water levels of the past few days, are known. In reality, this is difficult to guarantee, since only some flows and pressures are measured on-line; to compute the flows which are not measured, a calibrated network model is required, as well as estimations of the demand flows at each node in the network. An example of how to calibrate a network model of a real distribution network and how to model its quality dynamics, is presented in [107, 108]. Calibrating the hydraulic dynamics of a network model requires field data gathering to collect hydraulic measurements, such as pressures at various locations in the network. In addition, field data from tracer studies are used to calibrate the quality dynamics. Furthermore, historical data of total water demands are analyzed to determine long-term trends throughout the years, seasonal effects within a year, as well as weekly and daily changes; in practice, such demands may change significantly in time [107, 108]. For the solution methodology presented in this work, we assume that the water utility has constructed a calibrated hydraulic model of the water distribution network based on some recent historical and field data. As this calibrated hydraulic model may change in time, the model should be updated in time to capture the new system dynamics.

A lookup algorithm can also be used in the computing of the possible contaminant injection nodes and time; however this method can only be applied in the case when it is known that the hydraulic dynamics in the distribution system, which correspond to a certain previous time duration, are similar to the current hydraulic dynamics. For instance, consider the case where we can describe the water demands at each node by using a known periodic signal, with a period of 24 hours. First, we construct the finite set $\mathcal{S} \subset \mathcal{S}^*$ of contamination fault functions which start within one day (e.g. midnight to midnight, every 5 minutes), and at every node possible; we then use a simulation software to compute, for each contamination fault function, the time when each contamination fault arrives at every node in the network. As a result, we can construct the contamination fault arrival matrix, of size $|\mathcal{S}| \times |\mathcal{V}|$, where $|\cdot|$ is the cardinality of a set, whose (i, j)-th element is the time within the day (between 0 and 24) when the

contaminant arrives at the j-th node, due to the i-th contamination fault function. Therefore, when a contamination fault is detected, a lookup algorithm is activated, which searches the contamination fault arrival matrix, in order to find those rows (and in extend those contamination fault functions in $\mathcal{S}$) which correspond to the nodes and times where detection has occurred, while taking into account some uncertainty, so as to construct $\mathcal{S}^0$. The drawback of this method is that the database has to be recalculated when the dynamics of the system change, and in addition, it may require large computational power in computing the dataset.

6.3.2 Contamination Impact and Propagation Path

In the following paragraphs, we introduce an algorithm which selects a node to conduct manual sampling $f_\mu : \mathcal{V}_m \times \{0,1\} \mapsto \mathcal{V}_m$, as used in (6.4). The algorithm initiates when a contamination fault alarm is activated and a quality fault is confirmed at the i-th node. Expanded sampling nodes are computed iteratively through the function $\mu^j = f_\mu(\mu^{j-1}, \nu^{j-1}; \mathcal{S}^{j-1})$, for $j \geq 1$ the manual sampling iteration and $\mu^j \in \mathcal{V}_m$ the sampled node index; $\nu^j \in \{0,1\}$ is the binary contamination flag assigned for the sampled node μ^{j-1} and $\mathcal{S}^j$ the set of feasible contamination fault function for the j-th manual sampling iteration, such that $\mathcal{S}^j \subseteq \mathcal{S}^0 \subset \mathcal{S}^*$. As initial conditions, we consider $\mu^0 = i$, where i is the confirmed as contaminated node index, for which $\nu^0 = 1$; $\mathcal{S}^0$ is computed using a suitable algorithm, as discussed in the previous section.

In the next paragraphs we introduce three parameters which are computed at the j-th sampling iteration and are used in the decision tree algorithm: the normalized impact metric vector W^j, the contamination impact label vector L^j and the contaminant propagation time matrix M^j.

The damage caused by a contamination fault within a certain time period can be measured using various impact metrics, such as the number of people affected or volume of contaminated water consumed. From a set of possible contamination fault functions we can compute the maximum damage caused by the worst-case contamination fault; we can use this to normalize the impact metrics. In specific, for $\mathcal{S}^j$ the set of possible contamination fault functions for the j-th sampling iteration, we define the normalized impact metric vector $W^j \in [0,1]^{|\mathcal{S}^j|}$ which is computed using a suitable function $W^j = f_w(\mathcal{S}^j)$, where $f_w(\cdot)$ is a function which calculates the impact increase between two time instances.

To assist the decision-making process, we may assign a label to each contamination fault function which characterizes the fault's impact to describe the severity of a certain contamina-

tion fault based on the normalized impact damage vector. Labels may be words with semantic meaning which indicates the severity of a contamination fault. Let Λ be the set of the risk-level labels considered; a typical example used in the literature is

$$\Lambda = \{\text{'Extreme'}, \text{'High'}, \text{'Moderate'}, \text{'Low'}, \text{'Minimal'}\}.$$

For each element in the normalized impact metric vector W^j, we compute the corresponding impact label; we define $f_\lambda : [0,1]^{|\mathcal{S}^j|} \mapsto \Lambda^{|\mathcal{S}^j|}$, as the function which maps the normalized impact damage vector W^j, for the j-th sampling iteration, to the impact label vector $L^j = f_\lambda(W^j)$. Function $f_\lambda(\cdot)$ and the impact label set Λ can be specified by the utility operators.

By computing the hydraulic and quality dynamics of the network model under the influence of each contamination fault function in $\mathcal{S}^j$, e.g. by using EPANET, we construct the contaminant propagation binary matrix M^j of size $|\mathcal{S}^j| \times |\mathcal{V}_m|$. In this matrix, $M^j_{(i,l)} = 1$ if the arrival time of the contaminant at the l-th node in $\mathcal{V}_m$, due to the i-th contamination fault in $\mathcal{S}^j$, is smaller than a certain time (e.g. the discrete time of most recent hydraulic data measurements); otherwise, $M_{(i,l)} = 0$.

6.3.3 Decision Tree Algorithm

A decision tree has a hierarchical structure, comprised of elements which correspond to some conditional statement, such as, "check the i-th node for contaminant traces"; depending on the existence, or not, of contaminant traces, the algorithm will continue to other conditional statements, or return an impact label.

Typically in decision trees, the conditional statement is computed at each iteration using a "splitting algorithm" to select the variable (in this case one network node) which will give the most information possible, with respect to the actual label [4]. Various metrics have been utilized in decision trees to measure information; for example, a metric commonly used in decision trees is the "information gain", which describes the change in the information entropy when examining a certain variable.

Below, we demonstrate how to compute the information gain for the impact label set L^j computed in the j-th sampling iteration. It is useful to first define the information entropy metric, which can be used to measure the homogeneity of a set; we let $f_e(\cdot)$ be the function which computes the information entropy of a set of impact labels. For Λ the set of risk-level labels considered and L^j a set of impact labels computed for each contamination fault function

in the set $\mathcal{S}^j$ for the j-th sampling iteration, the entropy of the set L^j is given by

$$f_e(L^j) = -\sum_{i=1}^{|\Lambda|} \frac{|c_i^j|}{|L^j|} \log_2 \frac{|c_i^j|}{|L^j|} \tag{6.6}$$

where $\frac{|c_i^j|}{|L^j|}$ is the *probability mass* or the *relative frequency* of the i-th label (where $|c_i^j|$ is the number of times the i-th label in Λ appears in the set L^j).

For example, for $\Lambda = \{\text{'High', 'Low'}\}$ and $L^0 = \{\text{'High', 'Low', 'High'}\}$, $|c_1^0| = 2$ and $|c_2^0| = 1$; the probability mass of the label 'High' is $\frac{|c_1^0|}{|L^0|} = \frac{2}{3}$ and of the label 'Low' is $\frac{|c_2^0|}{|L^0|} = \frac{1}{3}$; therefore, the information entropy of L^0 is $f_e(L^0) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183$.

Let $Z \in \mathbb{R}^{|\mathcal{V}_m|}$ be the vector of the information gain; for the i-th node in $\mathcal{V}_m$, the i-th element of the information gain vector Z_i^j for the j-th sampling iteration is computed by

$$Z_i^j = f_e(L^j) - \sum_{l\in\{0,1\}} \frac{|z_i^l|}{|L^j|} f_e(z_i^l), \tag{6.7}$$

where z_i^l is the set of impact labels corresponding to those contamination fault functions in the set $\mathcal{S}^j$, for which contaminant traces have arrived ($l = 1$) or not ($l = 0$) at the i-th node. To find this, we utilize the M^j binary matrix described in the previous section.

Although the information gain metric alone is useful in many classification problems, to address the security aspects of the problem, we propose to give greater weight on high-impact contamination faults, i.e. those which may cause significant damage on the consumers, rather than low-impact contamination faults. In this work we propose to select a node to conduct manual sampling such that if it is contaminated, lower-impact faults are to be excluded, rather than to exclude higher-impact faults if no contamination is detected at those locations.

For this reason, at the j-th manual sampling iteration we compute the maximum impact metric vector O^j of size $|\mathcal{V}_m|$. Its i-th element, O_i^j, is equal to the maximum value of all the impact metrics computed for those contamination fault functions for which the contaminant has arrived at the i node before a certain time. To compute this we find those rows in the matrix M^j for which the element in the i-th column is equal to one.

Thus, at the j-th manual sampling iteration, the splitting algorithm which is responsible for constructing the conditional statement of the decision tree, is based on the information gain vector Z^j and the maximum impact vector O^j.

Let $f_p : \mathbb{R}^{|\mathcal{V}_m|} \times \mathbb{R}^{|\mathcal{V}_m|} \mapsto \mathcal{V}_m$ be the algorithm which computes the node which will be used in the conditional statement of the decision tree at the j-th sampling iteration, such that $\mu^j = f_p(Z^j, O^j)$. At each sampling iteration, a node from the set $\mathcal{V}_m$ is selected by the

algorithm, which corresponds to the maximum value in both Z^j and O^j. It sometimes occurs that a single optimal solution for the two objective metrics cannot be found; therefore a multi-objective algorithm is required to compute the Pareto front, i.e. the subset of node attributes which are non-dominant to each other. From the set of Pareto solutions, a single solution is selected using some heuristic (e.g. the smallest Euclidean distance from elements with maximum values in Z^j and O^j).

After the conditional statement of the j-th manual sampling iteration has been constructed, the response team is instructed to conduct manual quality sampling at the location selected by the algorithm. If contaminant traces are detected at that location, then we let $v^j = 1$, whereas if no contaminant traces are detected, $v^j = 0$.

