A study on optimal sensor placement strategies for water quality monitoring in water distribution networks

Thesis submitted in partial fulfilment
of the requirements for the degree
of
DOCTOR OF PHILOSOPHY

Submitted by

DINESH KUMAR GAUTAM
(Reg. No. 186107104)

DEPARTMENT OF CHENICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI-781039, ASSAM, INDIA

March 2024

This thesis is dedicated to

"Bai & Ma"

Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India
Department of Chemical Engineering

DECLARATION

I hereby certify that the work presented in this thesis entitled **" A study on optimal sensor placement strategies for water quality monitoring in water distribution networks"** is the outcome of my original research work performed at the Department of Chemical Engineering, Indian Institute of Technology Guwahati, under supervision of **Prof. Senthilmurugan Subbiah** and **Dr. Prakash Kotecha.** The results documented in this thesis are not submitted to any other university or institute for the award of any degree or diploma. Due acknowledgement has been made wherever the work described is based on the findings of investigations of others with supporting references.

Date:
Place:

Dinesh Kumar Gautam
Roll No. 186107104
Department of Chemical Engineering
Indian Institute of Technology, Guwahati

Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India
Department of Chemical Engineering

CERTIFICATE

This is to certify that the work contained in this thesis entitled **"A study on optimal sensor placement strategies for water quality monitoring in water distribution networks",** is being submitted by **Dinesh Kumar Gautam (Roll No. 186107104)** for the award of Ph.D. degree, is a record of bonafide original research carried out by him at Department of Chemical Engineering, Indian Institute of Technology Guwahati, under our guidance and supervision. This work embodied in this thesis has not been submitted to any other University or Institute for the award of any other degree or diploma.

Date:

Place:

Prof. Senthilmurugan Subbiah
Professor
Department of Chemical Engineering
Indian Institute of Technology Guwahati

Dr. Prakash Kotecha
Associate Professor
Department of Chemical Engineering
Indian Institute of Technology Guwahati

ACKNOWLEDGMENT

I would like to express my deepest gratitude towards my supervisors, Prof. Senthilmurugan Subbiah and Dr. Prakash Kotecha, for providing me with the opportunity to work on an exciting topic at IIT Guwahati. Their invaluable insights, encouragement and hunger for improvement have helped me acquire skills that will be helpful throughout my life. Their constant support in terms of ideas, resources and motivation has been a boon for PhD tenure. I want to thank my supervisors for their trust in me throughout the period.

I want to thank my doctoral committee members, Prof. Pallab Ghosh, Dr. Deepak Sharma, Dr. Sumit Kumar and Dr. Resmi Suresh, for their constructive suggestions during the progress seminars, which greatly enhanced the quality of my work. I am grateful to Prof. Kaustubha Mohanty, Head of the Department of Chemical Engineering, and Prof. Anugrah Singh, Head (Former) Department of Chemical Engineering, for providing me with all the facilities during the course of my work. I would also like to thank the staff of Department of Chemical Engineering for their valuable assistance.

I want to take these opportunities to thank all my seniors and lab mates from each research group. Firstly, from Prof. Senthilmurugan's research group (Water-and-Energy Nexus Lab), I would like to thank Dr. Vishal K. Verma, Dr. Vigneshwaran K., Dr. Habtom Teklu, Dr. Arunkumar Chandrasekaran, Dr. Aanisha Akhtar, Dr. Viswanth R., Dr. Nivedhitha S., Dr. Muniraja Tippa, Dr. Surendhar G., Dr. Senthil S., Mr. Munubarthi Kranthi K., Dr. Ananya Bardhan, Mr. Balakumara V., Dr. Priyamjeet, Ms. Neelam Dutta, Ms. Seema, Mr. Shanmugam, Mr. Bijoyendra, Mr. Venkatesh, Mr. Koushik Ghosh and Mr. Sai Mukesh Reddy for their constant support, laughter, and motivations. I would also like to thank Mr. Debasis Maharana, Mrs. Remya Kommadath, Dr. Swarna Makkitaya and Mr. Rajani from Dr. Prakash Kotecha's research group. I would also like to thank our lab assistants, Mr. Banajit Saloi, Mr. Rupam, Mr. Bishnu Sarkar, and Mr. Krishna Sarkar for their timely assistance.

This work has been part of the Indo-European project LOTUS (Project No : DST/IMRCD/India EU/LOTUS/208/(G) and EU Grant No 820881). I would like to convey my sincere thanks to the LOTUS project team for the technical assistance provided.

I am grateful to my friends Arnab, Shankar, Suraj, Sumita, Ashwin, Harish, Utla for their constant support, laughter, and love. I am indebted to my family, without whose unconditional support and sacrifices this study would have been impossible. I would like to express my

deepest appreciation to my wife, Mrs. Pallavi Sinha Roy for her commitments and enduring presence in every step of this academic endeavor.

(Dinesh Kumar Gautam)

Department of Chemical Engineering,

Indian Institute of Technology Guwahati

ABSTRACT

The primary objective of the United Nations (UN) sustainable development goal 6 is to ensure universal access to clean water and sanitation. Due to the ongoing degradation of water quality in various common water sources such as rivers, lakes, ponds, and groundwater, the demand for water treatment has significantly increased. Water treatment plants have a vital role in purifying water from these sources to make it safe for consumption. After treatment, the water is supplied to the end users through the piped water distribution network (WDN). A WDN is an interconnected nexus of pipes, pumps, hydraulic valves, reservoirs, overhead tanks, flow and quality sensors and end-user outlets. Piped water supply is implemented in developed countries while most of the developing countries are migrating towards 100% piped supply. Catering to a huge population with essential services such as water supply makes WDN a critical infrastructure. The safety of the water supply to the public is a top priority, and one way to monitor the quality of the supplied water is by strategically placing sensors along the WDN. This thesis focuses on developing strategies for optimal sensor placement in WDNs for water quality monitoring.

The water supplied through the WDN is disinfected prior to distribution. Low levels of disinfectant can result in incomplete deactivation of microbes, while excess levels can cause the formation of disinfectant byproducts. The WDN often has pipes located underground or passing through culverts. In these areas, stagnant water can seep into the network if pipes with small holes are exposed to a vacuum during non-operating periods. This contaminated water can then flow downstream and affect all the consumers connected to the network. Also, as the WDN are critical infrastructure, they are susceptible to terrorist attacks. That is the intended intrusion of toxic chemicals into the WDN. The sensors installed in the WDN should help in overcoming these water quality issues.

In the first chapter, the placement of wired water quality sensors in WDN is investigated. It is evident that when water gets contaminated at a certain point, all the downstream nodes receive the contaminated water. The monitoring of the WDNs requires the placement of sensors at strategic locations to detect maximum contamination events at the earliest. Detection of maximum contamination events requires the sensors to be placed at downstream nodes, while early detection requires sensors to be positioned near the contamination source. The competing objectives necessitate multi-objective optimization (MOO) of sensor placement. The sensor

placement problem (SPP) is a complicated problem owing to its combinatorial nature, interconnected and large WDN sizes, and temporal flows producing complex outcomes for a given set of contamination events. This study proposes a novel method to reduce the complexity of the problem by condensing the nodal search space. This method first segregates the nodes based on intrusion events detected, using k-means clustering, followed by selecting nodes from each group based on the improvement observed in the objectives, namely, contamination event detection, expected detection time, and affected population. The selected nodes formed the decision variable space for the MOO study.

The performance of the developed strategy was evaluated on two benchmark networks, BWSN Network1 and C-town network, and compared to the traditional method using the hypervolume contribution rate (CR) indicator and the number of Pareto points. The optimal subset of nodes generated twice as many Pareto points as the complete set of nodes for placing 20 sensors and had a 10% higher CR indicator than the traditional method. For placing 5 sensors, the proposed solutions performed better at higher detection likelihood values, which are necessary for maximum detection. The proposed sensor placement algorithm can be easily applied to large WDNs and is expected to provide a superior solution regardless of network size compared to the traditional approach.

In the initial chapter, the available flow data was utilized to create pollution matrices and identify nodes with significant impact. However, in developing nations, acquiring accurate water demand and flow data for distribution systems is often impeded by utility challenges and infrastructure upkeep, which presents obstacles to the effective deployment and adaptability of water quality sensors. This situation is also observed for newly commissioned WDNs. With the uncertainty involving flow data, these WDNs require a flexible monitoring solution. Wireless sensor networks (WSNs) are gaining attraction for their simple and swift installation, efficient data collection, and data processing, as well as for indicating the overall health of water distribution systems. Wireless sensors for water quality monitoring transmit the measured parameters via GSM, GPRS, or LoRaWan. A WSN architecture made of node stations and a single base station with multi-hopping data transmission is considered in this study. The optimal placement of WSN is a combinatorial problem of placing 'N_s' sensors in 'N' nodes, as discussed in the first objective, but additionally constrained by the transmission range (TR) of the sensors. Since there is a lack of information, WSN placement is demonstrated for maximization of objectives: network coverage and sensor detection redundancy, as only network and flow direction data, are prerequisites for its estimation. The study proposes

efficient greedy algorithms to generate sensor placement solutions for a given N_s and transmission range. However, optimizing network coverage often leads to a decrease in SDR and vice versa. The optimization of the competing objectives is performed using the NSGA-II algorithm. The initial population for the NSGA-II algorithm is populated with the solutions from the greedy algorithms, which are then compared to Pareto solutions obtained from the NSGA-II algorithm without using the initial greedy solutions. The results of the study show that incorporating the initial solutions from the greedy algorithm improves the performance of the NSGA-II algorithm, surpassing all the Pareto points achieved by the standalone NSGA-II algorithm.

Single base station architecture does not perform well when the transmission range of the sensors is very small compared to the WDN area. The narrow transmission range constricts the spread of the sensors and thereby restricts monitoring to only a small portion of the WDN. Deploying a multi-base station can facilitate the monitoring of extreme portions of WDN. The third algorithm developed in this study attempts to generate a sensor placement solution for this case. A mixed integer linear programming (MILP) formulation is developed with additional modifications to improve the computation time for solving. The objectives from the second chapter are re-iterated in this chapter. The results revealed that the additional modifications led to a more than 90% decrease in computation time for maximizing network coverage. The computation time for maximizing SDR with and without modifications was comparable. Further, MOO is performed using the augmented ε-constraint method and compared with the results of the ε-constraint method. The advantages of deploying multiple base stations are demonstrated by comparing with solving for single base station architecture. Finally, MILP formulation is adapted to typical scenarios observed in WDN, such as regular monitoring, critical node monitoring, and sensor failure.

Keywords: *Water distribution network, water quality monitoring, sensor placement, wireless sensor network, multi-objective optimization*

ABBREVIATIONS

WDN	Water Distribution Network
GOI	Government of India
BIS	Bureau of Indian Standards
DMA	District Metered Areas
DBP	Disinfection By-products
THM	Trihalomethanes
HAA	Haloacetic acid
IoT	Internet of Things
ICT	Information and Communication Technology
SCADA	Supervisory Control and Data Acquisition
SOP	Standard operating procedures
RTU	Remote transfer unit
TEVA-SPOT	Threat Ensemble Vulnerability Assessment – Sensor Placement Optimization Tool
CWS	Contamination warning system
SPP	Sensor placement problem
LOS	Level of service
SHS	Steady hydraulic simulation
EPANET	Environmental Protection Agency Network
BWSN	Battle of Water Sensor Networks
WQ SPP	Water quality sensor placement problem
LP	Linear Programming
MILP	mixed-integer linear programming
MOSPP	Multi-objective Optimization of Sensor Placement Problem
GRASP	Greedy Randomized Adaptive Search Procedures
PSO	Particle Swarm Optimization
NRW	Non-Revenue Water
WTP	Water Treatment Plant
NSGA	Non-dominated Sorting Genetic Algorithm
EDM	Event Detection Matrix

TDM	Time to detection Matrix
CR	Contribution Rate
PF	Pareto Front
OSN	Optimal Subset of nodes
CAS	Controllability Analysis Method
SGWDN	South Guwahati water distribution network
OTR	Overhead Tank Reservoir
CPM	Contamination Propagation Matrix
TR	Transmission Range
NC	Network Coverage (%)
SDR	Sensor Detection Redundancy (%)
MGA	Multi-Objective Genetic Algorithm
gMGA	Greedy initialized Multi-Objective Genetic Algorithm
AUGMECON	Augmented ε-constraint
ECON	ε-constraint

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview of Water Distribution Networks (WDNs)

A water distribution network (WDN) is a complex interconnected nexus of reservoirs, overhead tanks, junctions (demand outlets, fire hydrants, etc.), pipes, and valves built for the reliable distribution of water from the source to the end user. The source of WDN is generally a water treatment plant that lifts water from various sources such as groundwater, river water, reservoirs, etc. The treated water is then pumped into WDN through pipelines extending from WTP to end users. WDNs play a vital role in providing 'clean water for all' as it is considered the safest way to supply water to a large population that can be monitored and maintained.

The UN Sustainable Development Goals Chapter 6 aims at providing 'clean water and sanitation for all' with one of the targets to 'achieve universal and equitable access to safe and affordable drinking water for all by 2030'[1]. In accordance with this, the Government of India (GOI) has planned to increase the percentage of urban and rural households having access to safe drinking water. The National Sample Survey 76th round, December 2018, on 'Drinking Water, Sanitation, Hygiene and Housing Condition in India' showed that about 42% of urban homes consume 'piped water into dwelling' for water consumption. Moreover, recent data shows that about 62.55 % of rural households are catered with piped drinking water [2]. Though this data provides information about the number of piped connections, the quality of water supplied also needs to be quantified and addressed. The water quality set by the Bureau of Indian Standards (BIS) for potable water can be maintained when water is supplied through a healthy piped network or tanker with appropriate sensors and control system for quality control [3].

From ancient times, civilizations like the Greeks and Romans had water distribution setups comprising open channels, tunnels, and clay and lead pipes. However, it was not until the 16th and 17th centuries that a large-scale pumped water delivery system was established. Along with this, the principles of fluid flow developed by Bernoulli, flow regime by Reynolds, and subsequent headloss equations by Darcy and Weisbach and Hazen and Williams in the subsequent years paved the way for creating even more complex and efficient piped water distribution systems. The latest technologies employ these models to develop a digital twin of the WDN to monitor, manage, and predict anomalies in the WDN [4,5].

On the water quality front, water was considered drinkable as long as it was clear and pleasant-tasting. Early water treatment methods included filtration, boiling, storing in copper containers, and sedimentation to remove particles. Later, in the sixth century, distinctions between potable

and non-potable water resources were made, indicating advancements in water management standards. In the 17th century, more advanced methods such as percolation, distillation, and coagulation were studied. The breakthrough in the 'safe water' perspective came after the invention of the microscope, which enabled scientists to view the microscopic life in water. By the 19th century, many studies had led to the discovery of waterborne disease, and the disinfection of drinking water had been initiated. The disinfection process has been carried forward to this day with more focus on optimal disinfectant levels and swift detection of contaminants [4,5].

1.2 WDNs in Developed vs Developing Countries

WDNs in developed and developing countries have distinct architectures, operational practices, and challenges. In developed nations, WDNs generally operate 24/7, providing chemically disinfected water directly to households. These networks employ continuous pressurization to prevent external contamination and feature advanced monitoring systems equipped with flow, pressure, pH, and disinfectant sensors. If contamination occurs, the operators isolate the affected area and issue warnings to end-users.

In contrast, WDNs in developing countries like India often operate intermittently—perhaps for a few hours daily or twice a week. Households locally store water, making it susceptible to contamination even after chemical treatment. The use of decentralized treatment units is common before consuming the water for drinking purposes. These intermittent WDNs often lack advanced water quality monitoring solutions and are generally equipped with minimal instrumentation. Manual measurements at network entry points, such as primary storage tanks, are common practice.

The problems encountered while maintaining water supply are low pressure at end-user, failure of pumps, pressure surges, leakages, low disinfectant levels, and contamination due to seepage or intentional intrusion. The WDN management practices to overcome these issues can be segregated into proactive and reactive measures. In the former, measures are taken to maintain the WDN health and circumvent major issues generally observed in WDN. One of the primary methods in proactive measures is dividing the WDN into District Metered Areas (DMAs). This approach is widely recognized as an efficient way to optimize the operation of WDNs and to detect, localize, and mitigate leakages [6]. Surge tanks are installed at critical locations to release excess pressure build-up to overcome pipe burst scenarios [7].

Similarly, to maintain water quality, the water treatment facility ensures that pH, TDS, turbidity, E.coli level, and disinfectant level are within the standards of governing bodies when water is pumped into the WDN. Disinfection of drinking water has significantly reduced the fatalities caused by waterborne diseases. Among the disinfectants, free chlorine is the most commonly employed. As per the BIS standard, the free chlorine concentration at the tap of the end user should exceed 0.2 mg/l, with a maximum permissible limit of 1 ppm in the absence of an alternative water source [3]. The free chlorine content of the water supplied from the water treatment plant (WTP) may not be consistent with the free chlorine levels observed at the user end. This is due to the consumption of free chlorine while deactivating microbes, biofilms, reaction with the pipe material, and biological matter [8]. The lower disinfectant levels will lead to improper disinfection, generally observed at the periphery of the WDN.

On the other hand, elevated chlorine levels might react with Natural organic matter (NOM) present in water and produce disinfection byproducts (DBPs) like trihalomethanes (THMs) and Haloacetic acids (HAAs) [8]. These lead to problems like bladder cancer, miscarriages, etc. A study revealed that individuals consuming water with THM4 levels exceeding 50 μg/L had a 47% higher incidence of bladder cancer compared to those consuming water with THM4 levels below 5 μg/L [9,10]. The free chlorine levels in the network are maintained by estimating the requirement at the source and dosing accordingly. For large WDNs, additional disinfectant booster pumps are installed at crucial locations for intermediate dosing when the chlorine level drops.

Reactive measures are employed in situations when proactive measures fail. These measures include a) isolation of affected part of WDN and leakage detection in case of leakages and pipe bursts, b) optimal operation of chlorine booster units for dosing when chlorine is inadequate, c) optimal flushing operation during contamination, and d) alerting public regarding the threat and providing steps to follow. The developed countries use advanced monitoring systems, combining proactive and reactive measures, to monitor the WDNs. In developing countries, however, only proactive measures are implemented. Shortage of water supply, huge leakages, ignorant end-users, and continuous expansion force WDN operators to distribute water intermittently.

Operating the WDNs intermittently pressurizes and de-pressurizes the network continuously, inducing a ‘water hammering’ phenomenon. Over a prolonged duration, it has a very adverse effect on the infrastructure of WDN, which is typically designed for more than 20 years. Weak

infrastructure, characterized by loose valves and joints, pipe fracture, and internal corrosion, leads to water leakage and pipe bursts. In the non-supply hours, the seepage of water bodies surrounding the pipes transfers the contaminants into the pipes. When the supply regime resumes, the contaminants are transported throughout the network, affecting the people and, in turn, straining the medical infrastructure of the region [11,12].

The architectural backbone of intermittent WDNs is generally robust enough to support 24/7 operations. The advent of low-cost Internet of Things (IoT) and Information and Communication Technology (ICT) solutions is dramatically changing the landscape. Before 2020, implementing the Supervisory Control and Data Acquisition (SCADA) system was common but expensive, contributing to 10-15% of the total project costs and thus viable mainly for large networks.

Today, IoT technologies make it feasible to deploy sophisticated monitoring solutions at a fraction of the cost. Platforms like Microsoft Azure facilitate the integration of various sensors directly into cloud-based systems using GPRS, Wi-Fi, or LoRa technologies. This enables real-time operational and monitoring capabilities, offering better control over leak detection, water quality, and chemical dosing [13–16].

In summary, while developed countries have long benefited from advanced WDNs, the convergence of low-cost IoT and ICT solutions is beginning to narrow the gap, providing developing nations with the tools to upgrade their water infrastructure significantly.

1.3 Real-time Monitoring and Operational Aspects

Monitoring of WDNs has been widely researched with the application of sensors in the network, concentrating on collecting crucial supply data to facilitate swift mitigation procedures. Several standard operating procedures (SOPs) for water supply management have been developed to issue regulatory practices in the wake of undesirable events. Traditional WDNs are often plagued by inefficiencies stemming from their reliance on manual monitoring and control. However, the SOPs can be implemented online in real-time by detecting these anomalies [17,18]. The advancements in sensors for water quality measurements coupled with the efficient SOPs for water supply assurance have steadily shifted the focus towards monitoring and maintaining the quality of supplied water in WDNs [19].

The water quality sensors for continuous monitoring WDN measures various water quality parameters of supplied water such as turbidity, conductivity (indirectly gives total dissolved solids), pH, temperature, dissolved oxygen, and residual chlorine [20,21]. The advancements

in Nanomaterials and polymer science are enabling the development of specialized sensors for measuring contaminants like Arsenic, Iron, Fluoride, etc., emerging pollutants like microplastics, and microbial contaminants such as E.coli [22–24]. Investigations have led to various reports that shed light on terrorist organizations targeting critical infrastructure like WDNs. Sensors must also sense the presence of hazardous chemicals like cyanide, mercury, etc., which might be injected by terrorist groups to cause large-scale issues [25].

Real-time sensing of intrusion of such chemicals will alert the water management body for swift action and the public to avert usage of the contaminated water. Delays in detecting these events might lead to more people consuming contaminated water, severe hospitalization, and even death. It is of utmost necessity that all the contaminations are identified and mitigation procedures are in place well before widespread consumption of contaminated water.

The advent of IoT and ICT technologies has allowed for a more holistic and automated approach. In this new architecture, sensors placed throughout the WDN provide real-time data on various parameters like flow rate, pressure, pH levels, and disinfectant concentrations. These sensors are interconnected through various communication protocols such as GPRS, Wi-Fi, or LoRa, which connect to a centralized cloud-based platform.

Accurate water quality parameters measurement necessitates that the sensors are well calibrated for the respective parameters. Sensors for pH, conductivity, or chlorine are well established with standard products available in the market like conductivity transmitter AWT210 (ABB), Residual chlorine monitor AW401 (ABB), Depolox® 700 M Analyzer (Evoqua), and several others are listed in [26]. However, for niche contaminants, advanced sensors have been developed based on the concept of multiparameter sensor array. Multiparameter sensors are a set of arrays with lower selectivity and partial specificity (for cross-selectivity response) towards chemicals present in the solution, here water, and employ mathematical techniques like multi-variate analysis or pattern recognition. Though achievable, the calibration of such sensors is a complex and tedious process requiring a large design of experiments owing to complex water chemistry [27].

The sensor measurements produce continuous data that is sent to a central database for engineers to evaluate the health of the WDN. There are two methods for transmitting this data: (i) a conventional method where the sensors are connected to remote transfer units (RTUs) through wired connections, which then transfer the data to a database. The health of the WDN is visualized using a SCADA system in a control room. (ii) A modern wireless approach that

utilizes technologies like LTE, 4G/5G, LORA, and others. In this method, the sensors directly transmit the data to a database located in the cloud. The health of the WDN is then visualized using appropriate client tools [14]. A sample IoT architecture for monitoring WDN is shown in Figure 1.1. The cloud server receives data from all the installed sensors and controls the dosage and water supply by operating the control valves.

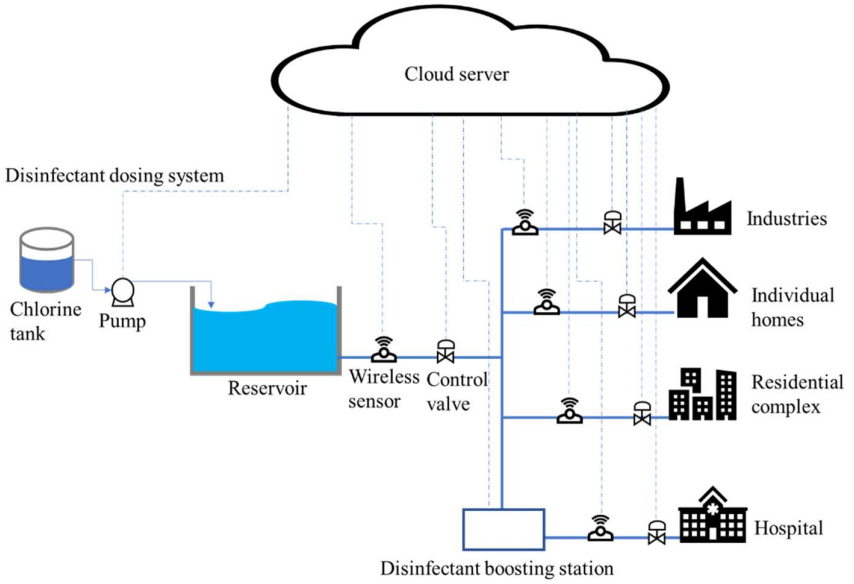


Figure 1.1 IoT architecture for monitoring and maintaining WDN

Traditional monitoring with SCADA generally requires laying out wires from the control centre to remote sensing units, which are then connected to monitoring devices. Installation of such an infrastructure is complex, laborious, and requires additional security for the wires laid throughout the network. Consequently, the restoration of damages is also an arduous process. RTUs with wireless communication features can overcome these issues. However, the data and corresponding control of attached sensors is limited. This limitation arises from non-interaction with other components, like different sensors, pumps, valves, etc., of WDN. Secondly, the power consumption at RTUs is much higher owing to bulky components, on-device processing, and data transmission.

In contrast, integrating IoT and ICT into WDNs ensures not only the consistent quality of water supply but also enhances the system's resilience against unforeseen disruptions. Various components in the WDN are connected and thus can communicate with each other. In case of any deviations from set parameters, instant alerts are generated, enabling immediate corrective actions. Automated controls allow for optimized pump operations, reducing energy consumption. Predictive maintenance, enabled by data analytics, reduces long-term operational

costs by pre-emptively identifying and addressing issues before they escalate into major problems. This architecture also offers benefits in terms of energy and cost-efficiency. IoT devices are developed with low-power components and can be operated with batteries powered by on-site solar panels. Moreover, in developing countries, where the constant expansion of WDNs is expected, reallocation and integration of sensors is much simpler than SCADA-based sensor deployment.

While cloud-based solutions offer benefits over conventional SCADA architectures, cybersecurity is crucial in cloud-based systems. Ensuring data integrity, encryption, authentication, malware protection, data latency, recovery strategies, and compliance with industrial data transmission standards are key security considerations for cloud server-based control systems. Additionally, incorporating fail-to-safe operation mode is essential in maintaining system security.

In summary, employing IoT and ICT technologies in WDNs revolutionizes their architecture, making them more robust, efficient, and responsive. The capabilities for real-time data collection, advanced analytics, and automated decision-making transform how these networks are operated and controlled, elevating them to modern standards of efficiency and reliability.

1.4 Significance of Sensor Placement

After the deadly 9/11 terrorist attack in the US, the Environmental Protection Agency (EPA) focused on the early identification of contamination in the water distribution system. A report on EPA's Threat Ensemble Vulnerability Assessment – Sensor Placement Optimization Tool (TEVA-SPOT) research program estimates that the expected fatalities due to WDN contamination can be reduced by 48% by the implementation of a robust Contamination warning system (CWS) [28].

Several additional independent studies have been conducted to monitor WDN in real-time and develop systems for event detection and decision support [29–32]. Given the vastness of the WDN and limited monitoring resources, the WDN design engineers have to choose the locations where the water quality sensors can be placed to achieve the monitoring requirements.

It should be noted that once a monitoring and mitigation system is in place, sensor location becomes essential. Optimal maintenance of disinfectant level requires sampling and testing of water in WDN. It might require intermediate disinfectant boosting to replenish the consumed free chlorine and bring it up to the accepted levels. The sampling and testing can be performed continuously by installing sensors along the WDN [33,34]. Similarly, any intrusion of

hazardous contaminants also requires sensors for swift detection. The placement of the sensors at niche locations is essential; however, due to budgetary constraints, only a limited number of sensors can be placed in a vast network [35]. This led to research in various sensor placement strategies to monitor WDNs with a limited number of sensors.

Even with IoT architecture, the mitigation procedures can only commence when the anomaly is detected. With the delay in detection, the contamination might spread through a large part of the network. Correspondingly, the response action comprising valve operations for isolation of contaminated segments and hydrant flushing of contaminated water will be extensive and extend the shutdown time. Similarly, in the context of economic viability, this translates to the allocation of excessive person-hours, additional clean water for flushing, and a delay in alerting the public to stop consuming water.

The design of SPP for WDNs often relies on information about the pipe architecture, flow, and pressure in the pipeline. However, in many cases, this information is not available for new networks or even for older networks in developing countries like India. This lack of information is due to inadequate maintenance, unaccounted leaks, unauthorized access, governance inefficiencies, and ongoing infrastructure enhancements. These limitations present challenges in ensuring the operational efficiency and reliability of WDNs and make it difficult to strategize for sensor placement [29,36–41]. The efficiency of monitoring data-deficient networks can be improvised by having flexible sensor placement solutions.

Wireless sensors offer this flexibility, which can be relocated according to monitoring requirements. The other advantages of wireless sensor networks are affordability, adaptive placement strategies, conducting measurements remotely, real-time optimization, and incorporation with IoT systems [13–16]. A WDN enabled with WSN consists of a set of water quality sensors with the feature of transmitting the measurements to a central data repository via wireless technologies like LoRaWAN, GPRS, ZigBee, etc. [14,42].

1.5 Complexities in Modern WDNs and Objective of the Thesis

The water supply systems generally operate by pumping water from the source to the overhead reservoirs, which are supplied to all the end users. The complexity in the WDN arises from thousands of kilometers of laid pipes, time-based demand patterns, customer-based demand consumption (individual apartments, offices, hospitals, schools, etc.), and looped connections that result in the internal mixing of supplied water [28].

WDNs are designed to accommodate population growth over a minimum service life of 20 years. Advanced medical facilities have increased life expectancy while depreciating the death rates. Consequently, the population catering to the water supply might exceed the population forecasted during the design phase. In addition to this, rapid urbanization leads to population influx and unexpected settlements in cities. Water consumption in these regions creates an uneven distribution of water demand within the WDN compared to the actual design [43]. Water consumption and supply patterns are also greatly affected by climatic conditions. Elevated temperatures trigger higher water usage, while intense monsoons might lead to flood water entering primary water sources and into the WDN through seepage at low pressure. For example, an algal bloom incident at Erie Lake led to a ban on drinking tap water, which catered to about half a million people [44,45]. Maintaining water quality is crucial in light of the uncertainties discussed above.

The sensor placement problem (SPP) for WDN can be defined as the "placement of sensors in the network that keep the contamination risks to a minimum." The measurement of the risks depends on the monitoring goals apt for a given WDN. These risks can measured in terms of regions or portions of the network monitored, detection of maximum number of contaminations, detection of events in a minimum time, and level of service (LOS) based risk assessment [46,47]. Typically, LOS-based risk assessments involve a combination of two risk measurements, such as the quantity of contaminant consumed prior to detection and the minimum time required for detection.

The evaluation of the risks requires information regarding the flow of contaminated water, such as flowrate, flow direction, contaminant concentration, etc. These details can be estimated by performing contaminant event simulations. The contaminant event simulations in the WDN considered for SPP can be broadly classified as steady hydraulic simulation (SHS) and dynamic hydraulic and water quality simulation. In the SHS approach, the network characteristics play a vital role in sensor placement, where the flow path determines the extent of risk containment. The pipe connectivity and demand locations, elevation, and pressure drop in the network determine the water flow direction. By deducing the flow direction, the path of the contaminant can be established, and sensors can be placed along this path.

The second approach involves simulating the network with the deliberate intrusion of contaminants, observing its effects on the network, and minimizing the observed effects by placing sensors at appropriate locations [28]. As stated previously, the water consumption

pattern in a WDN varies throughout the day. Furthermore, each type of consumer (residential buildings, apartments, schools, offices, hospitals, etc.) has a different water consumption pattern—the combination of these patterns leads to varying water velocities over time. Dynamic hydraulic and water quality simulations are required to capture the temporal complexities of WDNs.

WDN models are crucial for developing advanced operational tools in water distribution systems. Several software packages are available for modeling WDNs, including EPANET, WaterGEMS, WaterCAD, Pipe Flow Expert, Branch, HydrauliCAD, Pipe2018, Synergi Water, and HYDROFLO3. EPANET is the most popular among these options due to its robust and adaptable engine. EPANET has been publicly accessible since 2000 and has been extensively utilized in research and creating sophisticated applications such as event detection systems, extended dynamic simulation, and decision support systems [48,49].

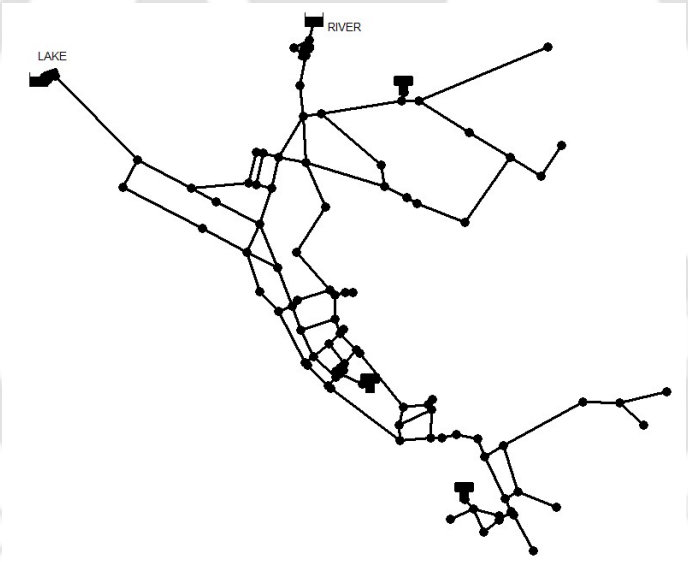


Figure 1.2 Example network from EPANET

Figure 1.2 displays an EPANET network as an illustrative example. Nodes represent demand points, splitters, and joints, while links represent pipe connections between these nodes. Additionally, EPANET allows for the inclusion of valves, pumps, emitters, overhead tanks, and reservoirs. Water quality parameters, such as water age and chemical concentration, can also be set. The choice of a chemical parameter depends on the simulation objective, such as chlorine dissociation or contaminant propagation study. There are two versions of EPANET, EPANET2.0 and EPANET2.2. The former utilizes a demand-driven approach that ensures that demands at all the nodes of WDN are fulfilled. However, this can lead to a negative available

head, resulting in error, although pressure does not go negative in reality. Instead, the flow to the node reduces to compensate for pressure loss. This attribute is incorporated in EPANET2.2, which follows a pressure-driven approach. The flow or supply to any node depends on the node's available head. Sufficient pressure ensures complete demand is met, but if pressure is low, supply is reduced according to the pressure-flow relationship [50].

1.6 Summary and objectives of the thesis

Overall, optimal placement of sensors facilitates efficient monitoring and management of WDNs. The critical challenge for placing sensors in a WDN arises in selecting the best solution from a vast number of feasible solutions available to utility managers. The efficiency of the 'best solution' has to be assessed and compared with other available solutions, and this 'efficiency' can be estimated based on the required goals or objectives for placing the sensors. Though the SPP can be constructed as an optimization problem, finding the optimal solution is computationally expensive when deterministic methods are employed, while stochastic approaches result in sub-optimal solutions. Also, the sensor placement problem is known to be computationally demanding and classified as NP-hard [51,52]. Further intricacies of the sensor placement are discussed elaborately in the next chapter.

The objective of the thesis is to explore and propose efficient strategies for sensor placement in WDNs. These strategies should address the challenges posed by the complexity of networks, data deficiency, and the need for optimization algorithms. The thesis aims to investigate different objective functions, estimation methods, and optimization algorithms for sensor placement. Additionally, the thesis seeks to analyze the placement of wireless sensor networks in WDNs, considering the unique challenges of dual-network optimization and limited flow data availability. The thesis also aims to explore the reduction of nodal search space using clustering techniques and the importance of multi-objective optimization for sensor placement. Ultimately, the objective is to propose practical and effective approaches for optimizing sensor placement in WDNs.

Chapter 2: Literature review

2.1 Foreword

WDNs play a crucial role in ensuring the reliable supply of clean water to communities. The efficient monitoring and management of these networks are essential to prevent contamination, identify leaks, and optimize operation. Water quality sensor placement is a crucial aspect of WDN monitoring, as it determines the locations where sensors should be installed to collect relevant water quality data. This chapter presents a comprehensive literature review on the water quality sensor placement problem for WDNs, discussing its challenges and the existing approaches and techniques to address them.

The chapter discusses the early approaches to water quality sensor placement, which focused on maximizing network coverage using steady-state hydraulics. However, these approaches simplified the complexity of WDNs and made certain assumptions that may not hold true in real-world scenarios. The limitations of these early approaches set the stage for the development of more advanced methods.

Objective functions and estimation methods used in water quality sensor placement are then explored, including the Battle of Water Sensor Networks (BWSN) design challenge [53], which compared different approaches for solving the water quality sensor placement problem. The conflicting objectives of WDN monitoring and the importance of multi-objective optimization are highlighted.

This chapter also covers various optimization techniques for determining the best sensor placements, including genetic algorithms and constrained mixed-integer programming. It also highlights a potential software tool developed by researchers that serves as a valuable resource for assessing contamination risks and designing sensor monitoring stations.

The chapter then delves into the simplification of WDNs to reduce the complexity of the sensor placement problem. Techniques such as graph trimming, topography-based clustering, and node selection based on connectivity and betweenness centrality are presented. However, the limitations of these simplification methods are also acknowledged, emphasizing the need for careful consideration of their applicability.

The water quality sensor placement problem for data-deficient networks, where limited information is available about the network, is also addressed. The challenges of data deficiency are discussed, and studies that have focused on maximizing network coverage and sensor detection redundancy using limited data are presented.

Finally, the chapter concludes by discussing the unique challenges posed by wireless sensor networks (WSNs) for WDNs. Specialized solutions are required to address these challenges, and existing studies on WSN placement in WDNs are highlighted.