The decision tree algorithm terminates automatically if one of the following occurs: a) the calculated entropy is zero, $f_e(L^j) = 0$, i.e. all the elements in L^j have the same label value; b) the maximum element in the information gain is zero $\max_i Z_i^j = 0$, i.e. the solution cannot be improved any further by sampling at some node in the network.

When the decision tree is terminated, the algorithm returns an evaluation of the water contamination impact with the worst-case risk-label within the set L^j. In addition, the algorithm isolates the possible source areas, by constructing a set of unique node indices which corresponds to the injection locations of each fault contamination function set $\mathcal{S}^j$.

6.4 Simulations

In this section, we present two simulation examples to illustrate the formulation and solution methodology for the water contamination impact evaluation and the source-area isolation problem. The first example is a simple network, which is used to demonstrate how to construct a decision tree after a fault has been detected, while the second example is a real-world drinking water distribution network benchmark.

6.4.1 Illustrative Example

In this example, we demonstrate the solution methodology on a water distribution network with four constant demand nodes, as seen in Fig. 6.1. The directed graph is $(\mathcal{V}, \mathcal{A})$, where the node set is $\mathcal{V} = \{1, 2, 3, 4\}$ and the edge set is $\mathcal{A} = \{(1, 2), (1, 3), (3, 4), (2, 4)\}$. Both the consumer outflow demand $d(k)$ at each node and velocity $b(k)$ at each pipe are con-

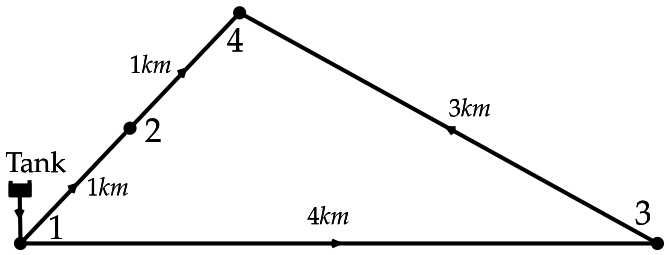


Figure 6.1: A trivial water distribution network with four demand nodes, supplied by a tank. A quality sensor is placed at node '4'.

stant at each discrete time, such that

$$d(k) = [5, 10, 20, 20]^\top \ \frac{m^3}{hr}$$

and

$$b(k) = [0.278, 0.278, 0.278, 0.278]^\top \ \frac{m}{s}$$

respectively; in addition, the pipe lengths are $[1, 4, 3, 1]^\top$ km. All other parameters in the network are considered known. One quality sensor is installed at $\mathcal{V}_s = \{4\}$, while $\mathcal{V}_m = \{1, 2, 3\}$ is the subset of nodes which can be used for manual sampling. We consider a time step of one hour, $\Delta t = 1$ hr; the length of one hydraulic period is 24 hours.

Consider the case where the contaminant concentration at the 4-th node, $x_4(k)$, is greater than a detection concentration threshold for $k \geq 10$; thus a contamination fault alert is triggered at $k = 10$ at the monitored node. This in turn activates the source-area isolation and impact evaluation algorithm.

In the following paragraphs we demonstrate how to select the first node to perform manual sampling. To compute the contamination fault functions set $\mathcal{S}^0 \subset \mathcal{S}$, we utilize a simple backtracking algorithm which follows the inverse in time contaminant propagation path. As an example, consider the pipe connecting nodes 3 and 4 of length 3 km; since the velocity at that pipe is $b_3(k) = 0.278 \ \frac{m}{s}$ (or $1 \ \frac{km}{hr}$), then the propagation time delay within that pipe is 3 hours. Therefore, one contamination fault corresponds to a contaminant injection at node 3, at discrete time $k = 7$. By repeating the reverse propagation, we compute the first set of contamination fault functions, as

$$\mathcal{S}^0 = \left\{\phi^1(k-8), \phi^2(k-9), \phi^3(k-7), \phi^4(k-10)\right\}.$$

For a certain contamination fault function $s \in \mathcal{S}^0$, we consider that the impact damage $\xi_i(k; s)$ at the node i is the volume of contaminated water consumed, and that the overall impact $\psi(k; s)$ is the volume of contaminated water consumed. To compute the normalized

impact damage W^1, we measure the impact increase from the start of the contamination fault, up until one hour after the detection time, at $k = 11$, the time of the most recent hydraulic sample available. Consider the first contamination fault function $\phi^1(k-8)$, which corresponds to a continuous contaminant injection at node 1 which initiates at discrete time $k = 8$; within 3 hours (until $k = 11$) the contaminant has propagated at nodes 2 and 4, but not yet at node 3, which is affected at a later time, since the propagation time delay between nodes 1 and 3 is 4 hours (thus reaching node 3 at $k = 12$).

For the first contamination fault function, where $\mathcal{S}^0 = \{s_1, ..., s_4\}$ and $s_1 = \phi^1(k-8)$, the normalized impact damage is calculated using function $f_w(\cdot)$ as follows: compute the overall impact, i.e. the volume of contaminated water consumed in the network up until time $k = 11$, such that $\psi(11; s_1) = d_1(k)\cdot(11-8)+d_2(k)\cdot(11-9)+d_4(k)\cdot(11-10) = 5\cdot3+10\cdot2+20\cdot1 = 55$; compute the maximum contaminated water volume from all contamination fault functions, in this case $W_3^1 = 100$; calculate the normalized impact damage for s_1, $W_1^1 = \frac{55}{100} = 0.55$. The normalized impact damage vector is $W^1 = [0.55, 0.40, 1.00, 0.20]^\top$.

We further define the set of three impact risk-level labels as $\Lambda = \{$'High','Moderate','Low'$\}$; for the i-th contamination fault function, we consider that if $0 \leq W_i^j < 0.33$, a 'Low' impact label is assigned; for $0.33 \leq W_i^j < 0.66$ and $0.66 \leq W_i^j \leq 1$, a 'Moderate' and a 'High' impact label are assigned respectively, for $j \geq 1$. In the first iteration, the algorithm $f_\lambda(\cdot)$ maps W^1 to the impact label set L^1, based on the above specification; therefore, the impact label set is $L^1 = \{$'Moderate', 'Moderate', 'High', 'Low'$\}$.

The fault propagation binary matrix M^1 of size 4×3 is computed. To illustrate on how to construct this matrix, consider the first contamination fault function $s_1 \in \mathcal{S}^0$, $s_1 = \phi^1(k-8)$; the contaminant propagates in the network and appears for the first time at each node in the manual sampling node set $\mathcal{V}_m$ as follows: node 1 at time $k = 8$, node 2 at time $k = 9$, node 3 at time $k = 12$. As in the previous paragraphs, we are interested in examining contamination faults at $k = 11$. The algorithm $f_m(\cdot)$ computes the elements of the matrix M^1 given the set of possible contamination fault functions; for the first contamination fault function s_1, $M^1_{(1,1)} = 1, M^1_{(1,2)} = 1$ and $M^1_{(1,3)} = 0$. Finally, the contamination fault propagation binary matrix is given by

$$M^1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{6.8}$$

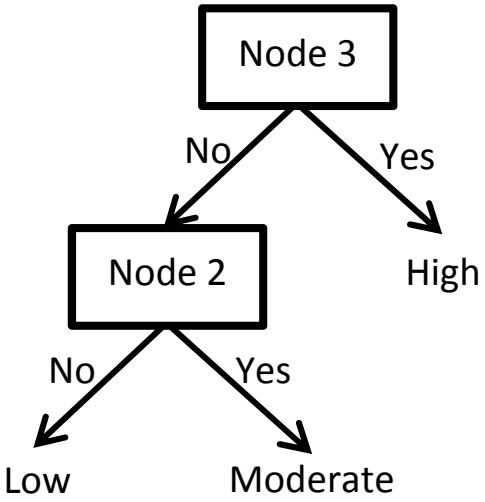


Figure 6.2: The decision tree for the impact evaluation of a contamination detected at node 4.

The entropy of L^1 is computed as $f_e(L^1) = 1.50$. The information gain for each node in $\mathcal{V}_m$ is $Z^1 = [0.311, 1.00, 0.811]^\top$ and the maximum worst-case impact is $O^1 = [0.55, 0.55, 1.00]^\top$. The i-th element in each vector Z^1 and O^1 corresponds to the i-th node where sampling should be conducted; in this case, nodes 2 and 3 have solutions on the Pareto front, $(1.00, 0.55)$ for node 2 and $(0.811, 1.00)$ for node 3. A decision can be made by selecting the Pareto solution with the smallest Euclidean distance from the maximum value of each vector, i.e. from $(1, 1)$; node 3 is therefore selected as the node to examine for contamination, i.e. $\mu^1 = 3$.

For illustrative purposes, the algorithm for computing the next sampling node is solved for all cases, i.e. in the existence or absence of a contamination fault. The complete decision tree is constructed for all possible cases.

In Fig. 6.2 the complete decision tree for this example is depicted. In summary, starting at time $k = 11$, an inspection team should check for contaminant traces at node 3; if there is indication of water contamination, then the fault impact is labeled as 'High' and the contamination source is at node 3. If there is no indication of water contamination, the fault's worst-case impact is labeled as 'Moderate', and possible contamination sources are nodes 1, 2 and 4. To further evaluate the water contamination impact and isolate the contamination source more efficiently, node 2 is examined next. If there is indication of water contamination at node 2, then the fault impact is labeled as 'Moderate' and possible contamination sources are nodes 1 or 2. If there is no indication of water contamination at 2, then the fault impact is labeled as 'Low' and the contamination source is at node 4.

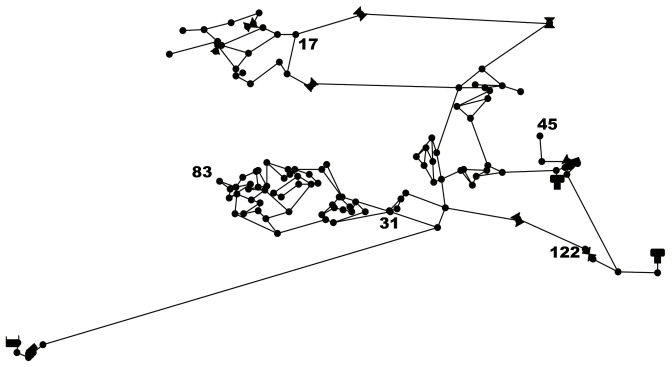


Figure 6.3: The real water distribution network benchmark. The indices correspond to the locations of online quality sensors.

6.4.2 Real Water Distribution Network Benchmark

In this example, we demonstrate the solution methodology on a real water distribution network benchmark. The task is to examine the performance of the proposed methodology in computing an expanded sampling scheme for the impact evaluation of a certain contamination fault which has appeared in the system.

Figure 6.3 depicts one of the benchmark networks in the "Battle of the Water Sensor Networks" design competition [123]. This network is composed of 178 pipes connected to 129 nodes (126 junctions, two tanks and one reservoir). All the network parameters are consider to be known. Each junction node is assigned with a daily average consumption volume as well as a discrete signal describing the rate of water consumption within 48 hours, with a 30-minute time step. These are assumed to describe the normal operation. The hydraulic dynamics are computed using the EPANET solver [142]; a daily period initiates at 8 A.M. and terminates in 24 hours. We assume that the contamination substance does not react with other substances flowing in the water distribution network. Five nodes in the network are monitored using fixed on-line quality sensors, and are indicated in Fig. 6.3 as $\mathcal{V}_s = \{$'17', '31', '45', '83', '122'$\}$; this is an optimal solution for the multiple objectives described in [90, 123]. We further consider that the water flows are approximately periodic and known, the time is discretized with $\Delta t = 5$ *min*, while the daily period is 288.

A lookup algorithm is utilized to construct the set $\mathcal{S}$, which is comprised of $28\,076$ contamination fault functions, which initiate at each node and at each discrete time in one day. A constant contaminant injection rate is considered for each fault function.

We simulate each contamination fault in $\mathcal{S}$ using a calibrated software model, and calculate the time the contaminant first arrives at a sensor node in $\mathcal{V}_s$, where it is detected. Figure

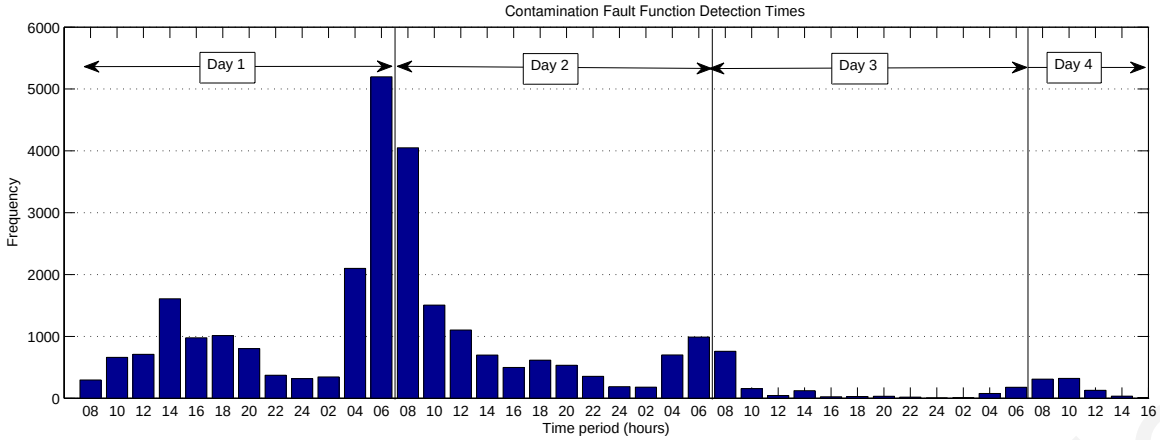


Figure 6.4: Detection times for contamination fault functions.

6.4 depicts the frequency of contamination fault detection, with respect to the time. For the system examined, most of the contamination fault functions are detected at sensor nodes between 6 A.M. and 10 A.M., whereas between midnight and 4 A.M. the detection frequency is at the lowest. A large number of faults are detected at the end of the first day and at the beginning of the second day. An important characteristic of the system is that some contamination faults may be detected a few days after their occurrence; in fact, some contamination faults are first detected at the sensor nodes in the fourth day of propagation within the distribution network.

Let $\Lambda = \{$'High','Moderate','Low'$\}$ be the set of three impact risk-level labels; for the i-th contamination fault function, we consider that if $0 \leq W_i^j < 0.33$, a 'Low' impact label is assigned; for $0.33 \leq W_i^j < 0.66$ and $0.66 \leq W_i^j \leq 1$, a 'Moderate' and a 'High' impact label are assigned respectively, for the j-th manual sampling iteration. As impact metric we consider the volume of contaminant mass consumed. From the set $\mathcal{S}$, 65% of the contamination faults have 'Low' impact, 21.4% have 'Moderate' impact, while 13.6% have 'High' impact.

To evaluate the effectiveness of the proposed solution methodology, we conduct an experiment in which we simulate 10 000 contamination fault functions, randomly selected from the contamination fault function set $\mathcal{S}$. Each contamination fault function is simulated, and the proposed decision tree algorithm is applied, in order to evaluate the impact damage and isolate the contamination source area.

Confusion matrix is a visualization tool which illustrates how well the actual risk-level labels of each contamination fault were classified; in this case, it summarizes the frequencies of misclassifications across each label. For example, in Table 6.1, when the actual contamination fault impact was 'High', all 1372 faults were correctly classified as 'High'; when the actual contamination fault impact was 'Moderate', 2117 faults were correctly classified as 'Moderate',

Table 6.1: Confusion Matrix for 10 000 Contamination Fault Functions

High	Moderate	Low	Classification
1372	36	154	as High
0	2117	275	as Moderate
0	0	6046	as Low

and 36 of them as 'High'. Finally, when the actual contamination fault impact was 'Low', 6046 faults were correctly classified, while 429 were classified with higher impacts. Overall, from all the simulations, the impact labels were correctly evaluated for 95.35% of all the contamination fault functions considered, 1.90% contamination faults were misclassified as 'High' impact faults, and 2.75% as 'Moderate' impact faults.

Even though the percentage of misclassification may appear significant, for some contamination fault it may not be possible to improve the classification of the risk-label by sampling in a different location in the network. In addition, it is important to note that the misclassifications never underestimate the contamination fault, and will always assign the highest impact-label, when there is no other information available, which is important with respect to the security perspective. In accordance to our solution methodology, it is preferable to misclassify a lower-impact fault as a higher-impact fault, instead of the opposite.

Detailed accuracy metrics for each label are presented in Table 6.2. We define the metrics used: a) The *True Positive Rate* for the i-th label is the number of contamination fault functions which have been correctly classified with the i-th label, over the number of all contamination fault functions which have been classified with the i-th label; b) The *False Positive Rate* for the i-th label, is the number of contamination faults misclassified with the i-th label, over the number of all contamination fault functions which have not been classified with the i-th label; c) The *Precision* metric for the i-th label is the number of contamination fault functions which have been correctly classified with the i-th label, over the number of all contamination fault functions which have been classified in that label, correctly or not; d) the *F-measure* metric for the i-th label is a statistical measure of the harmonic mean of *Precision* and the *True Positive Rate* for the i-th label, which describes the accuracy of the classifier, taking values from 0 to 1 (with 1 being the best value).

In the results presented in Table 6.2, True Positive Rate of the 'High' label has the maximum, equal to one. No contamination faults were misclassified with the 'Low' label, which has the minimum False Positive Rate, equal to zero; it also has the highest Precision among

Table 6.2: Detailed Accuracy by Impact Label

True Positive Rate	False Positive Rate	Precision	F-Measure	Label
1.000	0.022	0.878	0.935	High
0.983	0.035	0.885	0.932	Moderate
0.934	0.000	1.000	0.966	Low

all labels. For the network examined, most of the misclassifications were towards 'Moderate' and 'High' labels. For all labels, the F-measure shows that the classification is better for the 'Low' label rather than the 'Moderate' and 'High' labels; nevertheless, all have relatively high classification accuracies.

The statistical analysis has shown that the proposed methodology is suitable for the evaluation of the impact risk-level of contamination faults; however, we have not yet examined the difficulty in achieving so, with respect to the number of manual samplings required. Analysis of the results has shown that on average, 1.84 sampling iterations were necessary for the evaluation of the impact of a fault using decision trees. The cumulative distribution of the number of manual samplings with respect to the number of contamination fault functions evaluated is shown in Fig. 6.5; for 80% of the contamination fault functions considered, no more than two manual samplings were necessary in evaluating the impact, and for 90%, no more than four manual samplings. We should note that in practice manual sampling can stop at any time, provided that the accuracy of the impact evaluation and source-area isolation is considered adequate by the utility operator.

Source-area isolation is the secondary objective of the proposed methodology. When the algorithm terminates, it returns a set of contamination fault functions to the utility operator which confirm the observed quality measurements. From that set, by using a simple lookup algorithm we can construct a set of unique node locations, corresponding to the nodes where the contamination could have started. We expect that there may be a large number of nodes within a large-scale water distribution network, where the contamination fault could have occurred. For this reason, the nodes in this set may be segmented into groups, depending on the network connectivity and geographical location of each node; this information can be used by the utility operator to isolate the possible contamination source to one or more areas in the distribution network.

Figure 6.6 depicts the histograms of the number of possible source nodes for each contamination fault function, before and after the expanded sampling methodology is applied. For

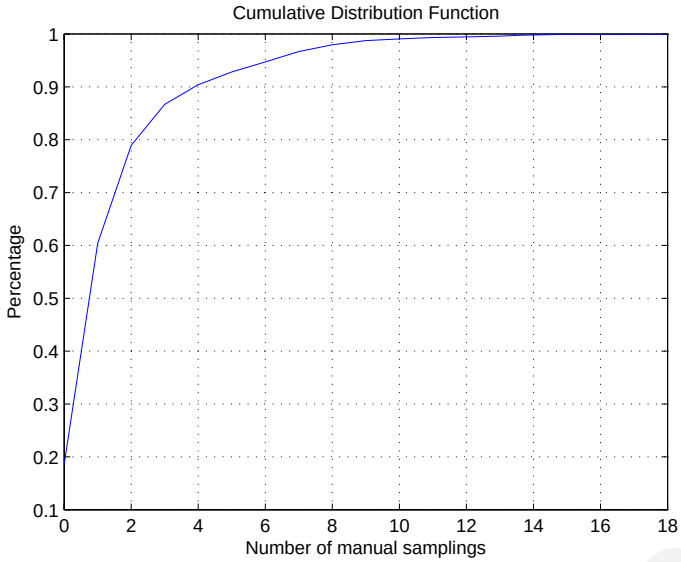


Figure 6.5: Cumulative function of the number of manual samplings required to evaluate the impact for each contamination fault function.