By providing an overview of the water quality sensor placement problem for WDNs, the associated challenges, and the existing approaches and techniques used to solve it, this literature review serves as a valuable resource for researchers and practitioners in the field of water distribution network monitoring and management. The insights gained from this review will contribute to the development of more effective and efficient water quality sensor placement strategies, ultimately improving the reliability and safety of WDNs. This chapter discusses the state-of-the-art sensor placement methodologies for optimal water quality monitoring of WDNs.

2.2 Sensor placement problem for WDN

The early approaches to the water quality sensor placement problem (WQ SPP) were developed based on maximum coverage of the WDN using steady-state hydraulics (SHS). The approach considered that if water passing through a selected node can be deemed safe, it can be assumed that water in the upstream nodes was safe. Thus, the sensors were placed such that the maximum number of nodes with maximum demand were covered [54,55]. The optimization problem in this approach is expressed as a linear programming (LP) model utilizing the network's adjacency matrix. The adjacency matrix is a square matrix that represents the connections between nodes in a graph. In water quality sensor placement analysis, this matrix is crucial for optimization algorithms. It provides information about the network's topology, showing how components such as junctions or tanks are interconnected. Using the adjacency matrix, the sensor placement problem can be formulated as an LP model to maximize node coverage while considering various constraints. The efficiency and applicability of the adjacency matrix method deem it extremely valuable for solving complex sensor placement problems.

Berry et al. [56] formulated the SPP into a facility location problem. In this approach, the sensor nodes are treated as facilities, while the connecting nodes are regarded as sites that are linked to these facilities. A distance matrix is created based on the contamination risks associated with each sensor and node pair. The problem is then solved using mixed-integer linear programming (MILP) to achieve a global optimal solution.

The use of SHS methods to optimize sensor placements in WDNs takes a reductionist approach that fails to adequately capture the complexities of WDNs. One significant limitation is the assumption of a unidirectional hydraulic flow, which does not align with the operational realities of multi-faceted WDNs with overhead storage tanks and pumping stations. For instance, diurnal variations in supply sources, with pumping stations predominant in the early hours and overhead tanks in the evening, contradict the unidirectional flow assumption of SHS models. Additionally, the SHS approach does not consider the temporal variability in hydraulic flow rates, a phenomenon observed in continuous and intermittent water supply systems. This variability leads to peak flow conditions during morning and evening hours, affecting the transit time of contaminants to specific points within the network [47].

Given these complexities, there is a need for more advanced contaminant transport simulations that accurately depict the spatiotemporal distribution of contaminants and their residence times in the network. These simulations offer a robust framework that enhances the reliability and effectiveness of sensor placement strategies, especially in WDNs with variable flow dynamics and multiple supply sources. This understanding emphasizes the importance of moving away from SHS-based models and towards comprehensive simulation techniques for a thorough and dependable optimization of sensor placement in complex WDNs.

2.2.1 Objective Functions and Estimation

The main components of the sensor placement problem that have been extensively researched are objectives, optimization algorithm, and decision variables. As described earlier, studies [54,55] focused on maximizing the demand coverage of WDN by optimally placing the sensors. The amount of water supplied through a node is equal to the nodal demand; thus, securing a node is equivalent to securing the demand of the node. Estimation of water anomalies is vital, but the safety of the public is ensured only by quick detection of the contamination. Kumar et al. [57] incorporated the time to detect contamination by considering only those events that are detected within a fixed from the onset of contamination. The time to detection is defined as the time between the occurrence of contamination and its detection by the sensors. Lower time to detection translates to early detection of the contaminants in the WDN. Rathi et al. [58] defined a level of service (LOS) criterion as ‘detection of contamination with a time limit’. Stricter LOS or shorter time limits increased the number of monitoring stations (sensors) and vice versa. Violation of LOS meant that the contaminated water was being served to the public for a more extended period. With this idea, Ostfeld et al. [59] defined

LOS based on the ‘maximum amount of contaminated volume consumed before detection’. Any contamination event that exceeds this LOS is a potential threat and has to be detected. In this study, random contamination scenarios were simulated, and the corresponding water quality and consumption values were recorded. The study defined the objective detection likelihood as the ‘percentage of the contamination events detected’. Following this, the sensor design focused on maximizing the detection likelihood, which violated the LOS limits. This study also discussed redundancy in the detection of events. Similar to detection likelihood, the sensors detection redundancy is given by the ‘percentage of events detected by at least two sensors’. The demand at the nodes is the cumulative consumption of the population fetching water from that node. Berry et al. [60] used population densities at corresponding nodes and minimized the overall population exposed to contamination for a given set of contamination events. The TEVA-SPOT software tool was developed by the U.S. EPA to evaluate the contamination risks and design sensor monitoring stations for WDN. TEVA-SPOT allows the users to choose an objective among the following: a) number of failed contaminated detections, b) time of detection, c) estimated population exposed, d) estimated population killed, e) extent of contamination, f) mass of contaminated water consumed, g) length of pipe contaminated and h) volume of water contaminated. Objectives c, d, f, and h correlate with the detection time. In contrast, objectives e and g are associated with objective a. TEVA-SPOT performs single objective optimization of the above objectives with user-defined constraints on other objectives [61].

From the above-discussed objectives, it is evident that all the objectives derived based on the time spent by the contaminant in the WDN have a specific LOS factor for estimating sensor placement design. Sensor solutions generated to minimize the detection time have two biases emerging from objectives associated with detecting events, such as network coverage and detection likelihood. The first case is observed while estimating the time taken only for detected events. In this case, all the sensors will be placed near the source. Second, consider the time taken for both detected and undetected events are accounted for. Here, the sensor solution will be forced to detect all the events, as the undetected events will either incur a penalty or be assigned the maximum simulation time value. Though the objective is to minimize detection time, the sensor placement design will be generated to maximize detection likelihood or network coverage. Similarly, placing two sensors nearby to improve sensor detection redundancy competes with placing two sensors at two dead-end nodes to improve detection likelihood. Thus, assessing the effect of one objective over the other is crucial for achieving

optimal monitoring of WDN [62]. Watson et al. [63] analyzed the correlation of various sensor placement objectives like detection likelihood, detection time, population affected, number of failed detections, etc. The study revealed that optimization of one of the objectives can yield highly sub-optimal values of other objectives for small to medium numbers of sensors. The correlation between the objectives improves for a large number of sensors as the similarities in the sensor placement design of various objectives will be high. Following this, the concept of a trade-off solution focused on optimizing one of the objectives with limits on the other objective.

To further examine the significance of objectives in sensor placement for water quality monitoring, the Battle of Water Sensor Networks (BWSN) design challenge was conducted to compare various approaches to solving SPP for WDNs [53]. In their study, the following data were provided to benchmark different objectives performance: the network flow configuration and nodal demands of a WDN to estimate flow, contaminant concentration at any node, and the time taken for the contaminant to travel in the network. Based on this data, The SPP problem was solved for the following four objectives: (i) minimization of expected detection time, (ii) minimization of the expected population affected, (iii) minimization of the expected amount of contaminated water consumed, and (iv) maximization of detection likelihood. Fourteen research groups submitted their approach for solving the SPP for two different networks of 126 nodes and 12523 nodes. The contamination intrusion pattern was the injection of the contaminant at a concentration of 230 g/L at 120 L/h flow rate for a time period of 2 hours. The time step considered in the design was 5 min.

The results of this challenge summarized the findings of 14 research groups involved [53]. The study highlighted the conflict between the first three objectives, 'affected population,' 'contaminated water consumed', and 'detection time,' with detection likelihood. Since then, many researchers have explored the Multi-objective Optimization of SPP (MOSPP) for WDN, focusing on detecting events and either of the other three objectives [46].

Furthermore, the literature discusses using simulation tools such as EPANET 2.0 to estimate objective function values. These tools provide a means to evaluate the performance of different sensor placement strategies. The effect of contamination in WDN can be projected by simulating a series of contamination scenarios in EPANET 2.0 (https://www.epa.gov/water-research/epanet) [64]. However, simulating the network during optimization is computationally expensive as all the events were simulated for each set of solutions to estimate

the objective function value [28]. An alternative to this process is to "observe" the effect of contamination events on all the nodes and record the response compactly in a matrix. Then, based on these nodal observations, the nodes are selected so that overall risk is minimized; based on these nodal observations, the nodes are selected to reduce overall risk [53]. The pollution matrices have the number of rows equal to the number of nodes in the network and columns equal to the number of contamination events. Each value represents information like contamination detection, time taken for detection if contamination is detected, etc., by the corresponding node (row) for the corresponding event (column). So, each row provides values for estimating the objective function if a sensor is placed at the corresponding node. Thus, by adopting a similar procedure for all the nodes (rows), the values required for the estimation of the objective function for each node can be stored efficiently. This step will reduce computation costs during optimization, as the pollution matrix will be utilized to estimate the objective function rather than simulating all the contamination events to evaluate the objective function for each sensor solution. The recorded information in pollution matrices depicted contamination presence, time of detection, water consumed before detection, or the number of persons affected before detection, etc. [53,65]. A detailed explanation of developing pollution matrices is described in Appendix A1.

2.2.2 Optimization methods for finding the optimal sensor locations

Various optimization algorithms have been implemented to solve the single objective WQ SPP and WQ MOSPP including mixed-integer MILP formulations [54,66], genetic algorithm (GA) and its variants [51,67,68] greedy randomized adaptive search procedures (GRASP) [66,69], heuristics [70,71] and Particle swarm optimization (PSO) [72,73]. Initial MIP approaches considered only SHS for deriving the contamination transport within the WDN. The order of the corresponding data size was equal to the number of nodes in the network. As temporal flows in WDN are expected, sensor solutions generated with this method are applicable only to the sub-optimal state of the WDN. The temporal nature of the WDN can be assessed by performing hydraulic and water quality simulations in conjunction with a set of contamination events. Incorporating dynamic analysis increases the size of the problem, and solving it becomes a cumbersome process. Secondly, given the nature of WQ SPP, a solution that balances the competing objectives is desired. Many researchers have employed stochastic techniques or meta-heuristic optimization methods to solve WQ SPP. Formulation of stochastic SPP is straightforward, with decision variables denoting the locations of nodes, and the objective function is estimated based on the information derived from the simulations.

Stochastic techniques generate near-optimal solutions with low memory requirements and CPU time for SPP and MOSPP. In contrast, the MILP formulations consume huge memory to provide optimal solutions for SPP and constrained MOSPP. Greedy search techniques have been presented by [69,74,75] to generate sensor design solutions for large WDNs with a vast number of contamination events. These techniques exploit the submodular nature of the objective function and the approximation guarantees that the generated solution will be within 63.2% of optimal value. This technique was extended for MOSPP by defining the greedy search objective as the normalized weighted sum of the competing objectives. The weights are incorporated according to the trade-offs fixed for the considered objectives. Based on the estimation of objectives, the sensor placement can be categorized into mean and robust. In the former, the objectives are estimated at all the nodes and their corresponding contamination events, while the latter focuses on minimizing the maximum damage observed for a set of contamination events. Though the latter approach reduces the computational burden, it might incur a higher cumulative impact of contamination events.

Overall, irrespective of the methods applied for optimizing sensor locations, the complexity of WDN due to vastness, complex interflows, and temporal demands have kept the search for optimal sensor locations still a significant challenge [46,62].

2.2.3 Mathematical Simplification of WDN for SPP

The problem size of WQ SPP depends on the number of nodes considered for placing sensors, as a large number of nodes translates into a large number of node-sensor pairs. This makes the problem computationally and memory-wise expensive. The problem can be interpreted as a combinatorial problem of choosing a solution from ${}^{n}C_{k}$ combinations where 'n' is the number of nodes, and 'k' is the number of sensors [76]. However, if a few nodes are deemed non-potential and only 'm' such nodes are considered for sensor placement, the number of combinations reduces to ${}^{m}C_{k}$, which is less than ${}^{n}C_{k}$. For instance, if a network has 100 nodes, but only 90 nodes are chosen to place 5 sensors, then the number of combinations reduces by more than 41%. This led to the following studies that focused on improving the nodal search space by contracting the WDN using graph trimming methods and topography-based clustering, designing contamination events based on heuristics, and decreasing decision variables (nodes) search space considered for optimization.

Klise et al. [77] pruned large WDN using graph trimming techniques followed by a two-tiered optimization procedure. The first step involved grouping similar nodes based on topography

and flow into 'super nodes', finding the optimal 'super nodes' for placing sensors, and then fine-tuning the sensor locations using the original nodes forming the optimal 'super nodes'. While reducing the complexity of the network reduced the run time of the algorithm, a deviation of 5% was still observed in comparison with BWSN results. Secondly, the aggregation of the demands of nodes over-simplifies the network by considering only the cumulative effect of multiple nodes that are replicated. Due to this, a group of nodes where one node has a significant demand while others have zero and another group of nodes whose sum of demands is equal to the highest demand node of the former group get treated the same way.

Xu et al. [78] clustered the WDN into '*k*' regions for placing '*k*' sensors, and from each cluster, one node was selected based on 'Betweenness Centrality' and maximum connectivity with other nodes in the cluster. Betweenness Centrality of a node describes to what degree a node falls on the shortest path between other pairs of nodes. Selecting a node in the 'centre' may aid in quickly capturing the contamination events of the upstream nodes that are detected at the downstream nodes. Upon implementation of this idea for the whole WDN, only the nodes at the centre of the network will be selected. To ensure uniform distribution of sensors, the network was clustered on the basis of 'pipe connectivity' distance with groups equal to the number of the sensors. The placement of sensors between the nodes hints at the trade-off between event detection and detection time. The contamination simulations were not considered for sensor locations; instead, this approach provides an estimate of sensor placement in the absence of hydraulic data.

The selection of a single node from '*k*' clusters was then improvised by greedily selecting '*k*' nodes with better overall connectivity within the WDN, leading to better observability of the whole network [79]. The quality simulations were performed only at the peak periods to estimate the effect of contamination on each node. In this way, the complex and large computations required to select nodes for contamination risks were reduced to selecting nodes vastly connected in the network. Sun et al. [80] proposed clustering nodes based on pressure heads using the hierarchal clustering technique. The method was proposed for pressure monitoring of the WDNs, and thus, the impact of contamination was excluded. However, clustering the nodes based on the pressure variations provides insight into how clustering can be utilized to group similar behaving nodes.

The above methods oversimplified the SPP by overseeing the effects of temporal demands, reverse flows, and contaminant transport velocity, as their applicability was tested only on a small number of contamination events.

Diao et al. [81] implemented controllability analysis on BWSN Network1 to condense the decision variable search space by choosing nodes that facilitated maximum network observability in a given time frame. However, this prioritized the observability at the expense of swiftness in detections when a set of nodes is considered together for sensor placement. In contrast, Khorshid et al. [82] developed a node selection strategy based on Value Of Information (VOI) and Transinformation Entropy (TE) techniques. The nodes were compared pair-wise in terms of detection time and grouped according to the swiftness in detecting the given set of contamination events. The results of the VOI-TE method were compared with the results from the TEVA-SPOT tool, which uses GRASP-heuristics, but the comparison with the traditional multi-objective optimization is required to assess the efficiency of the VOI-TE method. Mandel et al. [83] defined quality zones within the WDN by clustering the nodes based on time-varying concentration data, but the study has not been extended to sensor placement problems. The quality zones were developed by tracing the water from the source nodes using EPANET2.0, but in a WDN, the contaminant intrusion can occur at any node. Moreover, the simulations were carried out at steady-state conditions, which are generally not observed in real WDNs.

Thus, there is a need for a pre-selection procedure to reduce the nodal search space that encompasses the dynamic water quality simulations with a focus on the sensor placement objectives. The observations or pollution matrices acquired from the contamination transport simulations provide the required data on the objectives, detection of contaminants, time of arrival of contaminated water, and population affected at the nodes. Based on this data, the nodes from the WDN can be clustered in terms of similarity in detecting events, followed by selecting nodes from these clusters with better objective function values.

2.3 Sensor placement for data-deficient network

Reduction of nodal search space requires nodal information with temporal variations. However, in many cases, this information is not available for new networks or even for older networks in developing countries. The lack of information is due to factors discussed here. The difference between the amount of water pumped into the WDN and the amount billed gives the non-revenue water (NRW). NRW is a direct indicator of inefficiencies in the supplied water. In

India, NRW accounts for about 50% of water supplied by water utilities. The significant contributors to NRW are water loss due to leakages, unmetered connections, and unplanned and illegal tapping [37,40]. Secondly, improper operation and maintenance lead to various discrepancies in the water supply. Bello et al. [39] argue that the problems of the WDN must be addressed within realistic time frames to restore the WDN operation at the earliest. Prolonged time frames will exert pressure on maintenance work, and correspondingly, the WDN infrastructure will be under enhancements for a longer duration. Thirdly, maintaining a record of water consumed by individuals is difficult when water is supplied intermittently, which is the norm in most developing countries [12].

Furthermore, the water quality data of the water supplied is also unavailable for most developing countries, as published in UN Water 2021 [84] (Figure 2.1). The water quality parameters are measured at the WTP outlet or via manual sample collection, followed by testing in the lab [41].

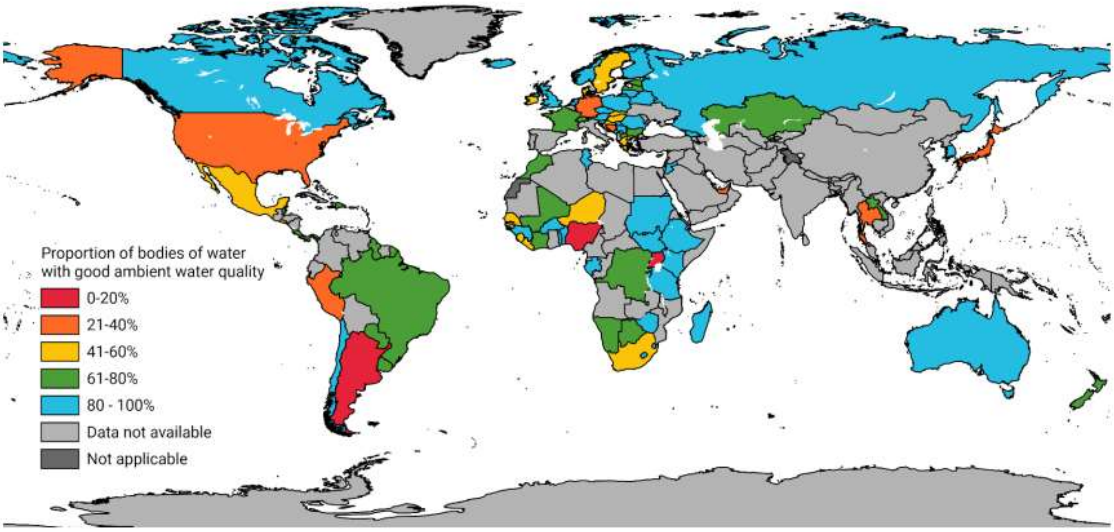


Figure 2.1 Water quality data across the world (UN Water 2021) [84]

Though many real networks have been carried out for sensor placement, very few are based on Indian WDN owing to the unavailability of data for large-scale networks. Rathi et al. [85] discussed a water distribution network from the Koriya district in Chhattisgarh, India, with 23 demand nodes and 25 pipe links. A single objective problem formulation where the weighted sum of detections and detection time was minimized. This work was also implemented in a more extensive network in Dharmpeth, Nagpur, India, where the number of sensors to be implemented was also part of the formulation [86]. This more extensive network consisted of

292 demand nodes and 377 pipes. They improvised the earlier method by including risk factors for each zone based on the end-users, like hospitals, schools, apartments, etc. [87]. These studies have developed demand data based on population and per capita daily water consumption rather than the SCADA data. Thus, the results have been estimated for an assumed demand and flow pattern. Though this approach is suitable for small case studies, its application to large networks becomes tedious as estimating the population density and consumption pattern varies widely across the network.