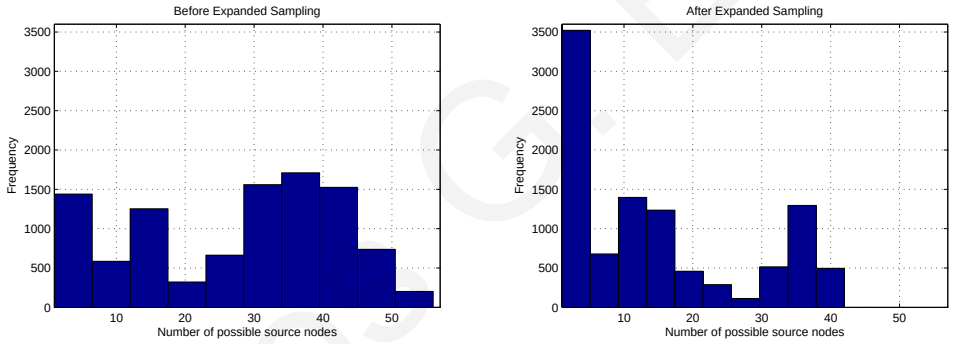


Figure 6.6: Histograms of the number of source nodes possible for all contamination fault functions, before and after expanded sampling.

this example, the average number of possible source nodes is reduced almost to half, from 29 source nodes to 15, before and after the expanded sampling respectively. In specific, we observe that in the first histogram, more than 50% of the contamination fault functions have more than 30 source node candidates (out of 129). By applying the proposed methodology, the histogram distribution is skewed towards zero, indicating that the number of possible source for a large number of contamination fault functions is reduced dramatically.

6.5 Concluding Remarks

In this chapter we have proposed a computational approach for choosing a sequence of nodes in the distribution network to perform expanded sampling, so that the water contam-

ination impact is evaluated and the source-area is isolated, with as few quality samples taken as possible.

Significant attention has been given to water security in the last few years. To provide a framework for water security, the US Environmental Protection Agency (EPA) has published guidelines for water utilities describing a *contamination warning system* architecture comprised of the standard operational procedures and the *consequence management plan*. As part of the consequence management plan, "Expanded Sampling" is recommended during the "Confirmed" determination stage, by conducting manual quality sampling at various locations in the water distribution network, aiming at evaluating the spread of the contamination, using a hydraulic model of the system. Selecting where to conduct manual sampling within a water distribution system is a challenging task, while the use of computational methodologies may reduce the reaction time and take more informed decisions.

The proposed method is based on constructing a decision tree using multiple objectives. To illustrate the solution methodology, we presented results based on a simplified and a real water distribution system benchmark. Future work will examine extending the problem formulation for multiple response teams, and address the computational complexity difficulties which may arise. In addition, future work will investigate how the demand and model uncertainty affect the impact evaluation and source-area isolation problem.

The contributions of this work is the mathematical formulation, the design and the implementation of an expanded manual sampling scheduling algorithm, which can serve as a tool for the decision maker within the scope of a contamination warning system; as added benefit, the proposed algorithm is in accordance to the guidelines specified by the EPA as part of the *consequence management plan* when a contamination is detected in a water distribution network. A journal paper on the results presented in this chapter is in a peer-review stage [57], and part of the results have been accepted to appear in a peer-reviewed conference [65].

Chapter 7

Conclusions

Fault tolerance and security issues in drinking water distribution networks are increasingly receiving more attention in the context of monitoring and control of critical infrastructure systems. In specific, the risk of hydraulic and quality faults in water distribution systems which may cause water losses or water contamination, with the danger of deteriorating the consumers' health, constituted the motivation of this research.

From a system theoretic point of view, monitoring and control of water distribution networks present important new challenges due to their large-scale interconnected dynamics, their structural uncertainties, the complex propagation and reaction dynamics, as well as the hydraulics uncertainties. Analytical results and algorithmic tools developed in the fault diagnosis research community could be employed, extending the mathematical formulation proposed, and especially on the issues of fault detection, isolation, identification and accommodation.

In this thesis we have addressed certain key issues on fault diagnosis and security in drinking water distribution networks, related to water systems modeling, quality faults monitoring, quality regulation, hydraulic faults detection and contamination faults isolation. In specific, we presented a formulation of a system-theoretic framework suitable for fault diagnosis and security monitoring in water distribution systems.

7.1 Relevant Approaches and Contribution

Regarding to water systems modeling, previous research has demonstrated how to construct mathematical models of hydraulic and quality dynamics, in pressurized water networks, as well as models for measuring the impact due to a contamination. In most of the previous research, the underlying dynamics are assumed to be described with software model (e.g.

through EPANET), and as a result, there was not an analytical mathematical framework which coupled the advection, reaction and impact dynamics. In the present work, a formulation of the monitoring and control problem of water distribution networks was presented, in a framework suitable for sensor placement and fault diagnosis.

Regarding to the quality fault monitoring problem, a significant part of previous research related with solution methodologies and algorithms to compute the sensor placement locations in a water distribution network. For example, the Battle of the Water Sensor Networks [123], which instigated research in quality monitoring through sensor placement, was an algorithmic competition where researchers developed algorithms for computing Pareto optimal solutions. In general, a few analytical problem formulations have been presented in the literature, which are related with a specific solution methodology. With respect to currently available software, it is important to mention the *Threat Ensemble Vulnerability Assessment and Sensor Placement Optimization Tool* (TEVA-SPOT), which has been developed by the U. S. EPA, Sandia National Laboratories, Argonne National Laboratory, and the University of Cincinnati [13]. TEVA-SPOT uses as a mathematical model the EPANET software, and its hydraulic and quality solver are used to simulate various contamination faults. The set of contamination fault scenarios is defined by the user, and afterwards, the health impact is computed based on a set of metrics and the contaminant type. The sensor-placement problem is formulated as a mathematical program of certain objective metrics (e.g. cost, population killed) with respect to certain objective functions (e.g. mean, worst), certain response times and with respect to certain constraints. Based on the developed framework, the sensor placement problem was examined, to find suitable locations in a water distribution network where on-line quality sensors ought to be installed, in order to minimize the risk of a severe damage on the population, in the case that a contaminant enters the water distribution network. The methodology for sensor placement proposed in this thesis, may be considered as a generalized and analytical formulation of the TEVA-SPOT methodology. Furthermore, in the present work, the manual quality sampling scheduling problem was examined, to find where and when to take water samples for quality monitoring.

Regarding to quality regulation, in some of the previous research the problem was address in a steady-state formulation, without adapting to changing dynamics. In some research, adaptive approximation methodologies have been proposed, which assumed that the hydraulic dynamics and controls exhibit periodicity, which may not hold in practice. In the present work, a disinfectant concentration regulation algorithm for water distribution networks was designed, using adaptive approximation to learn water demands.

Regarding to hydraulic fault detection, in previous research the problem was addressed within an optimization framework, as a inverse problem; however, information on the network parameters is required, which may not be available. In addition, transient pressures have been used for leakage detection; however, the use of high-frequency hydraulic pressure sensors may not be available in some water distribution networks. In addition, in previous work leakages were detected assuming fixed thresholds, which may not be realistic. In the present work, the detection of hydraulic leakage faults in District Metered Areas was examined, by using a fault detection method based on learning the periodic consumption dynamics.

Regarding to contamination fault isolation, in most of the previous research it was assumed that quality sensor measurements are considered for finding the source of a contamination; however in practice, quality sensors are installed at locations in the network for maximum coverage and for a small number of sensors, the redundancy in the measurements may be small, or there may be large time-delays between the measurements. Recently, a manual (grab) sampling approach was proposed, however the contamination impact has not been directly considered. In the present work, the impact of a contamination detected in a water distribution system was evaluated, and its source area was isolated, using a methodology based on decision tree induction.

The effectiveness of all the methodologies proposed in the present work, was illustrated with simulations using water distribution system models and historical hydraulic data.

7.2 Methodologies Application Guidelines

The results presented in this thesis could be applied to a real drinking water distribution network of a water utility. For this, the utility should first construct a model of the network, and calibrate its parameters using field tests and other techniques, in order to capture the hydraulic and quality dynamics are good as possible. What follows, is to construct estimates of the water consumption at each node; the use of specialized flow sensors at the residence level, and use real-time consumption monitoring is the optimal solution. When this information is not available, demand estimation from representative consumers, as well as DMA monitoring can be used. The utility can then compute the locations where quality sensors should be installed, as well as where to install chlorine booster stations controlled with a regulation algorithm, and in addition, compute where and when to conduct manual quality sampling. With respect to minimizing the risk due to a contamination fault, in collaboration with the authorities and policymakers, certain objectives should be specified, and based on those, compute the

optimal locations to install quality sensors. The calibrated model and the demand estimations can be used to simulate the various contamination fault scenarios, and solve the sensor placement and manual sampling scheduling problem, for a certain network topology and a certain water demand pattern. Due to changes in the consumption dynamics, these locations of these contamination/quality sensors may change, by recalculating new positions which minimize the objectives set by the decision makers. For leakage detection at a DMA, embedded software may be developed to be installed at the DMA entrance, or process the hydraulic data when they are received by the utility. Regarding the contamination isolation problem, specialized decision support software should be developed, which utilizes the calibrated model and water demand estimations, to compute the sampling locations when a contamination has been detected. In the future, the algorithms proposed in this work will be implemented in embedded or software systems.

7.3 Future Work

In this thesis, it was assumed that sensors are capable of measuring a known contaminant substance; water utilities, however, can sometimes afford to install only a small number of sensors measuring standard quality metrics, such as disinfectant concentrations (chlorine), pH and turbidity. The formulation proposed in this thesis could be extended by applying fault detection algorithms which use disinfectant concentration measurements, in order to deduce contaminant presence. In addition, in this thesis, no prior assumptions were made in relation to the probability of a contamination fault occurring. In fact, the contamination fault scenario construction makes no assumptions on the likelihood of a contamination occurring at some nodes; this is true in the case when considering malicious contamination attacks, which may occur anywhere and at anytime in the network. In future work we will examine the use of probabilistic models in the fault-scenario construction. With respect to the leakage detection problem, the work presented in this thesis can be extended to take into consideration pressure measurements for estimating the magnitude of the leakage fault, as well as to implement leakage detection algorithms in a plug-and-play embedded system framework, to be used by water utilities. Furthermore, we will investigate the effect of using "skeletonised" or "aggregated" water distribution models when solving the contamination source-area isolation and impact evaluation problem, as well as to address the problem of constructing representative contamination fault scenarios.