Under the circumstances of unavailability of flow data, the readily available data are often the system's piping architecture and demand locations or nodes. Utilizing this limited data, sensor placement strategies can be formulated to prioritize Network Coverage (NC), estimated as the percentage of nodes covered or monitored by the sensors [28]. On the other hand, the objective 'sensor detection redundancy' (SDR) facilitates the reliability and verification of sensor measurement data [59]. sensor detection redundancy enables reliability by balancing sensor locations among the far extreme of WDN and at the center of WDN. This behaviour is observed in most of the trade-off SPP solutions generated for WDN monitoring.

Thus, elevated network coverage enhances water quality oversight within the WDN, whereas sensor detection redundancy can aid in the localization of contamination events and the execution of mitigation procedures [73,88]. These two objectives have been proven to be competing in nature and, thus, require multi-objective optimization (MOO) to generate a set of non-dominating solutions or Pareto points [89]. In the context of systems with data deficiencies, deploying WSNs can enhance the monitoring capabilities of WDN operators by allowing the addition of new sensors or relocation of existing sensors as required, a feat difficult to achieve with wired sensor systems.

2.4 Wireless Sensor Networks for WDN

The earlier versions of water quality sensors transmitted data via wired communication protocols such as RS232, RS485, Modbus, etc. The new-age water quality sensors are enabled with wireless transmission of measured data via 3G/4G, GPRS, and LoRaWAN. These sensors are segregated based on their purpose in the wireless sensor network architecture: base stations, node stations, cluster heads, and relay nodes. The base stations aggregate and compile the data forwarded by all the node stations, cluster heads and relay nodes and transmit it to the cloud or central data repository. The node stations are sensors that transmit the measured data to the base station or the neighbouring node station. The cluster head sensors collect the data from

the node stations and pass it onto the data center. The cluster heads do not perform sampling but rather aid in the seamless transmission of data when the node stations are far away from the base station. Relay nodes, similar to cluster heads, aid in relaying the data from the node stations to the data center [90].

Most of the current WSNs employ a multi-hop strategy to transfer the data. The multi-hop strategy allows a node station to receive data from neighbouring node stations and transmit it with its own measured data. That is, the node stations effectively operate like relay nodes. In contrast, the single-hop strategy allows the node stations to forward the data to either the base station, cluster head, or relay node [13].

WSN placement in WDNs differs from standard WSN in non-piping domains and wired sensor placement in three main ways. First, the WSN is generally placed in a region of interest, and its locations are generated by creating grids or reachability spheres on the structure. WDNs restrict sensor placement to consumer tapping points, junctions, storage points (nodes in EPANET model) and pipes (links in EPANET model) of WDN. The sensing region of the water quality sensor is limited to the locations where it is installed, conventionally by tapping point. This reduces the sensing flexibility brought in by wireless sensors. Increasing the number of tapping points to aid flexibility will adversely increase the problem size. Second, the placement of WSNs in WDNs is a dual-network optimization problem, i.e. sensor connectivity network and piping network. The connectivity of the sensors in WSN ensures data transfer from the sensor to the data center. The piping infrastructure determines the contaminant flow from the source to the other nodes. Third, due to budget and design constraints, WDNs often have insufficient sensors for complete monitoring. It can be seen that to observe a network thoroughly, at least one sensor has to be placed at each dead-end. So, WSN for WDN would require a solution to achieve the maximum WDN monitoring objective while satisfying the wireless connectivity constraint. Therefore, optimizing wireless sensor placement in WDNs needs specialized solutions to address these unique challenges.

To overcome some of these challenges in WSN for WDN, Zeng et al. [35[35] formulated the SPP with mixed-integer non-linear programming (MINLP) to maximize the coverage with a constraint on the total sensor budget. The water quality sensors considered in this study have functionalities of either a node station or a base station. The ‘connectivity of sensors’ constraint was incorporated by assuming phantom flow between sensors followed by network flow balances. The ‘node station’ sensor nodes were sources, while the ‘base station’ sensor nodes

acted as the sinks. The MINLP formulation was solved using GA, as the formulation was found to be computationally exhaustive for large-scale WDNs. Given the combinatorial nature of the wireless WQ SPP, the example network used in the study was too small (10 to 24 nodes) to completely describe the performance of the method. Also, the direct application of meta-heuristic procedure on the MINLP formulation might not be an efficient method. Sankary et al. [91] developed a model to place wireless inline mobile sensors that traverse the pipelines and transmit data via a wireless protocol. The base station's range was assumed to be large enough to cover the entire network; thus, data from the wireless sensors can be directly transmitted. A similar study on mobile wireless sensors by Shahmirnoori et al. [92] proposed placing low-energy data sinks along the WDN to fetch the data transmitted by inline wireless sensors. Though the above two studies circumvent the transmission range constraint for inline mobile sensors, this is not always achievable, owing to the sheer size of WDNs.

The studies on the placement of WSNs for WDN applications are limited. However, the optimal placement of WSNs in non-piping domains, such as structural health monitoring, has been extensively studied [93]. The structures can be bridges, hospitals, etc., where the sensors are placed on the region of interest for maximal coverage of the system under consideration. These studies focus on maximum coverage, minimum energy consumption, and optimal sensor connectivity while tuning the node station and the base station locations [94]. The problem of OWSN has been described as NP-hard when optimal locations of all the nodes are to be placed. Both mathematical programming and meta-heuristic techniques have been incorporated to solve the Np-hard OWSN problem [95,96]. A few studies relevant to wireless WQ SPP are discussed in the following paragraphs.

Perez et al. [97] described a hybrid evolutionary algorithm for node station and relay node placement in WSNs, combining heuristics with NSGA-II. The connectivity of the WSN is assured by pre-selecting the plausible locations that satisfy the transmission range constraint based on the given base station location. The model minimizes the number of sensors and energy dissipation of the sensors to achieve 100% coverage. Li et al. [98] formulated a breadth-first search-based algorithm for the optimal placement of a single base station for a given set of node stations with location. The node stations are considered to be within the transmission range of other node stations or base stations. Hence, each node station can send data directly or via other node stations to the base station. The former study focuses on the placement of node stations for a given base station location, while the latter focuses on placing base stations for a given set of node stations.

Zhou et al. [99] formulated a MOO model for placing node stations using the modified multi-objective firefly algorithm to maximize the independent information gathered and energy consumption. The first objective is similar to the WQ SPP objective – maximum network coverage for a given number of sensors. Moreover, the independent information gathered is retrieved by performing a finite element analysis of the structure. This is analogous to simulating contamination events in WDN and estimating the sensor locations that will yield the maximum number of independent detections.

The connectivity of the WSN was identified using the judgment matrix given by,

$$D=\sum_{s=1}^{W-1}A^{s} \tag{2.1}$$

A^s – sth power of adjacency matrix 'A' of size 'WxW' where W is the number of node stations. Matrix 'A' describes whether two node stations are within transmission range and are capable of communication. However, the size of 'A' grows quadratically with the number of node stations (W); consequently, the computation requirement of 'D' increases exponentially with respect to 'W'. Also, the location of the base station is a prerequisite for implementing this model [99].

Flushing et al. [100] discuss the placement of relay nodes in WSN to improve performance in terms of data throughput using a heuristic approach. The locations of the node stations and base stations are pre-defined along with their data transmission path, and the optimal placement of relay nodes will aid in managing the data traffic. The crucial takeaway from this study is its connectivity constraint implementation. The connectivity between the nodes was modelled as a flow balance between the nodes, similar to that discussed in Zeng et al. [35]. This flow represented the amount of data flow between the sensor nodes, whose value depended on the routing paths. While exploring the locations of relay nodes, the routing paths also have to be optimized to improve the overall throughput of the system. In a subsequent study, two methods were discussed to address the complexity of the problem: a) integrating the MILP within NSGA-II and b) exchange of best solutions between MILP and NSGA-II. The study revealed that the MILP model had a considerable CPU time (>16000 s), with none of the instances converging to optimality. However, some instances were solved to optimality by hybrid models.

In the discussed studies, either the location of the base stations or node stations is known. But, in the dual-network problem of WSN for WDN, the locations of sensor nodes and their corresponding connectivity are estimated only during the optimization process. Secondly, the energy issue of WSN is due to the unavailability of the power supply, which generally arises from placing node stations in remote locations. In contrast, the sensors in WDN have access to the power supply as they are placed on the pipeline laid near buildings and residential complexes. While non-piping domains can benefit from existing optimization algorithms, the complexity of WDN sensor placement necessitates the development of new optimization techniques tailored to its specific challenges.

Furthermore, the above studies have a single and multiple base station architecture. Relaying the measured data via node stations allows for a swift and easy deployment of WSN in WDN with a single base station. However, deploying multiple base stations can realise the maximum benefits of WSN. Multiple base station architecture aids in the decentralization of the WDN monitoring system. In the event of sensor failure, only a part of the WDN under the failed sensor will not be monitored. Therefore, efficient strategies for optimal placement of both architectures are needed.

2.5 Literature Closure and Research Gaps

i) The placement of water quality sensors in WDNs is widely studied, focusing on monitoring objectives and optimization algorithms.

ii) The temporal and bi-directional flows observed in WDN make the SPP very complex. At the same time, the size of WDN leads to a factorial increase in the number of sensor placement solutions. Simplification of the SPP via reduction of problem size while retaining critical contaminant transport is essential.

iii) There is a need for water quality monitoring solutions in developing countries where data is scarce. Deployment of wireless sensor networks (WSNs) can be an efficient way to monitor data-deficient WDNs.

iv) Studies on the optimal placement of WSN in WDN are very limited. Thus, efficient algorithms addressing the requirements of developing countries, such as easy deployment and reallocation of sensors for continuous expansion, maintenance and sensor damages, are required.

2.6 Scope of the thesis

The literature study shows that monitoring crucial infrastructure such as WDN should be the primary goal for ensuring safe water supply and public safety. The sensors must be placed at crucial locations to monitor the water quality. These locations must have high coverage, while the overall response time to the contamination events should be minimal. Any relaxation in these objectives will result in a significant portion of the public being exposed to contaminants. This study aims to solve the complex WQ SPP with different objectives for optimal monitoring of WDNs. The estimation of these objectives requires prior information on the water flow data in the WDN. However, data is not always available, especially in developing countries. That being the case, the objectives must be modified to cater to data-deficient WDNs' needs. Also, the monitoring solutions should be flexible to absorb the intricacies observed in the WDNs operation. The uncertainty in data can be tackled with the placement of WSN in WDN. But this leads to complications of the already complex problem. The placement of WSN in WDN incorporates two network problems: the piping connections in WDN and wireless transmission in WSN. This dual-network problem has to be solved for networks with data deficiency. The next phase of sensor placement would require placing wireless sensors with reallocation features. This scenario will be observed when the water engineers want to place sensors at specific locations for mandatory monitoring or repositioning of already available sensors to monitor a critical region or when the sensors are damaged and require repositioning to fix the connectivity issue.

Overall, the main objective of the study is to develop novel strategies to generate sensor placement solutions for optimal monitoring of WDNs. In the due process, the study focuses on improving the nodal search space, defining objectives and constraints for placing wireless sensors data-deficient WDNs and implementing greedy search techniques and MILP cuts to generate SPP solutions efficiently. The scope of this thesis wholly pertains to securing the water supplied in WDNs with the help of strategically positioned monitoring stations.

Chapter 3: Optimal placement of wired sensors in WDNs

3.1 Foreword

The study presents a new approach for simplifying the process of selecting sensor locations in a WDN through multi-objective optimization. The approach combines k-means clustering and cluster-wise greedy selection to narrow down the search space for optimization. In contrast to traditional methods that solely focused on connectivity, this approach accounts for nodal observation and the transportation time of contaminants in the network. The selected nodes are then used for multi-objective optimization (MOO) to generate multiple sensor design solutions that achieve a balance between different objectives. The three objectives considered in this study are detection likelihood, detection time, and affected population. The MOO study was performed for the conflicting objectives: a) maximizing detection likelihood while minimizing detection time, and b) minimizing the affected population while minimizing detection time. The pattern search algorithm, a direct search method for optimization, is used to find the Pareto points. The effectiveness of the proposed node selection strategy is demonstrated by testing it on two benchmark networks, BWSN Network1 and C-town network, and comparing the resulting Pareto Front with the traditional method of performing multi-objective sensor placement study.

3.2 Methodology – Water Quality Sensor Placement for WDN

A typical WDN is generally spread over a vast area, and the corresponding hydraulic model consists of a large number of nodes (N_{node}). The complete set of nodes in the WDN are represented as positive integers, i.e. 1, 2 .. N_{node}. During the operation, the supplied water can be contaminated by intrusions at weak joints or deliberate intrusion along the WDN. Water quality sensors are placed in the WDN to identify these contamination events. Each sensor monitors the WDN at a certain level by indicating the water quality in its upstream node. A group of such sensors might monitor a larger portion of WDN based on their placement. Let the number of sensors be N_s, then the number of decision variables is N_s, with each variable having the range [1 N_{node}]. The placement of the sensors should achieve maximum WDN monitoring measured in terms of objective functions such as detection likelihood, detection time and population affected. These objective functions depend on the extent and speed of contamination propagation within the WDN. A series of contamination events are designed to capture its randomness, defined by the node, time of day, and quantity of contaminant. Overall, the sensor placement problem for a given WDN is to determine the set of N_s nodes (for deploying sensors) that achieve optimal values of the above-said objective functions estimated

over a given series of contamination events. The methodology, contaminant event simulations, and objective functions are described in detail in the subsequent paragraphs.

The proposed methodology for solving the sensor placement problem in water distribution networks (WDNs) is illustrated in Figure 3.1. The methodology consists of three steps: contamination transport simulation, node selection through clustering, and multi-objective optimization.

In the first step, various contamination events are simulated to analyze the impact of contamination in the WDN. MATLAB provides the conditions of each event (time, node, etc.) to EPANET2.0, which carries out extended hydraulic and water quality simulation. The network parameters, such as nodal concentration and time of contamination observance, are then used to develop pollution matrices in MATLAB.

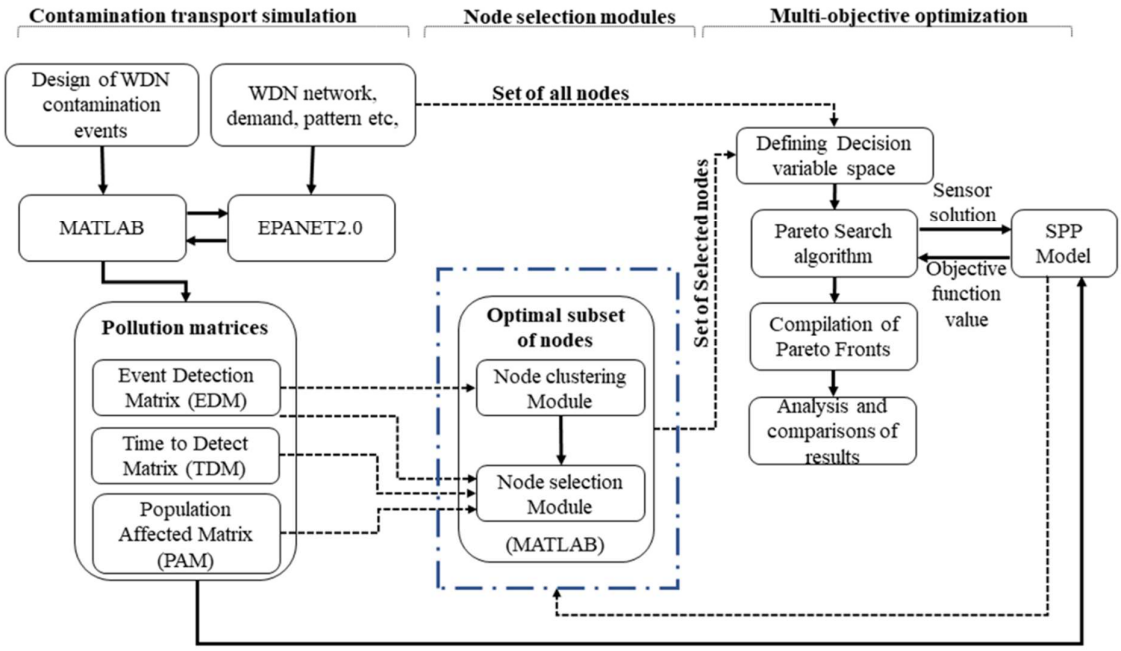


Figure 3.1 Overview of sensor placement problem for water distribution networks

The second step involves identifying a crucial set of nodes with higher detection likelihood and lower detection time. The nodes are first clustered using a node clustering module implemented in MATLAB. The cluster-wise selection of nodes is then performed in the node selection module, considering the objectives.

In the third step, a multi-objective optimization study is conducted to determine the optimal placement of sensors. The dotted lines in the figure represent the decision search space for the

optimization. Typically, only the first and third steps are carried out to find the optimal Pareto front, implying that all nodes in the WDN are potential sensor locations. The traditional method of estimating the Pareto front is also employed to compare it with the proposed strategy.

3.2.1 Contaminant Transport Simulation and Pollution Matrix

The two benchmark networks considered as test cases for the developed node selection strategy, BWSN Network1 and C-town network, are shown in Figure 3.2. The contamination scenario for BWSN Network 1 was a subset of BWSN Challenge Case A [53] with an intrusion at every half an hour interval. That is, the contamination events occur at a time gap of half an hour. Each contamination occurs for two hours. The contamination event is defined by the parameters: injection flow rate = 125 L/h, contaminant concentration = 230,000 mg/L, and injection duration = 2 hours. This translates to an intrusion of 479167 mg/minute of contamination for two hours. It is assumed that the ingested contaminant remains inert throughout the network and is consumed only at the demand nodes. Meanwhile, in C-town, a 15-minute interval for each contamination event was set to capture the intricacies of contaminant flow within the network. The number of sensors to be placed was fixed to 5 and 20, translating to 3.8% and 15.5% of nodes in BWSN Network1 and 1.2% and 5% of the C-town network, respectively. It is assumed that sensors have a sensitivity limit of 0.01 mg/L, and there is no lag between detecting a contaminant by the sensor and transmitting information to the utility manager.

Figure 3.2 Water Distribution Network Layout a) BWSN Network1 b) C-town Network

Table 3.1 describes the contamination events design for both the WDNs. The contamination event simulations were carried out in EPANET2.0 integrated with MATLAB 2020b. The results of the contamination events were recorded in three pollution matrices, depicting the detection of the event, time to detection, and population affected.

Table 3.1 Contamination events design for BWSN Network1 and C-town Network

Features	BWSN Network 1	C-town Network
Nodal contamination	All nodes (129)	All nodes (396)
Time of intrusion	Every half an hour	Every 15 minutes
Intrusion duration	Two hours	15 minutes
Contaminant mass injected	479167 mg/minute	10000 mg/minute
Detection delay	Real-time	Real-time
Detection sensitivity	0.01 mg/L	0.01 mg/L
Time period for extended simulation	96h	72h
Simulation time step	5 min	5 min
Number of sensors	5 and 20	5 and 20

The events detections matrix (EDM) is binary, where '1' represents if the contamination was observed and '0' otherwise. The time to detection pollution matrix (TDM) stores the time delay in contaminant observation since the onset of contamination for each node for all the intrusion events. Similarly, the population affected matrix (PAM) indicates the number of people affected until the contamination is detected.

3.2.2 Generation of an optimal subset of nodes

This section discusses the objectives for the sensor placement study and the objective function-based node selection strategy.

3.2.2.1 Definition of objective functions:

The objectives evaluated for the MOSPP study are: a) Detection likelihood, b) Expected detection time, and c) Population affected.

a) The detection likelihood (Z_1) is defined as the percentage of contamination events detected for a given design of sensors.

$$Z_1 = \frac{100}{N}\sum_{i=1}^{N} d_i \qquad i = 1,2,3....N \tag{3.1}$$

where i = 1,2,3…N refers to the 'N' contamination events simulated for the network and d_i = 1 if the contamination event 'i' was detected otherwise '0'.

b) The time of detection for a particular contamination event 'i', t_d, is the minimum time to detect among all sensors (j) present in the design,

$$t_{d,i} = min(t_{i,j}) \tag{3.2}$$

and, the Expected time of detection (Z_2) is the expected value of t_d over N contaminations given by,

$$Z_2 = E(t_{d,i}) \tag{3.3}$$

c) The expected population affected prior to detection (Z_3) is calculated by estimating the average amount of contaminated water consumed followed by probit factor-based estimation of the population affected. This is represented as,

$$M_i = \varphi \Delta t \sum_{k=1}^{N} c_{lk} \frac{q_{lk}}{\overline{q_l}} \tag{3.4}$$

where, φ = mean amount of water consumed by an individual (L/day/person), Δt = evaluation steps (days); c_{lk} = contamination concentration for node l and time step k (mg/L), q_{lk} = water demand for time step k and node 'l'; $\overline{q}$ = average water demand at node l; and N = number of evaluation time steps prior to detection. The probability that a person who ingests contaminant mass M_l will become infected was depicted by R_l estimated as,

$$R_l = \Phi\{\beta \log_{10}[(M_l / W) / D_{50}]\} \tag{3.5}$$

where β = probit slope parameter (unitless); W = assumed average body mass (kg/person); D50 = dose that would result in a 0.5 probability of becoming infected or symptomatic (mg/kg) and Φ = standard normal cumulative distribution function. The population affected (Pa) for a particular contaminant event is given as,

$$P_{a,i} = \sum_{l=1}^{V} R_l P_l \tag{3.6}$$

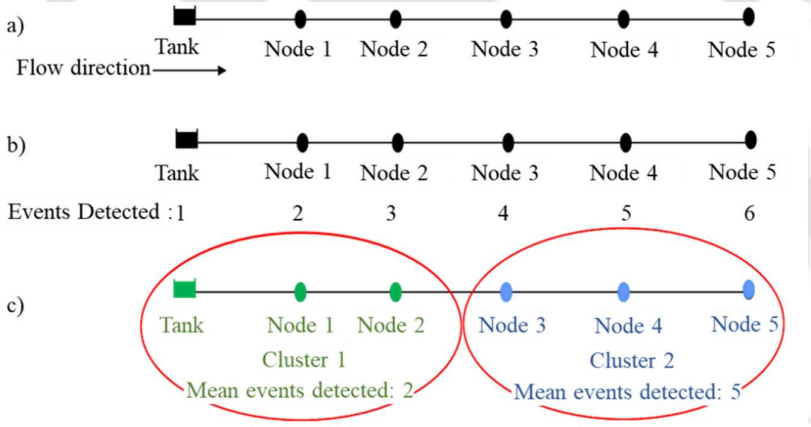
where P_l = population assigned to node 'l'; and V =total number of nodes. The objective function is the expected value of Pa computed over the assumed probability distribution of 'i' contamination events [53].

$$Z_3 = E(P_{a,i}) \tag{3.7}$$

The objectives are taken 'as is' from the BWSN challenge [40]. In the following paragraphs, the expected time of detection will be referred to as detection time and the expected population affected prior to detection as affected population for ease of discussion.

3.2.2.2 Generating optimal subset of nodes

The ideal case for complete observation of the WDN is to place sensors at all the nodes in the network, but it is not possible due to various constraints. The following best-case scenario is to design the sensor placement such that all the contamination events can be detected. However, a guarantee of 100% detection is inefficient if the time to detect is considerable or if a significant population gets affected before detection. It is necessary to assess the nodes based on swift detection of the contamination events. The objectives for improving the detection and detection time (or affected population) are competing in nature; thus, placing sensors for better detection will deteriorate the detection time.



d) **Pollution Matrix of Node 2 and Node 4 for 6 contamination scenarios**

Contamination scenarios →	1	2	3	4	5	6
Node 2	1	1	1	0	0	0
Node 4	1	1	1	1	1	0

Figure 3.3 Clustering and node selection for an example network a) Example network b) Events detected at each node c) Cluster formation based on event detection d) Pollution matrix example

Consider the simple example network with six nodes in a straight line, as shown in Figure 3.3a. The example is explained for the MOO study of detection likelihood vs detection time, but

it can also be extended to detection likelihood vs affected population. For placing only one sensor in the network, it can be observed that placing a sensor at node 5 can detect all the events, but the expected detection time will be maximum for this sensor placement design.

On the other hand, if the sensor is placed at the tank node, the expected detection time will be minimum, as the detection time is estimated based only on the detected events, and only those events will be detected that occur at the tank. Each objective necessitates a distinct set of nodes to achieve optimal value, leading to reduced value of other objectives due to their competing nature. However, suppose two sensors are to be placed in this simple network. In that case, the first sensor can be dedicated to maximizing the detection likelihood (Node 5), and the second sensor that results in minimum detection time in combination with Node 5 can be selected. Note that the number of combinations for placing two sensors in nodes, ${}^{6}C_{2} = 15$, is reduced to just 5 pairs of Node 5 and other nodes. This methodology can be extended to real WDNs by selecting all the nodes that improve the detections (predominantly dead-end nodes), followed by selecting the nodes that will improve the detection time of the previously selected nodes. The selection procedure thus has two phases. First, nodes will be selected based on event detections followed by detection time or affected population. However, real WDNs are much more complex in design than the above network in Figure 3.3a, and this will lead to the loss of important information when these nodes are considered for MOSPP. In real WDNs, multiple sources, such as reservoirs and overhead tanks, are installed for reliable water supply. The nodes that are connected to two sources might receive water from either of the sources. In cases where the nodes are situated between reservoirs and overhead tanks, a change in water flow direction is observed whenever the source of water supply changes. So, it must be noted that these nodes and the nodes downstream will only be contaminated by the current source of water supply. These issues of changing water flow direction can be overcome by segregating the network into groups or clusters based on the similarity in events detected and selecting the nodes from the group. The clustering and greedy selection procedures are further explained in the following paragraphs.

3.2.2.3 k-means clustering and node selection

The nodes of the WDN were segregated into clusters using k-means clustering based on the dissimilarity of events detected by the nodes rather than the time-variant contamination concentration. In k-means clustering, the data is clustered into ‘k’ groups based on the ‘distance metric’ between the ‘k’ centroids and the observation data points [72]. The difference in the

detection ability of the nodes is translated into the 'distance metric' using the event detection matrix (EDM). The EDM is a binary matrix (built of 0's and 1's); thus, the 'Hamming distance metric' was used to estimate the distance between two nodal observations. The Hamming distance is calculated based on the XOR operation and estimates the number of bits that are different in a given pair of strings. For example, the distance between 101011 and 110011 is provided by,

$$101011 \oplus 110011 = 011000;\ Hamming\ distance = 2$$

When this metric is applied to the detection matrix, nodes with a similar number of detections at the exact bit locations (events) will be grouped into one. In other words, this calculates the dissimilarity between the nodes in detecting a given set of contamination events. That is, the distance metric is not affected when both nodes detect or fail to detect the contamination. This clustering methodology prioritizes the transportation of contaminants to nodes during contamination events over connectivity or closeness of nodes, which is the norm considered in previous literature. The number of clusters can be estimated by clustering for different 'k' and then fixing 'k' based on statistical measures [83], or it can be fixed based on the number of sensors [78]. The former procedure requires multiple k-means clustering runs, while in the latter, only one node is chosen per cluster for sensor placement. These procedures result in higher computation time and the neglect of many potential locations. In this study, a heuristics-based approach is developed by assuming that all the nodes in the WDN are selected by choosing 'k' nodes from 'k' clusters for each objective considered for the MOO study. For two objectives in MOSPP, this results in the following equation

$$2k^2 = number\ of\ nodes \tag{3.8}$$

Upon rearranging the above, the number of clusters can be calculated as,

$$k = round\left(\sqrt{\frac{number\ of\ nodes}{2}}\right) \tag{3.9}$$

The number of clusters was estimated using(3.9), and from each cluster, the nodes were selected, as discussed below.

3.2.3 Selection of nodes

The node selection procedure is two-phased, and in each phase, the nodes are selected for each of the objectives in MOSPP. In order of importance, first, the nodes are chosen for event

detection, followed by the second competing objective, detection time (Z_2) or affected population (Z_3). The selection procedure relies on the pollution matrices to assess the importance of the nodes. In the first phase, nodes are selected from each cluster in the clustering module. The nodes belonging to each cluster are stored in sets G_i, where i=1,2..k, represents cluster 'i'. From each G_i set, nodes are selected based on the maximum number of events detected (Z_1) using the event detection matrix and stored in set $N_{G,i}$. The node selection for a

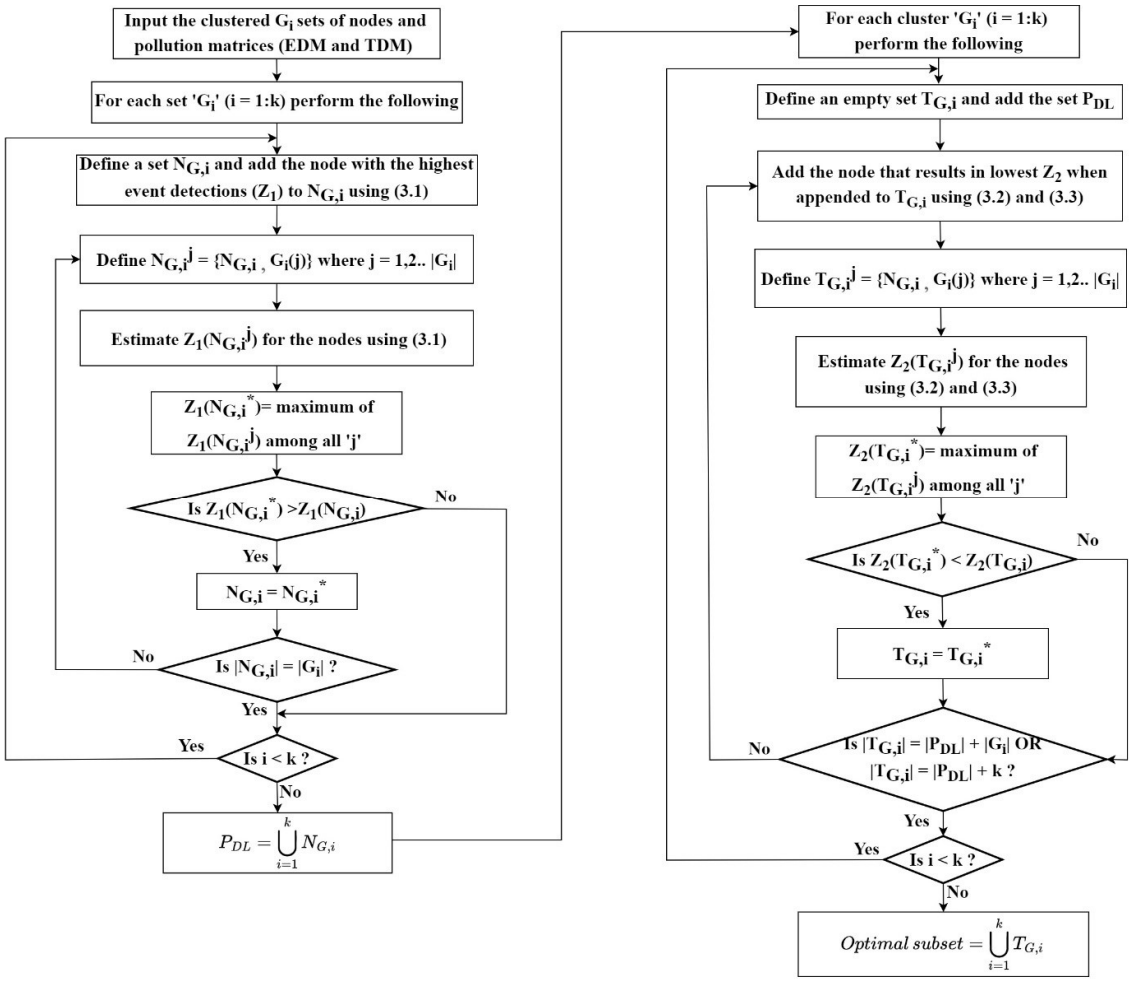


Figure 3.4 Objective based selection of nodes for multi-objective sensor placement study (Z_1 vs Z_2)

cluster is terminated when no further improvement in Z_1 is observed. In the example network (Figure 3.3a), Node 2 from cluster 1 will have the maximum Z_1 value, and no other nodes in the cluster can improve the Z_1 value. Similarly, only Node 5 from cluster 2 will be selected. Thus, for a given WDN, $N_{G,i}$, i = 1,2…k, are generated that comprise the nodes selected from

sets G_i, i = 1,2,…k, respectively. The nodes with higher sensitivity to detection likelihood (Z_1) are given by the union of the sets $N_{G,i}$, denoted as $P_{DL} = \coprod_{i=1}^{k} N_{G,i}$.

In the second phase, the nodes are selected based on the improvement observed in detection time. The nodes are selected from each set G_i, and their union will be defined as nodes with higher sensitivity to detection time. Firstly, sets $T_{G,i}$ are defined for each set G_i, and the nodes in P_{DL} are added to each $T_{G,i}$ set. The inclusion of set P_{DL} in the set $T_{G,i}$ before selecting nodes ensures that all the contamination events are detected, and the selected node will improve the overall detection time.

The detection times corresponding to the nodes were estimated using the time to detect matrix (TDM). Then from each set G_i nodes are appended one by one to set $T_{G,i}$ which results in the lowest value Z_2. The nodes are appended till no further decrease in the detection time is observed or a maximum of 'k' nodes are selected. The nodes in a WDN will provide the lowest detection time when they are the source of contamination and a sensor is placed at them. However, the trade-off in improving Z_2 for each new node selected is very low when the number of nodes in set $T_{G,i}$ is large. Thus, to control the number of selected nodes, the second terminating criterion mandates the choosing of a maximum of only 'k' nodes from each group. The set of nodes selected based on detection time is given by the union of set $T_{G,i}$ as $P_{DT} = \coprod_{i=1}^{k} T_{G,i}$.This is the optimal subset of nodes considered for MOO of Z_1 vs Z_2. Note that the nodes for improving detection likelihood have already been added to sets $T_{G,i}$ and thus do not require to be added separately. During the implementation of the above selection procedure, the nodes that were already part of $N_{G,i}$, P_{DL} and $T_{G,i}$ were removed from sets G_i to avoid repeated computations of already selected nodes. The cluster-wise greedy selection of nodes results in choosing nodes with lower detection time (which are generally located upstream of the cluster) as well as nodes with maximum observability (which are typically located downstream of the cluster). Thus, the combination of k-means clustering and greedy selection results in a set of selected nodes targeted towards both objectives. In contrast, the previous studies only focused on network connectivity, closeness to other nodes for trimming the network or just maximum observability at the expense of detection time. The flowchart for the node selection procedure is shown in Figure 3.4.

3.2.4 Multi-objective Optimization (MOO) and Contribution rate indicator

The pattern search algorithm from MATLAB2020b was implemented to find the Pareto front. This algorithm uses a derivative-free methodology by extending direct search methods to multi-objective problems, and is thus termed as direct multi-search (DMS) methodology [101]. Similar to the direct search techniques for single objective optimization, first, an initial feasible solution and step size of the algorithm are defined. This initial solution is added to the list of non-dominated solutions. Then, new solutions are generated using the initial solution and step size in a manner similar to the direct search methods. The objective function values are estimated for the new solutions, and non-dominated sorting is performed along with the solutions present in the non-dominated solutions list. If a new non-dominated solution is obtained, then the previous step of generating new solutions is continued. It must be noted that once a solution is utilized to generate new solutions, it is moved to the end of the list. Thus ensuring that only newly obtained non-dominating solutions are used to generate new solutions. If an iteration does not yield any new non-dominated point, then the step size is reduced to half, and the process is repeated. The stopping criteria for DMS are a maximum number of iterations and a minimum value for the step size.

The traditional method would evaluate the objective function for a maximum of 'Nruns x Np x T' times. Meanwhile, the objective functions would be first evaluated for selecting nodes in the node selection procedure. Therefore, to provide a fair comparison between the traditional and proposed method, the number of objective function evaluations during the selection procedure is reduced from the maximum number of function evaluations for the proposed method.

Pareto fronts of the traditional and proposed methods were compared in terms of contribution rate (CR) indicator based on hypervolume, developed by Cao et al. [102]. The Pareto fronts (PF_i) to be compared are first compiled to generate a surrogate true Pareto Front (PF_s). Then, the contribution from each of the PF_i's in PF_s is estimated using hypervolume. Hypervolume measures the size of the space enclosed by all points on the Pareto front and a user-defined reference point, and it is indicated as $I_H(PF,r)$ for a given Pareto front (PF) and a reference point (r). The reference point for hypervolume calculation was estimated based on [102], where the reference point is recommended to be within $\left[\frac{\sqrt{2}}{|PF|-1}, \sqrt{2}\right]$ distance from the nadir point after normalizing the Pareto Front. A nadir point is derived from the extreme points of a Pareto front, which indirectly describes the range of the non-dominated solution set. After evaluating the

reference point, the CR indicator of Pareto Fronts from each method was calculated. The contribution rate (CR) associated with each of the Pareto Front is estimated as follows:

$$CR_i = \frac{I_H\left(PF'_i, r\right)}{I_H\left(PF_s, r\right)} \tag{3.10}$$

where PF_i' is the set of non-dominating points of Pareto Front 'i' present in the surrogate true Pareto PF_s, 'r' refers to the reference point and 'I_H' refers to hypervolume.