An interesting question is whether there can be a unified mathematical model to address

the different fault diagnosis and security monitoring problems discussed. We believe that it is possible to view the various problems, as a general problem seen from various viewpoints. For example, the quality monitoring and quality control can be considered as a single optimization problem; in addition, the sensor placement problem, the manual sampling problem and the contaminant concentration problems can also be considered within a unified framework as a general multi-objective optimal selection problem with respect to certain risk metrics. Future research will address the problem of a unified approach for fault diagnosis and water security, within a common mathematical framework and problem formulation.

In the end, the problems addressed within this thesis, require, in general, the synergy of various research fields, such as control and systems engineering, water engineering, hydro-informatics, optimization and risk management. The multi-disciplinary nature of these problems is by-itself a challenge, however, it is also a fertile ground for innovation, in an infrastructure delivering one of society's most important human right, water.

Bibliography

[1] S. R. Agha, "Use of goal programming and integer programming for water quality management—a case study of Gaza Strip," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1991–1998, Nov. 2006.

[2] M. A. Al-Zahrani and K. Moeid, "Locating optimum water quality monitoring stations in water distribution system," in *Proc. ASCE World Water and Environmental Resources*, 2001, pp. 393–402.

[3] H. Alegre, J. Baptista, E. C. Jr, F. Cubillo, P. Duarte, W. Hirner, W. Merkel, and R. Parena, *Performance Indicators for Water Supply Services*. London, UK: International Water Association, 2006.

[4] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA, USA: MIT Press, 2004.

[5] A. An, C. Chan, N. Shan, N. Cercone, and W. Ziarko, "Applying knowledge discovery to predict water-supply consumption," *IEEE Expert*, vol. 12, no. 4, pp. 72–78, Jul./Aug. 1997.

[6] M. M. Aral, J. Guan, and M. L. Maslia, "Optimal design of sensor placement in water distribution networks," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 1, pp. 5–18, Jan./Feb. 2010.

[7] K. Astrom and B. Wittenmark, *Adaptive Control*. Reading, MA, USA: Addison-Wesley, 1995.

[8] M. Basseville, A. Benveniste, G. Moustakides, and A. Rougee, "Optimal sensor location for detecting changes in dynamical behavior," *IEEE Transactions on Automatic Control*, vol. 32, no. 12, pp. 1067 – 1075, Dec. 1987.

[9] M. Basseville and I. Nikiforov, *Detection of abrupt changes: theory and application*. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

[10] M. Basseville, "Detecting changes in signals and systems — a survey," *Automatica*, vol. 24, no. 3, pp. 309–326, May 1988.

[11] J. Berry, W. Hart, C. Phillips, J. Uber, and J. Watson, "Sensor placement in municipal water networks with temporal integer programming models," *ASCE Journal of Water Resources Planning and Management*, vol. 132, no. 4, pp. 218–224, Jul./Aug. 2006.

[12] J. Berry, H. Lin, E. Lauer, and C. Phillips, "Scheduling manual sampling for contamination detection in municipal water networks," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, paper 95, pp. 95.1–95.10.

[13] J. Berry, E. Boman, L. A. Riesen, W. E. Hart, C. A. Phillips, and J.-P. Watson, *User's Manual TEVA-SPOT Toolkit*, 2nd ed., U.S. Environmental Protection Agency, Cincinnati, OH 4525, Oct. 2010, EPA 600/R-08/041B.

[14] J. Berry, R. D. Carr, W. E. Hart, V. J. Leung, C. A. Phillips, and J.-P. Watson, "Designing contamination warning systems for municipal water networks using imperfect sensors," *ACSE Journal of Water Resources Planning and Management*, vol. 135, no. 4, pp. 253–263, Jul./Aug. 2009.

[15] J. Berry, L. Fleischer, W. Hart, and C. Phillips, "Sensor placement in municipal water networks," in *Proc. ASCE World Water and Environmental Resources*, 2003, pp. 40–49.

[16] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J.-P. Watson, "Sensor placement in municipal water networks," *ASCE Journal of Water Resources Planning and Management*, vol. 131, no. 3, pp. 237–243, May/Jun. 2005.

[17] P. Biswas, C. Lu, and R. Clark, "A model for chlorine concentration decay in pipes," *Water Research*, vol. 27, no. 12, pp. 1715–1724, Dec. 1993.

[18] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder, *Diagnosis and Fault-Tolerant Control*. Berlin, Germany ; New York, USA: Springer, 2003.

[19] D. L. Boccelli and W. E. Hart, "Optimal monitoring location selection for water quality issues," in *Proc. ASCE World Environmental and Water Resources*, 2007, pp. 522–527.

[20] D. L. Boccelli, M. E. Tryby, J. G. Uber, L. A. Rossman, M. L. Zierolf, and M. M. Polycarpou, "Optimal scheduling of booster disinfection in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 124, no. 2, pp. 99–111, Mar./Apr. 1998.

[21] D. L. Boccelli, M. E. Tryby, J. G. Uber, and R. S. Summers, "A reactive species model for chlorine decay and THM formation under rechlorination conditions," *Water Research*, vol. 37, no. 11, pp. 2654–2666, Jun. 2003.

[22] J. Bougadis, K. Adamowski, and R. Diduch, "Short-term municipal water demand forecasting," *Hydrological Processes*, vol. 19, no. 1, pp. 137–148, Jan. 2005.

[23] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 1994.

[24] M. A. Brdys, T. Chang, and K. Duzinkiewicz, "Intelligent model predictive control of chlorine residuals in water distribution systems," in *Proc. ASCE Water Resource Engineering and Water Resources Planning and Management*, 2001, pp. 391–401.

[25] M. A. Brdys and B. Ulanicki, *Operational control of water systems: structures, algorithms, and applications*. New York, USA: Prentice Hall, 1994.

[26] M. Brdys and T. Chang, "Robust model predictive control under output constraints," in *Proc. IFAC World Congress*, 2002, pp. 21–26.

[27] M. A. Brdys, T. Chang, K. Duzinkiewicz, and W. Chotkowski, "Hierarchical control of integrated quality and quantity in water distribution systems," in *Proc. ASCE Water Resource Engineering and Water Resources Planning and Management*, 2000, pp. 193–202.

[28] S. G. Buchberger and G. Nadimpalli, "Leak estimation in water distribution systems by statistical analysis of flow readings," *ASCE Journal of Water Resources Planning and Management*, vol. 130, no. 4, pp. 321–329, Jul./Aug. 2004.

[29] S. G. Buchberger and L. Wu, "Model for instantaneous residential water demands," *ASCE Journal of Hydraulic Engineering*, vol. 121, no. 3, pp. 232–246, Mar. 1995.

[30] F. Camacho and C. Bordons, *Model Predictive Control*, 2nd ed. London, UK ; New York, USA: Springer, 2004.

[31] A. Caputo and P. Pelagagge, "An inverse approach for piping networks monitoring," *Journal of Loss Prevention in the Process Industries*, vol. 15, no. 6, pp. 497–505, Nov. 2002.

[32] W. Carus, *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents since 1900*. Amsterdam, The Netherlands: Fredonia Books, 2002.

[33] S. Chakrabarti, D. Eliades, E. Kyriakides, and M. Albu, "Measurement uncertainty considerations in optimal sensor deployment for state estimation," in *Proc. IEEE International Symposium on Intelligent Signal Processing*, 2007, p. 6.

[34] S. Chakrabarti, E. Kyriakides, and D. Eliades, "Placement of synchronized measurements for power system observability," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 12–19, Jan. 2009.

[35] T. Chang, M. A. Brdys, and K. Duzinkiewicz, "Decentralized robust model predictive control of chlorine residuals in drinking water distribution systems," in *Proc. ASCE World Water Congress*, 2003, pp. 255–264.

[36] ——, "Quantifying uncertainties for chlorine residual control in drinking water distribution systems," in *Proc. ASCE World Water and Environmental Resources*, 2003, pp. 28–37.

[37] T. Chang, "Robust model predictive control of water quality in drinking water distribution systems," Ph.D. dissertation, The University of Birmingham, Birmingham, UK, 2003.

[38] W. Cheng and Z. He, "Calibration of nodal demand in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 137, no. 1, pp. 31–40, Jan./Feb. 2011.

[39] S. Chick, S. Soorapanth, and J. Koopman, "Inferring infection transmission parameters that influence water treatment decisions," *INFORMS Management Science*, vol. 49, no. 7, pp. 920–935, Jul. 2003.

[40] R. Church and C. ReVelle, "The maximal covering location problem," *Papers in Regional Science*, vol. 32, no. 1, pp. 101–118, Dec. 1974.

[41] R. M. Clark, L. A. Rossman, M. Sivaganesan, and K. M. Schenck, *Controlling Disinfection By-Products and Microbial Contaminants in Drinking Water*. Cincinnati, OH, USA: U.S. Environmental Protection Agency, 2001, ch. Modeling Chlorine Decay and the Formation of Disinfection By-Products (DBPs) in Drinking Water, pp. 265–298.

[42] A. F. Colombo, P. Lee, and B. W. Karney, "A selective literature review of transient-based leak detection methods," *Journal of Hydro-environment Research*, vol. 2, no. 4, pp. 212 – 227, Apr. 2009.

[43] S. Constans, B. Bremond, and P. Morel, "Simulation and control of chlorine levels in water distribution networks," *ASCE Journal of Water Resources Planning and Management*, vol. 129, no. 2, pp. 135–145, Mar./Apr. 2003.

[44] D. Covas and H. Ramos, "Case studies of leak detection and location in water pipe systems by inverse transient analysis," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 2, pp. 248–257, Mar./Apr. 2010.

[45] P. S. P. Cowpertwait and A. V. Metcalfe, *Introductory Time Series with R*. New York, USA: Springer, 2009.

[46] C. D. Cristo and A. Leopardi, "Pollution source identification of accidental contamination in water distribution networks," *ASCE Journal of Water Resources Planning and Management*, vol. 134, no. 2, pp. 197–202, Mar./Apr. 2008.

[47] G. Dandy and C. Hewitson, "Optimizing hydraulics and water quality in water distribution networks using genetic algorithms," in *Proc. ASCE Water Resource Engineering and Water Resources Planning and Management*, 2000, pp. 211–220.

[48] A. E. De Sanctis, F. Shang, and J. G. Uber, "Determining possible contaminant sources through flow path analysis," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, pp. 124–137.