The procedure for estimating the Contribution rate indicator for each Pareto Front is as follows:

i. Combine all the Pareto fronts, PF_1, PF_2, ..., PF_n, according to the definition of Pareto domination, into a single set of non-dominated points, denoted by PF_s. PF_s is regarded as a surrogate for the true Pareto front.
ii. Obtain approximation Pareto fronts PF_1', PF_2', …, PF'_n such that PF'_i only contains those criterion vectors from PF_i which are not collectively dominated by PF_s.
iii. Calculate the hypervolume indicator (I_H) for each of the approximation Pareto fronts (PF_i') and PF_s using the reference point estimated for PF_s.
iv. Compute the *contribution rate* (CR) associated with each approximation using (3.10).

The value of CR_i close to '1' means that PF_i is closest to replicating the Surrogate true Pareto front. Compared to other PFs, a higher CR indicator translates into higher hypervolume dominance.

3.3 Results and Discussion

3.3.1 BWSN Network1

The contamination event simulations for BWSN Network1 yielded three pollution matrices of dimensions 129x6192, as 48 contamination events were simulated for each of the 129 nodes. It was observed that 622 events remained undetected due to zero flow conditions at the nodes during contaminant intrusion. The nodes of BWSN Network1 were clustered into 8 groups based on (3.9) and 21 nodes were selected based on EDM, 51 based on TDM and 48 based on the PAM. Most dead ends have been selected (except Junction 13 and Junction 36, due to zero flow), ensuring that all the 'observable events' will be detected. The 8 clusters with the average number of detections are depicted in Figure 3.5a. The average number of detections for clusters ranged from 203.4 (3.6% of observable events) at the network entrance nodes and low flow

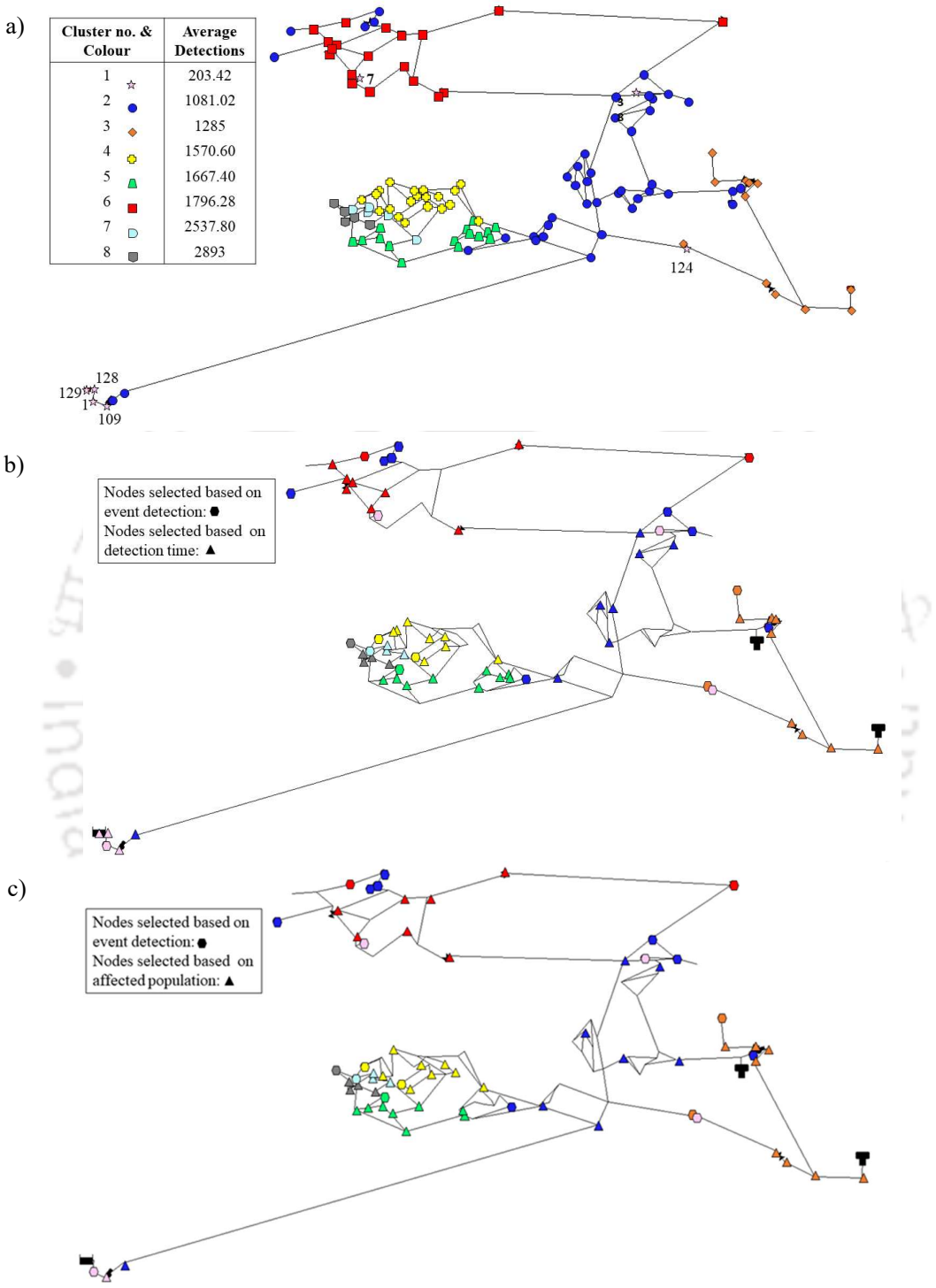


Figure 3.5 a) Clusters formed for BWSN Network1 b) Nodes selected for Z_1 Vs Z_2 MOSPP study c) Nodes selected for Z_1 vs Z_3 MOSPP study

nodes to 2893.0 (51.94% of observable events) at the downstream nodes. The 72 nodes selected for detection likelihood vs detection time and 69 nodes selected for detection likelihood vs affected population are shown in Figure 3.5b and Figure 3.5c, respectively. The first cluster had the lowest mean detections and consisted of Junctions 1, 7, 38, 109, 124, 128, and reservoir node (129). Junctions 1, 7, 38, and 124 were selected during the first phase of selection, leaving junctions 109, 128, and 129 as no improvement in detections was observed when appending these nodes.In the second phase of selection, all three nodes improved detection time and were thus added to the optimal subset of nodes. In contrast, only Junction 109 was selected for the affected population. It must be noted that all three nodes have zero base demand, but as the algorithm chooses at least one node from each group, the node at the downstream was selected. Out of 48 nodes selected in the affected population, 13 had zero base demands. These nodes were selected either due to their position between two nodes with non-zero base demands (like Junction 91) or due to a condition that makes the algorithm choose at least one node from each cluster, as observed for Junction 109. It can be observed that five nodes in cluster 2 are disjointed from cluster 2 and are separated by cluster 6. This is because, though the disjointed nodes are closer to cluster 6, the events detected by these nodes are more similar to the events detected by nodes of cluster 2. This is verified by the average number of events detected by the disjointed nodes =1181.8, which is closer to the average number of events detected in cluster 2 (=1081.02) than in cluster 6 (=1796.28).

The matching rate is defined as the percentage of nodes from BWSN results appearing in the pre-selected nodes, and it is provided in Table 3.2. The sensor locations (nodes) in the BWSN challenge were given by 14 research groups using different methods to optimize these locations. The efficiency of the proposed method is tested by comparing the optimal subset of nodes with the possible sensor nodes provided in BWSN. The matching rate of the BWSN 5 SPP solution nodes with the optimal subset of nodes (set OSN) was 78.8% and 75.6% for Z_1 vs Z_2 and Z_1 vs Z_3 study, respectively. In comparison, only a 26.9% matching rate was observed in the controllability analysis method (set CAS) described in [81]. For BWSN 20 SPP, 64.71% and 67.06% of the BWSN solution nodes matched with set OSN compared to 57.4% for set CAS. The lower matching rate of set CAS might be due to the absence of the cumulative effect of selected nodes in improvising the detection time. Meanwhile, the proposed methodology weighs the nodes in terms of their significance in detecting events and reducing detection time or exposed population. The nodes of set OSN and BWSN results are enumerated in Table 3.2 below.

Table 3.2 Nodes selected and matching rate in various strategies

MOO Study	**Sensor nodes**	**Matching rate %** [81]	**Matching rate (This study) Z_1 Vs Z_2**	**Matching rate (This study) Z_1 Vs Z_3**
5 sensors (B5) Total: 33	1, 10, 17, 20, 21, 22, 29, 30, 31, 34, 45, 46, 58, 68, 71, 79, 81, 82, 83, 84, 97, 98, 100, 101, 102, 109, 112, 116, 117, 118, 122, 123, 126	23.9	78.8	75.76
20 sensors (B20) Total: 85	0, 1, 3, 4, 5, 7, 8, 10, 11, 12, 14, 17, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 31, 32, 34, 35, 37, 38, 39, 40, 42, 45, 46, 49, 51, 52, 53, 58, 61, 63, 64, 65, 68, 69, 70, 71, 72, 73, 74, 75, 76, 78, 79, 80, 81, 82, 83, 85, 88, 89, 90, 93, 95, 96, 97, 98, 99, 100, 102, 103, 106, 109, 111, 112, 114, 116, 117, 118, 122, 123, 124, 125, 126, 128	57.4	64.71	67.06
Nodes for Z_1 vs Z_2 (SO) Total: 72	1, 3, 4, 5, 7, 10, 11, 12, 14, 16, 19, 20, 21, 31, 32, 34, 35, 38, 40, 41, 45, 46, 49, 51, 59, 61, 62, 64, 67, 68, 70, 71, 72, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 90, 93, 98, 99, 100, 101, 102, 103, 104, 107, 109, 112, 113, 114, 116, 118, 119, 120, 121, 122, 123, 124, 126, 127, 128	-	-	-
Nodes for Z_1 vs Z_3 Total: 69	0, 1, 3, 5, 7, 9, 10, 11, 16, 17, 19, 20, 24, 28, 30, 31, 32, 34, 35, 38, 39, 45, 46, 51, 59, 65, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 88, 90, 94, 95, 96, 98, 99, 100, 102, 103, 104, 107, 108, 109, 113, 114, 115, 117, 118, 120, 121, 122, 123, 124, 126	-	-	-

The percentage of events detected with respect to mean detection time for set OSN and set CAS [81] are shown in Figure 3.6. About 66.16 % of events were detected by OSN, while CAS detected only 41.36% within 5 minutes of contamination. Also, the set CAS observed a maximum of 94.5% of events, while OSN observed 100% of the events. This shows that the selected nodes have a quicker reaction time to the contamination events than DN nodes. For comparison purposes, the sensor solutions from the BWSN 5 sensor (set B5) and the BWSN 20 sensor (set B20) were also plotted). The order of improvement of event detection percentage with respect to detection time was B5, CAS, OSN, and finally, B20 sets. The set OSN was swift in detecting events than the sets B5 and CAS but lagged behind set B20. It must be noted that only the combination of nodes from the set OSN is required to place the sensor. Thus, multi-objective optimization was performed to verify the efficiency of selected nodes.

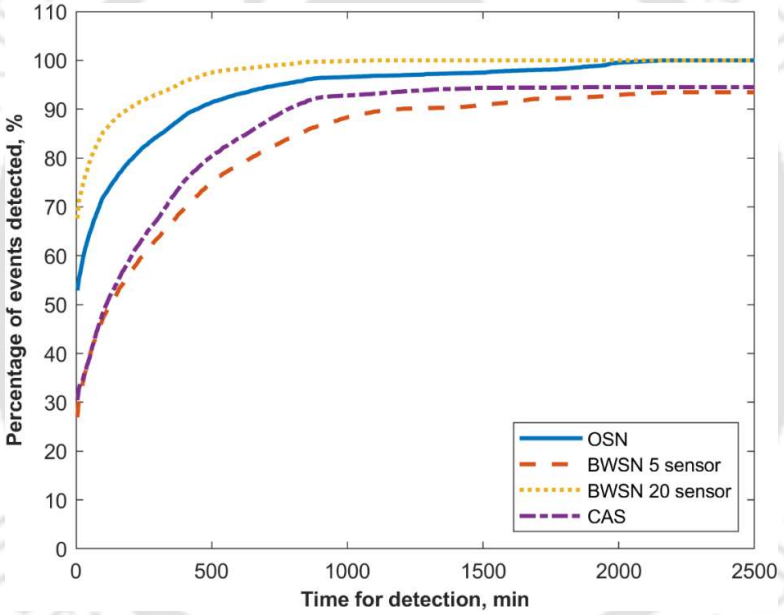


Figure 3.6 Percentage of events detected with respect to mean detection time for BWSN Network1

The parameters for the multi-objective optimization study were: number of runs=30, maximum function count =50000, tolerance = $1*10^{-6}$, final Pareto Front = Compiled non-dominating set of all 30 runs. The decision variable bounds for the complete set of nodes will be [1 N_{nodes}], and for the subset of nodes, it will be [1 $|\coprod_{i=1}^{k} T_{G,i}|$]. The 'paretosearch' tool of MATLAB2020b, which incorporates a pattern search technique, was used to generate Pareto fronts. The legend sensor locations while the legend 'Proposed' refers to the method developed in this study where only the optimal subset of nodes (OSN) are considered for placing sensors.

Figure 3.7 represents the process to estimate the surrogate true Pareto front for placing 20 sensors for objectives Z_1 and Z_2. Figure 3.7a depicts the complete set of Pareto points obtained from 30 runs of the optimization. A clear separation of the Pareto points between traditional and proposed procedures was observed, with the latter dominating the former in the 50% to 90% detection likelihood range.

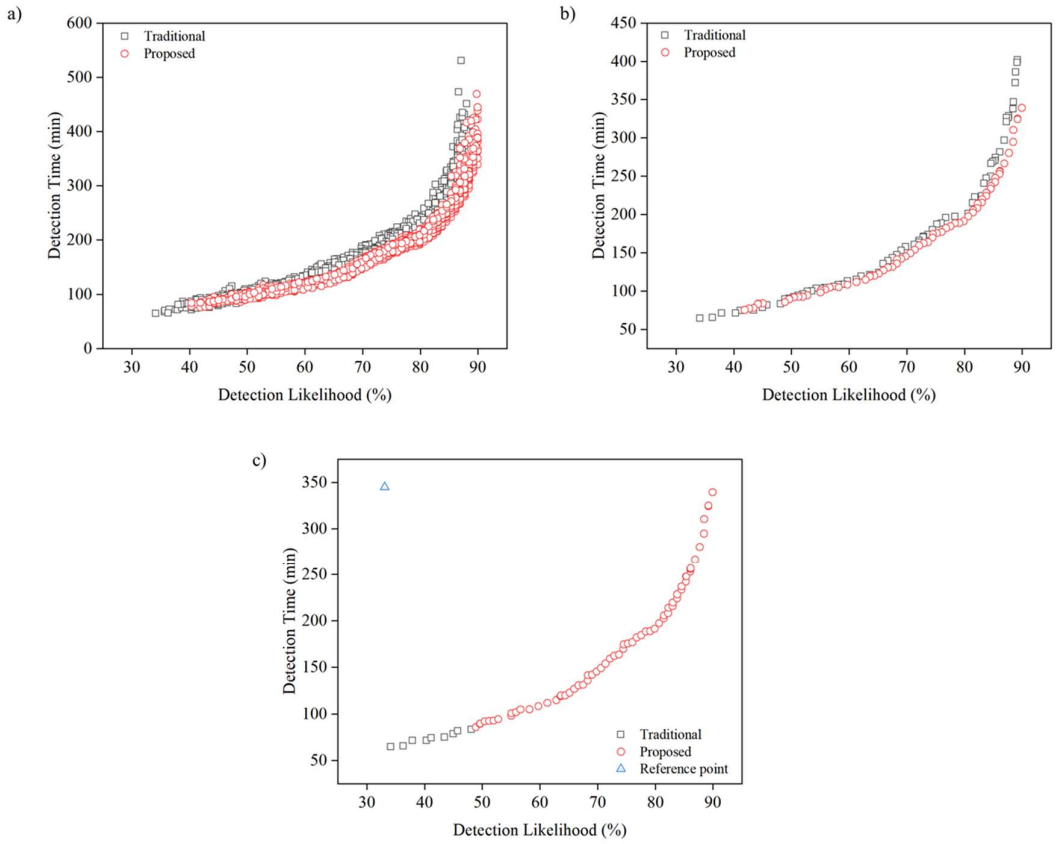


Figure 3.7 Generation of surrogate true Pareto front for BWSN Network1 - 20 sensors problem a) Pareto fronts of all 30 runs b) Compiled final Pareto c) Surrogate true Pareto front with reference point

This effect is reiterated when Pareto points from 30 runs are compiled and compared in Figure 3.7b. Based on these Pareto fronts, the surrogate true Pareto front and reference point were obtained and depicted in Figure 3.7c. The Surrogate true Pareto fronts with the reference points for 5 and 20 sensors placement for objectives Z_1 vs Z_2 and Z_1 vs Z_3 study are shown in Figure 3.8. An overview of Pareto points for the Z_1 vs Z_2 study revealed that the OSN performs better than the complete set. In the Z_1 vs Z_3 study, for placing 5 sensors, the optimal subset generated fewer Pareto points than the traditional method, and most of these Pareto points had higher

detection likelihood values. This could be due to the exclusion of nodes with very low base demand that were not selected.

This trend was again pronounced in 20 sensor results shown in Figure 3.8d, where all the Pareto points at Z_1<65% were obtained from the traditional method and at Z_1 > 75% were from the proposed method. The comparison of both approaches in terms of CR indicator is shown in Table 3.3.