[49] ——, "Real-time identification of possible contamination sources using network backtracking methods," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 4, pp. 444–453, Jul./Aug. 2010.

[50] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[51] M. Dodd, N. Vu, A. Ammann, R. Kissner, H. Pham, M. Berg, and U. von Gunten, "Kinetics and mechanistic aspects of As (III) oxidation by aqueous chlorine, chloramines, and ozone: Relevance to drinking water treatment," *Environmental Science & Technology*, vol. 40, no. 10, pp. 3285–3292, May 2006.

[52] G. Dorini, P. Jonkergouw, Z. Kapelan, and D. Savic, "SLOTS: Effective algorithm for sensor placement in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 6, pp. 620–628, Nov./Dec. 2010.

[53] K. Duzinkiewicz, M. Brdys, and T. Chang, "Hierarchical model predictive control of integrated quality and quantity in drinking water distribution systems," *Urban Water Journal*, vol. 2, no. 2, pp. 125–137, Jun. 2005.

[54] D. G. Eliades, E. C. Kyriakides, and M. M. Polycarpou, "Enhanced robustness in model predictive control of water quality using adaptive estimation methods," in *Proc. Computing and Control for the Water Industry*, Leicester, UK, 2007, p. 8.

[55] D. G. Eliades and M. M. Polycarpou, "Multi-objective optimization of water quality sensor placement in drinking water distribution networks," in *Proc. European Control Conference*, Kos, Greece, 2007, pp. 1626–1633.

[56] ——, "Leakage fault detection in a district metered area of a water distribution system," 2011, submitted for publication.

[57] ——, "Water contamination impact evaluation and source-area isolation using decision trees," 2011, submitted for publication.

[58] D. G. Eliades, M. M. Polycarpou, and B. Charalambous, "Security issues in drinking water distribution networks," *Journal of Information Assurance and Security*, vol. 4, no. 6, pp. 500–508, 2009.

[59] D. G. Eliades, M. M. Polycarpou, and B. Charalampous, "A security-oriented manual quality sampling methodology," in *Proc. European Water Resources Association*, Limassol, Cyprus, 2009, p. 8.

[60] ——, "A security-oriented manual quality sampling methodology for water systems," in *Proc. Electrical and Computer Engineering Student Conference*, Thessaloniki, Greece, 2009, p. 6.

[61] ——, "A security-oriented manual quality sampling methodology for water systems," *Water Resources Management*, vol. 25, no. 4, pp. 1219–1228, Mar. 2010.

[62] D. Eliades and M. Polycarpou, "Iterative deepening of pareto solutions in water sensor networks," in *Proc. ASCE Water Distribution Systems Analysis*, Cincinnati, USA, 2006, paper 114, p. 14.

[63] D. G. Eliades and M. M. Polycarpou, "Security issues in drinking water distribution networks," in *Proc. Computational Intelligence in Security for Information Systems*, 2009, pp. 69–76.

[64] ——, "Security of water infrastructure systems," in *Critical Information Infrastructure Security*, ser. Lecture Notes in Computer Science, R. Setola and S. Geretshuber, Eds. Berlin / Heidelberg, Germany: Springer, 2009, vol. 5508, pp. 360–367.

[65] ——, "Fault isolation and impact evaluation of water distribution network contamination," in *Proc. IFAC World Congress*, 2011, p. 6, to appear.

[66] D. Eliades and M. Polycarpou, "A fault diagnosis and security framework for water systems," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 6, pp. 1254 –1265, Nov. 2010.

[67] European Commission, "Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption," Official Journal of the European Communities, pp. 32–54, Dec. 1998.

[68] ——, "Critical infrastructure protection in the fight against terrorism," Oct. 2004, COM(2004) 702 final.

[69] M. Farley, *Leakage management and control: A best practice training manual*. Geneva, Switzerland: World Health Organization, 2001.

[70] J. A. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control-Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Hoboken, NJ, USA: Wiley-Interscience, 2006.

[71] A. Fügenschuh, S. Göttlich, and M. Herty, "Water contamination detection," in *Wirtschaftinformatik Proceedings*, 2007, p. 18.

[72] B. Gabrys and A. Bargiela, "General fuzzy min-max neural network for clustering and classification," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 769 –783, May 2000.

[73] ——, "Neural networks based decision support in presence of uncertainties," *ASCE Journal of Water Resources Planning and Management*, vol. 125, no. 5, pp. 272–280, Sep./Oct. 1999.

[74] G. Germanopoulos, "A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models," *Civil Engineering Systems*, vol. 2, no. 3, pp. 171–179, Sep. 1985.

[75] O. Giustolisi, D. Savic, and Z. Kapelan, "Pressure-driven demand and leakage simulation for water distribution networks," *ASCE Journal of Hydraulic Engineering*, vol. 134, no. 5, pp. 626–635, May 2008.

[76] P. H. Gleick, "Water and terrorism," *Water Policy*, vol. 8, no. 6, pp. 481–503, May 2006.

[77] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley, 1989.

[78] W. M. Grayman, "A quarter of a century of water quality modeling in distribution systems," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, p. 12.

[79] J. Guan, M. M. Aral, M. L. Maslia, and W. M. Grayman, "Identification of contaminant sources in water distribution systems using simulation–optimization method: Case study," *ASCE Journal of Water Resources Planning and Management*, vol. 132, no. 4, pp. 252–262, Jul./Aug. 2006.

[80] M. Haestad, T. Walski, D. Chase, D. Savic, W. Grayman, S. Backwith, and E. Koelle, *Advanced Water Distribution Modeling and Management*. Waterbury, CT, USA: Haestad Press, 2003.

[81] S. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Operations Research*, vol. 12, no. 3, pp. 450–459, May/Jun. 1964.

[82] P. Harmant, A. Nace, L. Kiene, and F. Fotoohi, "Optimal supervision of drinking water distribution network," in *Proc. ASCE Water Resources Planning and Management*, 1999, pp. 52–60.

[83] W. E. Hart and R. Murray, "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 6, pp. 611–619, Nov./Dec. 2010.

[84] J. Hill, B. van Bloemen Waanders, and C. Laird, "Source inversion with uncertain sensor measurements," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, p. 13.

[85] F. Hua, J. West, R. Barker, and C. Forster, "Modelling of chlorine decay in municipal water supplies," *Water Research*, vol. 33, no. 12, pp. 2735–2746, Aug. 1999.

[86] J. J. Huang and E. A. McBean, "Data mining to identify contaminant event locations in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 135, no. 6, pp. 466–474, Nov./Dec. 2009.

[87] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods—an introduction," *Control Engineering Practice*, vol. 5, no. 5, pp. 639–652, May 1997.

[88] B. W. Karney, *Water distribution systems handbook*. New York, USA: McGraw-Hill, 2000, ch. Hydraulics of Pressurized Flow, pp. 2.1–2.43.

[89] J. H. Kim, S. H. Hwang, and H. S. Shin, "An optimal neural network model for daily water demand forecasting," in *Proc. ASCE Water Resources Planning and Management*, 1999, pp. 236–245.

[90] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *ASCE Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 516–526, Nov./Dec. 2008.

[91] A. Kumar, M. L. Kansal, and G. Arora, "Identification of monitoring stations in water distribution system," *ASCE Journal of Environmental Engineering*, vol. 123, no. 8, pp. 746–752, Aug. 1997.

[92] J. Kumar, E. M. Zechman, E. D. Brill, G. Mahinthakumar, S. Ranjithan, and J. Uber, "Evaluation of non-uniqueness in contaminant source characterization based on sensors with event detection methods," in *Proc. ASCE World Environmental and Water Resources*, 2007, pp. 513–520.

[93] E. Kyriakides and M. Polycarpou, *Studies in Computational Intelligence*. Berlin / Heidelberg, Germany: Springer, 2007, vol. 35, ch. Short Term Electric Load Forecasting: A Tutorial, pp. 391–418.

[94] C. D. Laird, L. T. Biegler, and B. G. van Bloemen Waanders, "Mixed-integer approach for obtaining unique solutions in source inversion of water networks," *ASCE Journal of Water Resources Planning and Management*, vol. 132, no. 4, pp. 242–251, Jul./Aug. 2006.

[95] C. D. Laird, L. T. Biegler, B. G. van Bloemen Waanders, and R. A. Bartlett, "Time dependent contamination source determination: A network subdomain approach for very large water networks," in *Proc. ASCE World Water and Environmental Resources*, 2004, pp. 469–480.

[96] C. Laird, L. Biegler, B. van Bloemen Waanders, and R. Bartlett, "Contamination source determination for water networks," *ASCE Journal of Water Resources Planning and Management*, vol. 131, no. 2, pp. 125–134, Mar./Apr. 2005.

[97] A. Lambert and J. A. E. Morrison, "Recent developments in application of 'bursts and background estimates' concepts for leakage management," *Water and Environment Journal*, vol. 10, no. 2, pp. 100–104, Apr. 1996.

[98] A. Lambert, "What do we know about pressure: Leakage relationships in distribution systems?" in *Proc. IWA System Approach to Leakage Control and Water Distribution Systems Management*, 2001, p. 8.

[99] K. Lansey and L. W. Mays, *Water distribution systems handbook*. New York, USA: McGraw-Hill, 2000, ch. Hydraulics of Water Distribution Systems, pp. 4.1–4.29.

[100] B. Lee and R. Deininger, "Optimal locations of monitoring stations in water distribution system," *ASCE Journal of Environmental Engineering*, vol. 118, no. 1, pp. 4–16, Jan./Feb. 1992.

[101] R. LeVeque, "Nonlinear conservation laws and finite volume methods," in *Computational methods for astrophysical fluid flow*, O. Steiner and A. Gautschy, Eds. Berlin, Germany ; New York, USA: Springer, 1998, pp. 1–160.

[102] J. A. Liggett and L.-C. Chen, "Inverse transient analysis in pipe networks," *ASCE Journal of Hydraulic Engineering*, vol. 120, no. 8, pp. 934–955, Aug. 1994.

[103] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown," *Journal of the American Statistical Association*, vol. 64, no. 325, pp. 387–389, Mar. 1969.

[104] L. Liu, "Real-time contaminant source characterization in water distribution systems," Ph.D. dissertation, North Carolina State University, Raleigh, North Carolina, 2009.

[105] L. Liu, E. M. Zechman, J. E. Downey Brill, G. Mahinthakumar, S. Ranjithan, and J. Uber, "Adaptive contamination source identification in water distribution systems using an evolutionary algorithm-based dynamic optimization procedure," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, pp. 123–133.

[106] J. Mashford, D. D. Silva, and D. M. S. Burn, "An approach to leak detection in pipe networks using analysis of monitored pressure values by Support Vector Machine," in *Proc. Network and System Security*, 2009, pp. 534–539.

[107] M. L. Maslia, J. B. Sautner, M. M. Aral, R. Gillig, J. J. Reyes, , and R. C. Williams, *Historical Reconstruction of the Water-distribution System Serving the Dover Township Area, New Jersey, January 1962-December 1996*. Georgia, USA: ATSDR — US Dept. of Health and Human Services, Oct. 2001.

[108] M. L. Maslia, J. B. Sautner, M. M. Aral, J. J. Reyes, J. E. Abraham, and R. C. Williams, "Using water-distribution system modeling to assist epidemiologic investigations," *ASCE Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 180–198, Jul./Aug. 2000.

[109] J. McClellan, D. Reckhow, J. Tobiason, J. Edzwald, and D. Smith, *Natural Organic Matter and Disinfection By-Products: Chacterization and Control in Drinking Water*. Washington DC, USA: American Chemical Society, 2000, ch. A Comprehensive Kinetic Model for Chlorine Decay and Chlorination Byproduct Formation, pp. 223–246.

[110] S. McKenna, B. Waanders, C. Laird, S. Buchberger, Z. Li, and R. Janke, "Source location inversion and the effect of stochastically varying demand," in *Proc. ASCE World Water and Environmental Resources*, 2005, p. 10.

[111] P. Menold, R. Pearson, and F. Allgöwer, "Online outlier detection and removal," in *Proc. Mediterranean Conference on Control and Automation*, Haifa, Israel, 1999, pp. 1110–1133.

[112] D. Misiunas, "Failure monitoring and asset condition assessment in water supply systems," PhD thesis, Lund University, Lund, Sweden, 2005.

[113] D. Misiunas, M. Lambert, A. Simpson, and G. Olsson, "Burst detection and location in water distribution networks," *Water Supply*, vol. 5, no. 3-4, pp. 71–80, 2005.

[114] R. D. Morris, A. M. Audet, I. F. Angelillo, T. C. Chalmers, and F. Mosteller, "Chlorination, chlorination by-products, and cancer: a meta-analysis." *American Journal of Public Health*, vol. 82, no. 7, pp. 955–963, Jul. 1992.

[115] J. Morrison, "Managing leakage by District Metered Areas: a practical approach," *Water 21*, pp. 44–46, Feb. 2004.

[116] S. R. Mounce, J. B. Boxall, and J. Machell, "An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation," in *Proc. Water Management Challenges in Global Change*, London, 2007, pp. 313–320.

[117] ——, "Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 3, pp. 309–318, May/Jun. 2010.

[118] S. R. Mounce, A. Khan, A. S. Wood, A. J. Day, P. D. Widdop, and J. Machell, "Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system," *Information Fusion*, vol. 4, no. 3, pp. 217 – 229, Sep. 2003.

[119] G. R. Munavalli and M. S. M. Kumar, "Optimal scheduling of multiple chlorine sources in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 129, no. 6, pp. 493–504, Nov./Dec. 2003.

[120] R. Murray, J. Uber, and R. Janke, "Model for estimating acute health impacts from consumption of contaminated drinking water," *ASCE Journal of Water Resources Planning and Management*, vol. 132, no. 4, pp. 293–299, Jul./Aug. 2006.

[121] A. Ostfeld and E. Salomons, "Solving the inverse problem of deliberate contaminants intrusions into water distribution systems," in *Proc. ASCE World Water and Environmental Resources*, 2005, p. 12.

[122] ——, "Conjunctive optimal scheduling of pumping and booster chlorine injections in water distribution systems," *Engineering Optimization*, vol. 38, no. 3, pp. 337–352, Apr. 2006.

[123] A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S.-T. Khu, D. Savic, D. Eliades, M. Polycarpou, S. R. Ghimire, B. D. Barkdoll, R. Gueli, J. J. Huang, E. A. McBean, W. James, A. Krause, J. Leskovec, S. Isovitsch, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. B. Trachtman, Z. Y. Wu, and T. Walski, "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," *ASCE Journal of Water Resources Planning and Management*, vol. 134, no. 6, pp. 556–568, Nov./Dec. 2008.

[124] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1-2, p. 100, Jun. 1954.

[125] R. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55 –63, Jan. 2002.

[126] M. M. Polycarpou, J. G. Uber, and U. Desai, "Design of a feedback controller for water distribution systems residual maintenance," in *Proc. ASCE Water Resources Planning and Management*, 1999, pp. 61–67.

[127] M. Polycarpou, J. Uber, Z. Wang, F. Shang, and M. Brdys, "Feedback control of water quality," *IEEE Control Systems Magazine*, vol. 22, no. 3, pp. 68–87, Jun. 2002.

[128] Z. Poulakis, D. Valougeorgis, and C. Papadimitriou, "Leakage detection in water pipe networks using a Bayesian probabilistic framework," *Probabilistic Engineering Mechanics*, vol. 18, no. 4, pp. 315–327, Oct. 2003.

[129] J. C. Powell, J. R. West, N. B. Hallam, C. F. Forster, and J. Simms, "Performance of various kinetic models for chlorine decay," *ASCE Journal of Water Resources Planning and Management*, vol. 126, no. 1, pp. 13–20, Jan./Feb. 2000.

[130] T. D. Prasad, G. A. Walters, and D. A. Savic, "Booster disinfection of water supply networks: Multiobjective approach," *ASCE Journal of Water Resources Planning and Management*, vol. 130, no. 5, pp. 367–376, Sep./Oct. 2004.

[131] A. Preis and A. Ostfeld, "A contamination source identification model for water distribution system security," *Engineering Optimization*, vol. 38, no. 8, pp. 941 – 947, Dec. 2007.

[132] ——, "Multiobjective Sensor Design for Water Distribution Systems Security," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, p. 17.

[133] ——, "Multiobjective contaminant sensor network design for water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 134, no. 4, pp. 366–377, Jul./Aug. 2008.

[134] A. Preis, A. J. Whittle, A. Ostfeld, and L. Perelman, "An efficient hydraulic state estimation technique using reduced models of urban water networks," *ASCE Journal of Water Resources Planning and Management*, 2011, to be published.

[135] M. Propato and J. G. Uber, "Linear least-squares formulation for operation of booster disinfection systems," *ASCE Journal of Water Resources Planning and Management*, vol. 130, no. 1, pp. 53–62, Jan./Feb. 2004.

[136] M. Propato, J. G. Uber, F. Shang, and M. M. Polycarpou, "Integrated control and booster system design for residual maintenance in water distribution systems," in *Proc. ASCE World Water and Environmental Resources*, 2001, pp. 390–399.

[137] R. S. Pudar and J. A. Liggett, "Leaks in pipe networks," *ASCE Journal of Hydraulic Engineering*, vol. 118, no. 7, pp. 1031–1046, Jul. 1992.

[138] R. Puust, Z. Kapelan, D. Savic, and T. Koppel, "Probabilistic leak detection in pipe networks using the scem-ua algorithm," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, p. 15.

[139] S. Rao, *Engineering Optimization: Theory and Practice*, 3rd ed. New York, USA: Wiley, 1996.

[140] D. Rosen and S. A. Zenios, *Handbook of Asset and Liability Management*. The Netherlands: North-Holland, 2006, ch. Enterprise-wide Asset and Liability Management: Issues, Institutions, and Models, pp. 1–23.

[141] L. A. Rossman, *EPANET Users Manual*, U.S. Protection Agency, Cincinnati, OH, Jan. 1994, (Version 1.1).

[142] ——, *EPANET 2 Users manual*, EPA/600/R-00/057, National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, Sep. 2000.

[143] L. A. Rossman and P. F. Boulos, "Numerical methods for modeling water quality in distribution systems: A comparison," *ASCE Journal of Water Resources Planning and Management*, vol. 122, no. 2, pp. 137–146, Mar./Apr. 1996.

[144] L. A. Rossman, P. F. Boulos, and T. Altman, "Discrete volume-element method for network water-quality models," *ASCE Journal of Water Resources Planning and Management*, vol. 119, no. 5, pp. 505–517, Sep./Oct. 1993.

[145] L. A. Rossman and W. M. Grayman, "Scale-model studies of mixing in drinking water storage tanks," *ASCE Journal of Environmental Engineering*, vol. 125, no. 8, pp. 755–761, Aug. 1999.

[146] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall/Pearson Education, 2003.

[147] A. B. A. Sakarya and L. W. Mays, "Optimal operation of water distribution pumps considering water quality," *ASCE Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 210–220, Jul./Aug. 2000.

[148] D. Savic, Z. Kapelan, and P. Jonkergouw, "Quo vadis water distribution model calibration?" *Urban Water Journal*, vol. 6, no. 1, pp. 3–22, Feb. 2009.

[149] F. Shang, J. Uber, and M. Polycarpou, "Particle backtracking algorithm for water distribution system analysis," *ASCE Journal of Environmental Engineering*, vol. 128, no. 5, pp. 441–450, May 2002.

[150] F. Shang, J. G. Uber, and L. A. Rossman, *EPANET Multi-Species Extension User's Manual*, National Risk Management Research Laboratory, Office of Research and Development, U.S. Enviromental Protection Agency, Cincinnati, OH 45268, Jan. 2008, EPA/600/S-07/021.

[151] F. Shang, J. G. Uber, B. G. van Bloemen Waanders, D. Boccelli, and R. Janke, "Real time water demand estimation in water distribution system," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, pp. 95.1–95.14.

[152] Y. Shastri and U. Diwekar, "Sensor placement in water networks: A stochastic programming approach," *ASCE Journal of Water Resources Planning and Management*, vol. 132, no. 3, pp. 192–203, May/Jun. 2006.

[153] L. Shvartser, U. Shamir, and M. Feldman, "Forecasting hourly water demands by pattern recognition approach," *ASCE Journal of Water Resources Planning and Management*, vol. 119, no. 6, pp. 611–627, Nov./Dec. 1993.

[154] (2008, Jul.) Finnish town faces fecal foul-up. Spiegel Online. [Online]. Available: http://www.spiegel.de/international/europe/0,1518,527057,00.html

[155] S. Sreepathi, K. Mahinthakumar, E. Zechman, R. Ranjithan, D. Brill, X. Ma, and G. von Laszewski, "Cyberinfrastructure for contamination source characterization in water distribution systems," in *Proc. International Conference on Computational Science*, 2007, pp. 1058–1065.