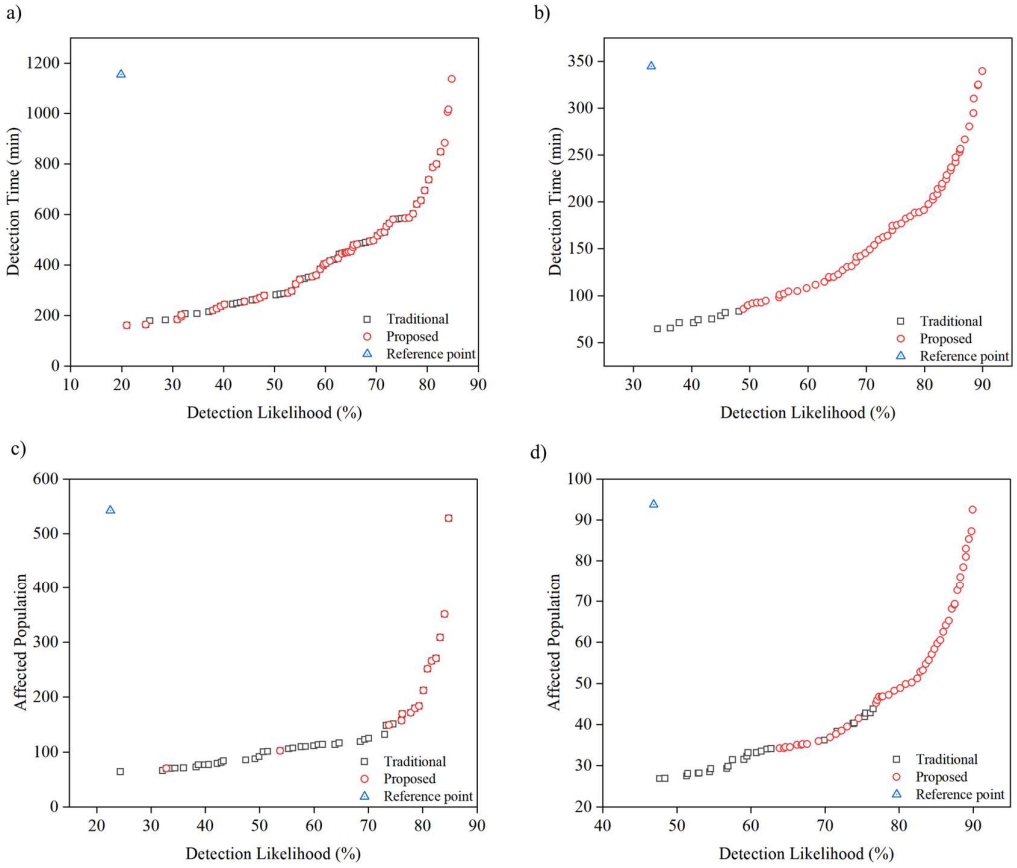


Figure 3.8 Surrogate true Pareto fronts and Reference point for BWSN Network1 a) Z_1 vs Z_2 5 sensors b) Z_1 vs Z_2 20 sensors c) Z_1 vs Z_3 5 sensors d) Z_1 vs Z_3 20 sensors

The CR indicator of both methods for 5SPP was the same in the Z_1 vs Z_2 study, indicating that both procedures dominate a significant portion of hypervolume at the same level. Whereas in the Z_1 vs Z_3 study, the traditional method was higher by 4.16%. The CR indicator for 20 SPP was 190% and 22.7% more than the traditional method in Z_1 vs Z_2 and Z_1 vs Z_3 study, respectively. This indicated that the optimal subset provided better Pareto points in terms of hypervolume dominance. Since the solutions with lower detection likelihood cannot be considered for placing sensors, a cut-off criterion was maintained to assess the number of

Pareto points contributed by both methods. Only those solutions whose Z_1 value was more than 50% of maximum observed events (50% of max(Z_1)) were considered as potential solutions. The number of these Pareto points observed for 5 sensors was higher for the traditional method, while for 20 sensors, the proposed method provided a higher number of solutions. The time taken for estimating the final Pareto Front using the proposed method for 5SPP was 25% (by average) less than the traditional method, while it was less by only 0.7% (by average) for 20SPP.

Table 3.3 BWSN Network comparison of the traditional and proposed method

	Detection likelihood Vs Detection time				**Detection likelihood Vs Affected population**			
	5 Sensors		**20 Sensors**		**5 sensors**		**20 sensors**	
Parameters	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**
Number of nodes	129	72	129	72	129	69	129	69
Contribution rate Indicator	0.99	0.99	0.21	0.60	1	0.96	0.44	0.54
Pareto points in PFs (Z_1>50%)	59	47	3	65	36	14	26	46
Time taken for Final Pareto, sec	433.10	318.84	5426.26	5415.68	679.49	418.45	6402.20	6324.27

3.3.2 Comparison with BWSN Results

The results from the BWSN competition were compared with the Pareto fronts obtained from the current MOSPP study. Before comparison, the contamination intrusion strategy was analysed by plotting the objective values from BWSN [53] and superimposing the objective function values for the sensor nodes solutions from BWSN [53] on the pollution matrix developed in this study. The variation in these two was less, which asserted that the pollution matrix developed in this study was close to the contaminant intrusion strategy considered in the BWSN. The results of BWSN provided in Figure 3.9 depict one solution provided by each of the participating groups. Except for 4 points (marked by green box), all other points are dominated by Pareto fronts of both traditional and proportional methods.

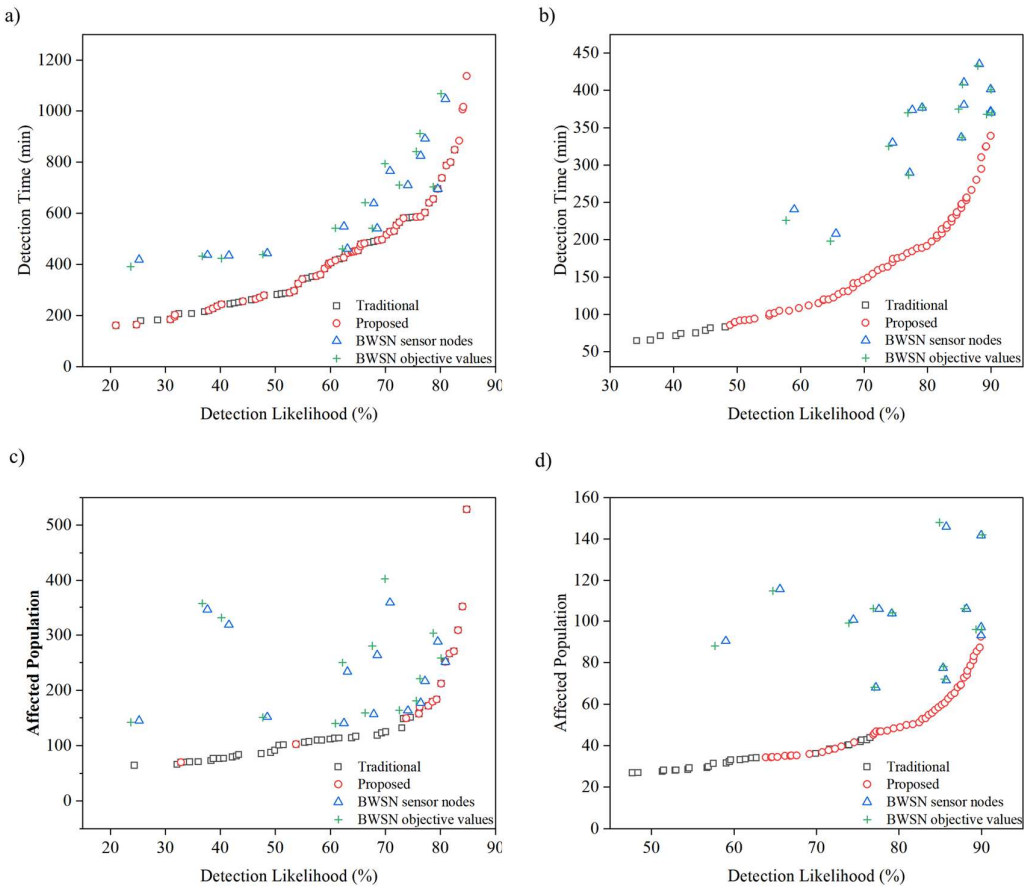


Figure 3.9 Comparison of solution with BWSN results a) Z_1 vs Z_2 5 sensors b) Z_1 vs Z_2 20 sensors c) Z_1 vs Z_3 5 sensors d) Z_1 vs Z_3 20 sensors

3.3.3 C-town Network

The pollution matrices for the C-town network had 38016 contamination events for 396 nodes, yielding a pollution matrix of dimension 396x38016. Three nodes in the network did not detect any event, and their respective contamination events were undetected due to zero flow. The C-town network was segregated into 14 groups: 263 nodes were selected for the Z_1 vs Z_2 study, and 261 nodes were selected for the Z_1 vs Z_3 study. The surrogate true Pareto front and reference point for CR indicator estimation for the C-town network are displayed in Figure 3.10. For 5SPP, comparable Pareto points are generated by both methods for the MOSPP studies. In the case of Z_1 vs Z_3, the proposed methodology generated better Pareto points for higher detection likelihood, similar to the results observed for BWSN Network1. However, for 20SPP, most of the Pareto points in PFs were generated using the proposed method. These observations have been reflected in the CR value of the corresponding problem provided in Table 3.4. The set OSN yielded 96.6% and 74.6% of Pareto points in PF_s for Z_1 vs Z_2 and Z_1

vs Z_3 respectively for 20SPP. The time taken for estimating the final Pareto Front using the proposed method for Z_1 vs Z_2 study was 3.85% (by average) less than the traditional method, whereas it was less by 9.75% (by average) for Z_1 vs Z_2 study.

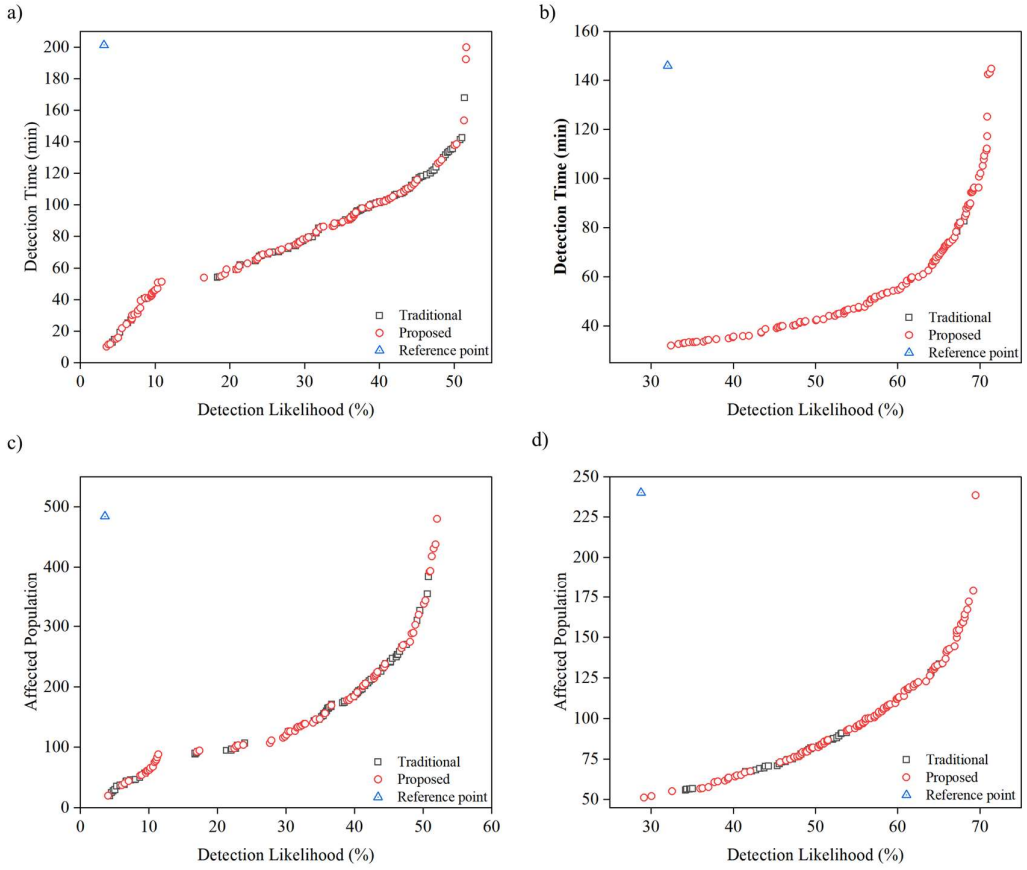


Figure 3.10 Surrogate true Pareto fronts and Reference point for C-Town Network a) Z_1 vs Z_2 5 sensors b) Z_1 vs Z_2 20 sensors c) Z_1 vs Z_3 5 sensors d) Z_1 vs Z_3 20 sensors

Table 3.4 C-Town Network comparison of the traditional and proposed method

	Detection likelihood Vs Detection Time				**Detection likelihood Vs Affected Population**			
	5 Sensors		**20 Sensors**		**5 sensors**		**20 sensors**	
Parameters	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**	**Traditional**	**Proposed**
Number of nodes	396	263	396	263	396	261	396	261
Contribution rate (CR) Indicator	0.615	0.617	0.543	0.650	0.691	0.697	0.576	0.628
Pareto points in PF_s (Z_1 > 50% Z_1)	77	55	4	117	50	49	36	106
Time taken for Final Pareto, sec	3012.18	3185.59	3799.45	3877.15	3678.44	3167.42	3629.90	3425.35

From the results of both the methods, it was observed that the pre-selection procedure performs well when the number of combinations is huge, as in 20SPP ($^{129}C_5 << {}^{129}C_{20}$ and $^{396}C_5 << {}^{396}C_{20}$). However, the deterioration in performance was only observed for the lower detection likelihood solutions, which are generally omitted.

3.3.4 Comparison of sensor placement solutions based on detection likelihood cut-off

The expected time for detection for various percentages of detections is described in Table 3.5. For BWSN 5SPP detection, the maximum Z_1 value from the traditional method was 84.01% and 84.78% for the traditional method and the proposed method, respectively and for 20 sensors, it was 89.21% and 89.95%. On the other hand, the detection time at the maximum Z_1 values was lower in the proposed method. But to provide a viable comparison, Z_2 and Z_3 are compared at values at four fixed values of Z_1, starting at a minimum level of detecting 50% of

Table 3.5 Comparison of objectives Z_2 and Z_3 at different Z_1 cut-off values

Detection likelihood (Z_1), %	Detection Time (Z_2), min				Affected Population (Z_3)			
	BWSN 5 Sensors		BWSN 20 Sensors		BWSN 5 Sensors		BWSN 20 Sensors	
	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed
50%	281.96	289.23	92.10	91.86	100.61	102.46	27.49	32.08
60%	406.31	406.31	114.91	111.70	113.1	118.61	33.18	34.03
70%	515.91	515.91	160.38	149.25	124.97	133.79	37.03	36.88
80%	737.38	737.38	201.02	197.38	212.07	212.07	50.32	48.80

the events and then increasing in the order of 10%. It was observed that there is no significant difference between the traditional and proposed method for 5 SPP, while for 20 SPP, the expected time of detection reduces by 2.8%, 6.94% and 1.8% for 60%, 70%, and 80% detections, respectively. The affected population metric was higher in the proposed method for 5 sensor design and lower for 20 sensor design at 60%, 70% and 80% detection likelihood. The lowest Z_1 value Pareto point for 20SPP was 58% in the proposed method; thus, at a cut-off of 50%, the affected population has an 18% higher value.

As shown in Figure 3.11, the sensor locations for 70% detection for Z_1 vs Z_3 study for 5SPP have 3 locations in common, and one node for each solution had zero base demand. In contrast, 12 sensor locations were similar for 20 SPP, with 5 and 7 zero demand nodes for traditional

and proposed solutions, respectively. The sensor locations of 5 SPP are not repeated in 20 SPP, which indicates that solutions of a smaller number of sensors cannot be replicated for a larger number of sensors. For each sensor design, optimization is required to estimate the best location. Similarly, Z_2 and Z_3 are compared at Z_1 values of 40%, 50%, 60% and 65%. This range of Z_1 was fixed owing to the maximum Z_1 value of 53% and 71% observed for placing 5 and 20 sensors. The expected time for detection for both 5 SPP and 20 SPP from the proposed

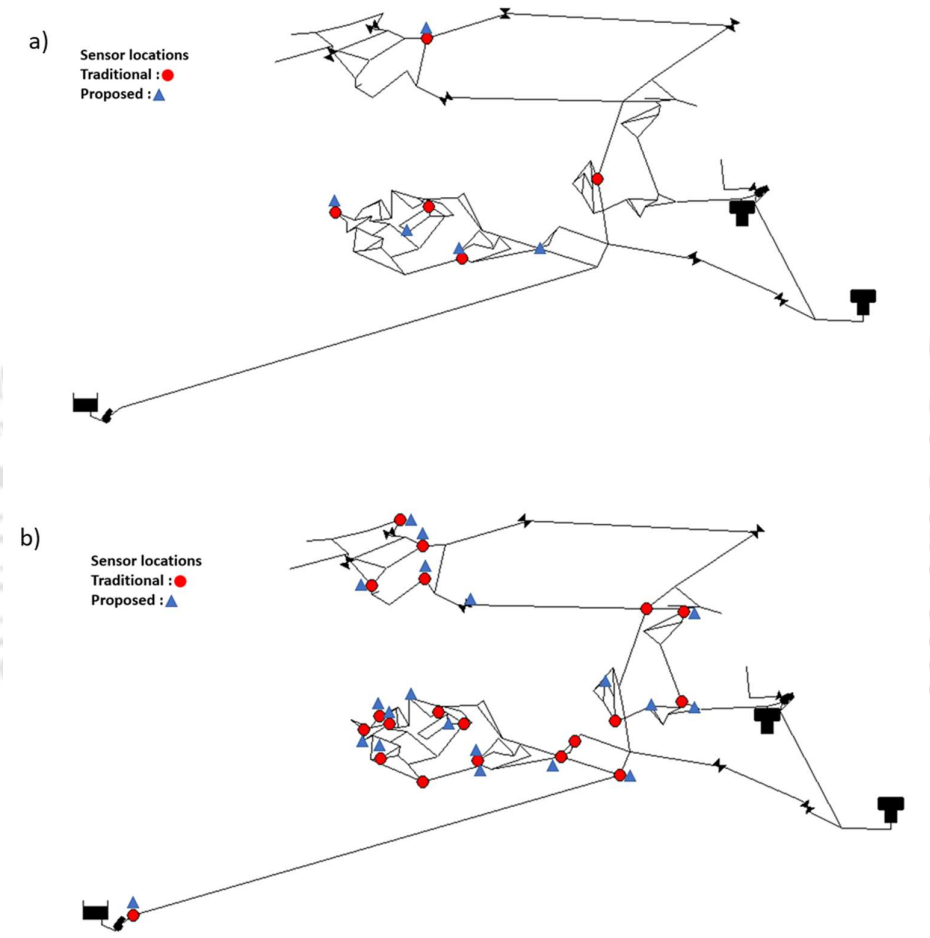


Figure 3.11 Sensor node location for 5 and 20 sensors in BWSN Network from Z_1 vs Z_3 study at Z_1 = 70%

method is comparable to values observed in the traditional method Table 3.6. All the objective values obtained from the proposed method for objectives Z_2 and Z_3 (except at Z_3 at 50% detection likelihood, where it is almost equal) are lower than the traditional method.