[156] S. Srirangarajan, M. Allen, A. Preis, M. Iqbal, H. B. Lim, and A. J. Whittle, "Water main burst event detection and localization," in *Proc. ASCE Water Distribution Systems Analysis*, 2010, p. 11.

[157] Y. Tachibana and M. Ohnari, "Development of prediction model of hourly water consumption in water purification plant," in *Proc. IEEE Industrial Electronics Society*, vol. 2, 1999, pp. 710–715.

[158] ——, "Prediction model of hourly water consumption in water purification plant through categorical approach," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, 1999, pp. 569–574.

[159] J. Thornton, R. Sturm, and G. Kunkel, *Water Loss Control*, 2nd ed. New York, USA: McGraw-Hill, 2008.

[160] E. Todini, "A more realistic approach to the 'extended period simulation' of water distribution networks," in *Proc. Computing and Control for the Water Industry*, London, UK, 2003, pp. 173–183.

[161] E. Todini and S. Pilati, "A gradient method for the analysis of pipe networks," in *Proc. Computer Applications for Water Supply and Distribution*, Leicester, UK, 1987, p. 20.

[162] C. Toregas and C. ReVelle, "Optimal location under time or distance constraints," *Papers in Regional Science*, vol. 28, no. 1, pp. 131–143, Dec. 1972.

[163] J. G. Uber, "Identifiability of contaminant source characteristics in steady-state and time-varying network flows," in *Proc. ASCE World Water and Environmental Resources*, 2005, pp. 25–28.

[164] K. A. Umberg, "Performance evaluation of real-time event detection algorithms," Master's thesis, University of Cincinnati, 2006.

[165] U.N. General Assembly, "International decade for action, 'water for life', 2005-2015," Feb. 2004, A/RES/58/217.

[166] ——, "The human right to water and sanitation," Aug. 2010, A/RES/64/292.

[167] S. Uryasev, *Probabilistic Constrained Optimization: Methodology and Applications*. Boston, USA: Kluwer Academic Publishers, 2000.

[168] U.S. Environmental Protection Agency, "Chemical summary of chlorine," Aug. 1994, EPA 749-F-94-010a.

[169] ——, "National primary drinking water regulations: Disinfectants and disinfection byproducts; final rule," Federal Register, pp. 69 390–69 476, 1998. [Online]. Available: http://www.epa.gov/OGWDW/mdbp/dbpfr.html

[170] ——, "Water security initiative: Interim guidance on developing an operational strategy for contamination warning systems," Sep. 2008.

[171] ——, "Water security initiative: Interim guidance on developing consequence management plans for drinking water utilities," Oct. 2008.

[172] U.S. Government, "National primary drinking water regulations - Title 40, Code of federal regulations, Part 141 - Enviromental Protection Agency (EPA)," 2002.

[173] B. G. van Bloemen Waanders, R. A. Bartlett, L. T. Biegler, and C. D. Laird, "Nonlinear programming strategies for source detection of municipal water networks," in *Proc. ASCE World Water and Environmental Resources*, 2003, p. 10.

[174] P. Vankayala, A. Sankarasubramanian, S. R. Ranjithan, and G. Mahinthakumar, "Contaminant source identification in water distribution networks under conditions of demand uncertainty," *Environmental Forensics*, vol. 10, no. 3, pp. 253 – 263, Sep. 2009.

[175] J. Vasconcelos and P. Boulos, *Characterization and Modeling of Chlorine Decay in Distribution Systems*. Denver, CO, USA: American Water Works Association, 1996.

[176] J. P. Vítkovský, A. R. Simpson, and M. F. Lambert, "Leak detection and calibration using transients and genetic algorithms," *ASCE Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 262–265, Jul./Aug. 2000.

[177] T. Walski, W. Bezts, E. T. Posluszny, M. Weir, and B. Whitman, "Understanding the hydraulics of water distribution system leaks," in *Proc. ASCE World Water and Environmental Resources*, 2004, pp. 476–485.

[178] T. M. Walski and J. W. Male, *Water distribution systems handbook*. New York, USA: McGraw-Hill, 2000, ch. Maintenance and rehabilitation/replacement, pp. 17.1–17.28.

[179] T. Walski, D. Chase, and D. Savic, *Water Distribution Modeling*. Waterbury, CT, USA: Haestad Press, 2001.

[180] L. Wang, *Adaptive fuzzy systems and control: design and stability analysis*. Englewood Cliffs, NJ, USA: Prentice Hall, 1994.

[181] Z. Wang, M. Polycarpou, and J. Uber, "Decentralized model reference adaptive control of water quality in water distribution networks," in *Proc. IEEE International Symposium on Intelligent Control*, 2000, pp. 127–132.

[182] Z. Wang, M. Polycarpou, J. Uber, and F. Shang, "Adaptive control of water quality in water distribution networks," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 1, pp. 149–156, Jan. 2006.

[183] Z. Wang, M. M. Polycarpou, F. Shang, and J. G. Uber, "Design of feedback control algorithm for chlorine residual maintenance in water distribution systems," in *Proc. ASCE World Water and Environmental Resources*, 2001, pp. 377–386.

[184] Z. Wang, M. Polycarpou, J. Uber, and F. Shang, "Adaptive periodic control for chlorine residual maintenance in drinking water distribution networks," in *Proc. IEEE Conference on Decision and Control*, 2001, pp. 4069–4074.

[185] Water Authorities Association, "Leakage control policy and practice," Water Authorities Association, Tech. Rep. 26, Jul. 1980.

[186] J.-P. Watson, H. J. Greenberg, and W. E. Hart, "A multiple-objective analysis of sensor placement optimization in water networks," in *Proc. ASCE World Water and Environmental Resources*, 2004, pp. 456–465.

[187] M. Weickgenannt, Z. Kapelan, M. Blokker, and D. A. Savic, "Risk-based sensor placement for contaminant detection in water distribution systems," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 6, pp. 629–636, Nov./Dec. 2010.

[188] A. J. Whittle, L. Girod, A. Preis, M. Allen, H. B. Lim, M. Iqbal, S. Srirangarajan, C. Fu, and W. K. J., "WATERWISE@SG: a tested for continuous monitoring of the water distribution system in singapore," in *Proc. ASCE Water Distribution Systems Analysis*, 2010, p. 15.

[189] World Health Organization, *Guidelines for drinking-water quality, Vol. 1, 3rd edition incorporating 1st and 2nd addenda*, 3rd ed. Geneva, Switzerland: World Health Organization, 2008.

[190] Z. Y. Wu, P. Sage, and D. Turtle, "Pressure-dependent leak detection model and its application to a district water system," *ASCE Journal of Water Resources Planning and Management*, vol. 136, no. 1, pp. 116–128, Jan./Feb. 2010.

[191] Z. Wu and P. Sage, "Water loss detection via genetic algorithm optimization-based model calibration," in *Proc. ASCE Water Distribution Systems Analysis*, 2006, p. 11.

[192] H. Yamauchi and W. Huang, "Alternative models for estimating the time series components of water consumption data," *Journal of the American Water Resources Association*, vol. 13, no. 3, pp. 599–610, Jun. 1977.

[193] E. M. Zechman and S. R. Ranjithan, "Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system," *ASCE Journal of Water Resources Planning and Management*, vol. 135, no. 5, pp. 334–343, Sep./Oct. 2009.

[194] Q. Zhang and S. Pehkonen, "Oxidation of diazinon by aqueous chlorine: kinetics, mechanisms, and product studies," *ACS Journal of Agricultural and Food Chemistry*, vol. 47, no. 4, pp. 1760–1766, Apr. 1999.

[195] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 4, pp. 576–593, Apr. 2002.

[196] S. L. Zhou, T. A. McMahon, A. Walton, and J. Lewis, "Forecasting operational demand for an urban water supply zone," *Journal of Hydrology*, vol. 259, no. 1-4, pp. 189 – 202, Mar. 2002.

[197] M. Zierolf, M. Polycarpou, and J. Uber, "Development and autocalibration of an input-output model of chlorine transport in drinking water distribution systems," *IEEE Transactions on Control Systems Technology*, vol. 6, no. 4, pp. 543–553, Jul. 1998.

Appendix A′

Algorithms

Iterative Deepening of Pareto Solutions

```
function IDPS_Search returns pareto_front

M= number of quality sensors

pareto_front= empty_list
network_nodes= all_nodes_in_network

for k= 1 to M
  for m= pareto_front
    for n= network_nodes
      current_node= [m, n]
      calculate objective functions for current_node
    end
  end
  if k=1
    network_nodes = network_nodes above T threshold
  end
  sort on demand_coverage
  select the first L solutions
  calculate pareto_front
end
```

Evolutionary Multi-Objective Optimization

```
function EMMO_Search returns pareto_front

Do
  parent_set = N random solutions
  pareto_ranks= measure non-dominance rank in parent_set
  offsprint_set=random selection from parent_set
  new_offsping_set = genetic operators on offspring_set
  parent_new_offspring= parent_set UNION new_offspring_set
  crowding_metric=measure  proximity in parent_new_offspring
  parent_set= N solutions from parent_new_offspring...
              ... considering pareto_rank and crowding_metric
Loop for L epochs
```

Quality Regulator

```
function Quality_Regulation returns input_signal

Do
  Measure real DMA_demands
  Compute estimation_error using DMA_demands
  Update adaptive_parameters of Fourier Series using estimation_error
  Simulate model using Fourier series with adaptive_parameters
  Construct process_model_matrix
  Compute next_inputs by solving the optimization problem
  input_signal = next_inputs for time k
Loop for each time step k
```

Leakage Fault Detector

```
function Leakage_Detection returns alarm

Do
  Measure real DMA_input
  Remove outliers
  Normalize with trend_seasonal signal
  Compute estimation_error
  Update adaptive_parameters of Fourier Series using estimation_error
  Compute CUSUM metric
  if CUSUM> threshold
    then alarm
  end
Loop for each time step k
```

Contamination Fault Isolation

```
function decision_tree returns next_sampling_node

Construct contamination_fault set
For each contamination_fault
   Compute impact_damage
   Compute impact_label
End For
Compute propagation_matrix

Do
   Compute entropy_vector
   Compute information_gain_vector
   Compute worst_impact_vector
   Compute pareto_front w.r.t.
       information_gain_vector and worst_impact_vector
   Select next_node from pareto_front
Loop until information_gain is zero
```