Table 3.6 Comparison of objectives Z_2 and Z_3 at different Z_1 cut-off values for C-town Network

Detection Likelihood (Z_1)	Detection Time (Z_2)				Affected Population (Z_3)			
	C-town 5 Sensors		C-town 20 Sensors		C-town 5 Sensors		C-town 20 Sensors	
	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed	Traditional	Proposed
40%	101.37	101.91	38.40	35.64	187.43	188.71	66.35	64.01
50%	137.81	137.81	44.68	42.22	342.15	337.30	82.76	82.30
60%	-	-	57.00	54.49	-	-	113.78	112.86
65%	-	-	116.60	102.03	-	-	133.57	133.97

The sensor nodes at 50% detections are displayed in Figure 3.12, where 3 out of 5 sensor locations are at the same nodes for 5 sensors (Figure 3.12a), and in the case of 20 sensors, only 6 nodes are common in solutions provided by both methods (Figure 3.12b). The sensor nodes from the proposed method are evenly distributed in the WDN, such that the nodes are placed at the farthest ends to obtain maximum detections. The set of nodes obtained from the proposed (Figure 3.12) is also present in the decision space of the traditional method. Still, due to the

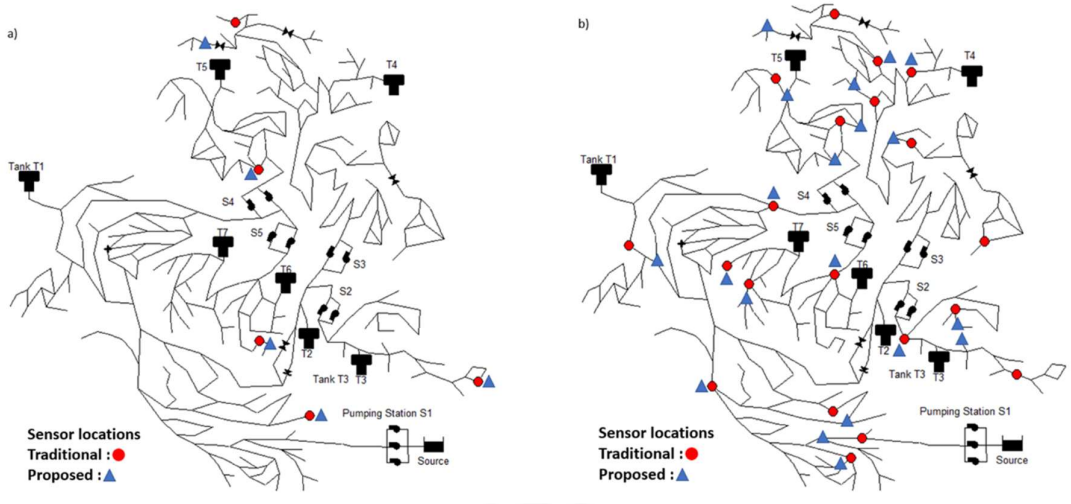


Figure 3.12 C town Sensor node location for 5 and 20 sensors for Z_1 vs Z_3 study at Z_1 = 50%

combinatorial nature of the problem, that particular solution was not obtained using the traditional method during the optimization process. The focused optimal sub-space developed in this study made it easy for the optimizer to determine a better combination of significant nodes, as observed in the results of the proposed method.

The repeatability of the pre-selection strategy was tested by running the selection module 100 times and estimating the percentage of the nodes from the final non-dominating Pareto Fronts selected in all 100 runs. On average, 80% of the nodes from Pareto Fronts were selected in all the runs, while 90.5% of nodes were selected in at least 85 of the runs (Table 3.7). The average time for finding the optimal subset for BWSN was 3.01 secs and for C-town network 156.57 secs.

Table 3.7 Repeatability test results

MOSPP Study (Number of runs)	**BWSN 5 SPP**	**BWSN 20 SPP**	**C-town 5 SPP**	**C-town 20 SPP**
Z_1 vs Z_2 (100)	88.9	86.2	73.7	83.6
Z_1 vs Z_2 (>85)	100	96.5	80.0	90.0
Z_1 vs Z_3 (100)	82.1	82.6	70.8	78.0
Z_1 vs Z_3 (>85)	100	94	80.2	84.0

3.4 Summary

In this chapter, the reduction of decision variable space for the sensor placement problem is investigated. An algorithm is developed to select nodes based on k-means clustering and greedy search techniques. It is tested on two networks BWSN Network1 and c-town network for the objectives, detection likelihood, detection time and population affected prior to detection. First, the algorithm's performance is analyzed by comparing it with benchmark results. Then, it is compared with results obtained when all the nodes are considered in the decision variable space in terms of the number of Pareto solutions and hypervolume contribution rate indicator. Finally, the sensor placement design solutions are presented for discussion.

Chapter 4: Optimal placement of WSNs in data deficient WDNs

4.1 Foreword

Chapter 3 discusses the analysis of the sensor placement problem for a Water Distribution Network (WDN) without considering the advantages and limitations of wireless sensor networks (WSNs) and the issue of data deficiency in the WDN. To address these concerns, this chapter presents efficient algorithms for the optimal placement of WSNs in data-deficient WDNs. These algorithms aim to maximize network coverage and provide redundancy in sensor detection. The main contributions of this study are summarized as follows:

i. Formulation of the model to evaluate the objectives, Network coverage (NC) and Sensor detection redundancy (SDR) and incorporating the sensor connectivity constraint simply and efficiently.
ii. Development of greedy search algorithms to generate sensor solutions that can be incorporated into the population of multi-objective optimization algorithms.
iii. Testing of the proposed method with different use cases focusing on various scenarios that might be observed in the placement of WSN in WDN.

The proposed study is tested on the newly commissioned South Guwahati water distribution system in Assam, India. It is assumed that the data transmission between nodes occurs through the multi-hopping mechanism. The greedy algorithms proposed are tested by incorporating them in the initial population of NSGA-II and compared to scenarios where the solutions of greedy search are not integrated.

4.2 Methodology – WSN for WDN

Similar to the previous chapter, in this study, N_s nodes are chosen from N_{nodes} to achieve maximum WDN monitoring using contamination event simulations. That is, the decision variables are positive integers representing the nodes of WDN and have values in the range [1 N_{node}]. Also, the objective function values are estimated over a series of contamination events. However, there are two primary modifications in this study. First, the WDN considered in this study lacks flow data. Correspondingly, the objectives have been changed, i.e. network coverage and sensor data redundancy, to achieve a similar distribution of sensors in WDN as observed in the previous chapter. Second, the sensors transmit the measurements via wireless communication. Each sensor has a transmission range, and it can send or receive data from other sensors within the communication range. It is necessary that they form a network for a seamless transmission of data. Overall, the WSN placement problem is defined as choosing N_s

nodes from the N_{node} of a WDN, such that the sensors placed in these nodes form a wireless network and yield maximum network coverage and sensor data redundancy.

4.2.1 Specification WDN used for case study

The proposed methodology for WSN design for WDN is tested on the newly commissioned South Guwahati water distribution network (SGWDN). The SGWDN has 326 nodes with one overhead tank reservoir (OTR) supplying water to all the nodes. The hydraulic and quality simulations were initiated by fixing demand at all the dead-end nodes of the network and then serving these demands from the OTR. Then, in EPANET simulation, the contaminant was introduced at each node. The nodes that showed a spike in chemical concentration were noted in the contamination propagation matrix (CPM), as described in Hart et al. [53]. The simulations were carried out in EPANET2.0 and MATLAB2019a academic version. 326 contamination simulations were performed corresponding to the 326 nodes of the WDN, and the results were stored in a CPM of size 326x326. Each row represented the nodes of the network, whereas the columns represented the node where the contaminant was introduced. An entry of '0' was made if there was no spike in contamination level, which means the contaminant never reached the particular node. Otherwise, it is fixed to 1. In this binary matrix, a cell represents whether the node 'row' detects the contamination event 'column'. Furthermore, as each column corresponds to a contamination event at one of the nodes, detecting a contamination event, in turn, means that the row node covers the corresponding column node. The goal is to find the set of nodes (rows) that cover the maximum number of nodes (columns). For network coverage, nodes have to be covered by at least one of the sensor nodes, while for sensor detection redundancy, the nodes have to be covered by at least two sensor nodes. The CPM is a reduced form of pollution matrix developed in the previous chapter. In the pollution matrix, the temporal flows of WDN are incorporated, while in CPM, only steady-state flow is considered.

In this study, it is assumed that the wireless water quality sensors transmit data through a multi-hop strategy. To realize the advantages of the multi-hop strategy, the placement of only node stations and a base station is considered. Also, we assume that the transmission range (TR) of all the sensors is equal, as observed in many studies [97]. The aim of this chapter is to generate a wireless water quality sensor placement solution that maximizes the above-discussed objectives while regarding the defined WSN architecture for seamless data transmission.

4.2.2 Stochastic optimization: Objective functions and constraints

Network coverage is defined by the percentage of nodes covered by the given sensor design. For the current system, network coverage is estimated as the percentage of contamination events detected by a given sensor placement design. It is calculated as,

$$NC = \frac{100}{N}\sum_{i=1}^{N} d_i \qquad i = 1,2,3\ldots..N \tag{4.1}$$

where i = 1,2,3….N refers to the 'N' contamination events or 'N_{nodes}' nodes, and d_i = 1 if the design detects event 'i', else 0. Similarly, the sensor detection redundancy (SDR) of a sensor placement design is the percentage of events detected by more than one sensor. This is estimated by the equation,

$$SDR = \frac{100}{N}\sum_{i=1}^{N} R_i \qquad i = 1,2,3\ldots..N \tag{4.2}$$

where R_i = 1 if an event gets detected more than once, else R_i=0 [68]. The value of d_i and R_i can be easily obtained by column-wise addition of CPM and checking if the value is greater than or equal to 1 (d_i) or greater than 1 (R_i), respectively.

The sensor connectivity constraint demands that sensor nodes be within the transmission range. The Euclidean distance between any two nodes is estimated using the x-coordinates and y-coordinates of the nodes (from the EPANET .inp file), given by,

$$S_{mn} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \tag{4.3}$$

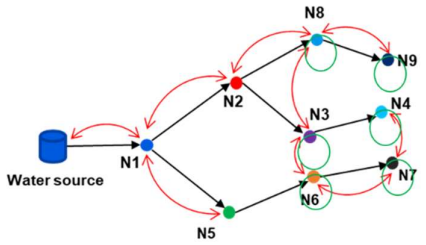
where S_{mn} is the distance between the node 'm' and node 'n' and, m and n = 1,2…N_s are the nodes selected for sensor placement. Based on the internodal distance, a connectivity matrix (M) is constructed to ensure connectivity between the nodes. The size of matrix 'M' is N_s x N_s. The matrix is filled with '0's and 1's according to the below set of equations,

$$M_{mn} = \begin{cases} 0 & if\ m = n \qquad \forall m,n \\ 1 & if\ S_{mn} \le Transmission\ range \quad \forall m,n\ and\ m \ne n \\ 0 & if\ S_{mn} > Transmission\ range \quad \forall m,n\ and\ m \ne n \end{cases} \tag{4.4}$$

A sensor cannot transmit data to itself, and it is incorporated by the first condition in (4.4). The second condition ensures that if node 'm' is within the transmission range of 'n', then M_{mn} =1, else M_{mn} =0 (3rd condition). Then, the number of non-zero elements in each row is estimated

to deduce the row (node) with maximum connections with other selected sensor nodes. The nodal connectivity array 'C_N', of size 1 x N_s, describes the interconnectivity among the selected nodes. Each value in C_N states the number of nodes connected with the column node. In other words, when all the values of C_N are greater than 1, a fully connected wireless network is achieved with the selected sensor nodes. When a node is not connected with other selected nodes, the corresponding value in C_N will be 0. C_N is estimated in an iterative procedure (hence, C_N^i), and at the 0th iteration, the values of the maximum non-zero row are copied to C_N^0.

a)



b) **Matrix M**

Nodes	N3	N4	N6	N7	N8	N9	# non-zero values
N3	0	0	1	0	1	0	2
N4	0	0	0	1	0	0	1
N6	1	0	0	1	0	0	2
N7	0	1	1	0	0	0	2
N8	1	0	0	0	0	1	2
N9	0	0	0	0	1	0	1

c) ***0th iteration***

$C_N^0 = [0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$ N3

#non-zero values in C_N^0, $|C_N^0 > 0|$ =2

Checks: $|C_N^0 > 0| < N_s$

d) ***1st iteration***

$C_N^1 = [0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$ N3

$+[1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]$ N6

$+[1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]$ N8

$C_N^1 = [2 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1]$

#non-zero values in C_N^1, $|C_N^1 > 0|$=5

Checks: $|C_N^1 > 0| < N_s$; $|C_N^1 > 0| > |C_N^0 > 0|$

e) ***2nd iteration***

$C_N^2 = [0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$ N3

$+[1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]$ N6

$+[0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]$ N7

$+[1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]$ N8

$+[0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0]$ N9

$C_N^2 = [2 \quad 1 \quad 2 \quad 1 \quad 2 \quad 1]$

#non-zero values in C_N^2, $|C_N^2 > 0|$ = 6 (= N_s)

Checks: $|C_N^2 > 0| \not< N_s$

Figure 4.1 Estimating wireless connectivity among the selected nodes

Consider the example presented in Figure 4.1, with connectivity shown by red-colored two-headed arrows. Assume that the nodes 3, 4, 6, 7, 8, and 9 are selected for placing sensors. The matrix M and corresponding row-wise addition of non-zero elements are given in Figure 1b. Following this, the value of the maximum connection row is assigned to C_N^0. The superscript '0' refers to the 0th iteration count. So, when the $|C_N^0 > 0| = N_s$, all the nodes are connected. If the number of non-zero elements in C_N^0 are less than N_s, then the algorithm proceeds to the next iteration. In the first iteration, all the rows of codes corresponding to the value of '1' in C_N^0 are added column-wise to yield C_N^1. Two checks are performed, firstly whether all the nodes are not connected, i.e. $|C_N^1 > 0|$<N_s and whether new nodes have been attached to C_N^1 (performed as: $|C_N^1 > 0| > |C_N^0 > 0|$). When both conditions are true, the algorithm proceeds to the next

iteration and performs the steps carried out in 1st iteration. The first condition terminates the algorithm when all nodes are connected, while the second algorithm terminates when there is no improvement in the number of connected nodes. After the iteration stops, if the total number of non-zero elements in the final row is less than N_s, a penalty is added to the objective function. Let C_{diff} give the number of elements greater than or equal to 1 in the final C_N, then the penalty is levied as

$$P_1 = \sigma\left(N_s - C_{diff}\right) \tag{4.5}$$

where 'σ' is the penalty coefficient, a very large positive value. A counter-example for the above example network is described in the appendix section for further clarification. It must be noted that, for a given transmission range, connectivity is ensured if two sensor nodes are within transmission range, estimated by Euclidean metric. The hindrances arising from elevated structures like buildings, trees, etc., are not considered in this study. A penalty is included to remove the trivial solution of placing two sensors at the exact location to maximize sensor detection redundancy. With the unavailability of data, distributed sensors will give more data to the utility manager. For a given 'S' set of sensor nodes and N_s number of sensors, this penalty is given as,

$$P_2 = \sigma(N_s - |distinct(S)|) \tag{4.6}$$

|distinct(s)| function would remove duplicate sensor nodes from the set S, leaving only the distinct sensor nodes.

Thus, the fitness functions for maximizing network coverage and sensor detection redundancy are given as,

$$Z_1 = NC - P_1 - P_2 \tag{4.7}$$

$$Z_2 = SDR - P_1 - P_2 \tag{4.8}$$

In the following sub-section, the algorithms for generating sensor placement are discussed.

4.2.3 Greedy search

4.2.3.1 Sensor location for Maximum Network coverage

The sensor nodes for network coverage are selected to cover the maximum number of nodes with given N_s wireless sensors. Let set S be the set of sensor nodes (initially empty) and s^* be

the node under evaluation for inclusion in the set S. The node with the maximum network coverage is estimated and added to set S. Next, we add the node (s*) to set S if it produces the highest network coverage when combined with the existing nodes in set S and is within the transmission range of at least one of the nodes in set S. This process is continued till N_s nodes are selected. This generic greedy selection procedure to select nodes for wireless sensor placement is referred to in this study as Cov_{max}. It is essential to acknowledge that when using the greedy search, multiple nodes may achieve the same maximum network coverage value. In this study, Cov_{max} is improvised by selecting the farthest node that provides the maximum network coverage. The farthest node is estimated by calculating the spread, given by,

$$spread = \frac{1}{|S|}\sum_{i \in S} C_{i,s^*} \tag{4.9}$$

The spread of node 's*' is estimated by summing the distance between node 's*' and all the nodes in S and dividing it by the total number of nodes in set S. The function to estimate the fitness of a given set of nodes is then modified for network coverage as,

$$Z_1^* = \frac{100}{N}\sum_{i=1}^{N} d_i + eps \times spread - P_1 - P_2 = Z_1 + eps \times spread \tag{4.10}$$

The above procedure describes the algorithm *s*Cov_{max}. Placing the sensors at the dead-end nodes will yield maximum network coverage, as water quality at these nodes reflects the quality at their corresponding upstream nodes. However, greedy selection is limited by sensor range constraint; that is, the subsequent sensor nodes are selected within the range of the previous sensor nodes. This limitation is more pronounced when the sensor transmission range is much

Algorithm1: sCov$_{max}$	**Algorithm2: sCov$_{cen}$**
$S = \phi$ $S = \underset{k \in N}{\text{argmax}}(Z_1(\{k\}))$ $while \ \|S\| <= N_s$ $s^* = \underset{k \in N}{\text{argmax}}(Z_1^*(\{k\} \cup S))$ $S = S \cup s^*$ end	$S = \phi$ $S = N_{centralnode}$ $while \ \|S\| <= N_s$ $s^* = \underset{k \in N}{\text{argmax}}(Z_1^*(\{k\} \cup S))$ $S = S \cup s^*$ end

smaller with respect to the WDN area. To overcome this, a variation of algorithm *s*Cov_{max} is proposed by defining the initial node as the central node of the WDN rather than the node with

maximum network coverage. In Algorithm *s*Cov$_{cen}$, the centroid of the WDN is first estimated by dividing the sum of the coordinates of all the nodes by number of nodes. Then, the node closest to the centroid is fixed as the initial node ($N_{centralnode}$). The rest of the selection procedure is the same as in *s*Cov$_{max}$.

4.2.3.2 Sensor location for Maximum Sensor detection redundancy

The second set of greedy solutions is generated to maximize sensor detection redundancy, defined as the number of nodes covered by more than one sensor. This necessitates that the nodes are covered by at least one sensor, and then the second sensor covers the nodes covered by the first sensor. The generic approach (termed as Red$_{batch}$)for maximizing sensor detection redundancy involves selecting half of the nodes for maximum network coverage and then selecting the remaining nodes for maximum sensor detection redundancy. For a given sensor node at the dead-end, choosing either the upstream prior node or the node at the adjacent branched-out dead-end will result in an equal sensor detection redundancy. However, selecting the latter will cover one extra node, whereas the former will not increase the network coverage as it is already covered by the dead-end node. The algorithm *r*Red$_{batch}$ incorporated this notion by selecting nodes for sensor detection redundancy with emphasis on network coverage using the function,

$$Z_2^* = \frac{100}{N}\sum_{i=1}^{N} R_i + eps \times \frac{Z_1^*}{Np \times 100} - P_1 - P_2 \qquad (4.11)$$

In another version, algorithm *r*Red$_{alt}$, the nodes are chosen alternatively for network coverage

Algorithm3: *r*Red$_{batch}$

$S = \phi$

$S = \underset{k \in N}{\operatorname{argmax}}(Z_1(k))$

while $|S| <= N_s$

  if $|S| <= \frac{N_s}{2}$

    $s^* = \underset{k \in N}{\operatorname{argmax}}(Z_1^*(\{k\} \cup S))$

    $S = S \cup s^*$

  else

    $s^* = \underset{k \in N}{\operatorname{argmax}}(Z_2^*(\{k\} \cup S))$

    $S = S \cup s^*$

  end

end

Algorithm4: *r*Red$_{alt}$

$S = \phi$

$S = \underset{k \in N}{\operatorname{argmax}}(Z_1(\{k\}))$

while $|S| <= N_s$

  if $|S|$ *is odd*

    $s^* = \underset{k \in N}{\operatorname{argmax}}(Z_2^*(\{k\} \cup S))$

    $S = S \cup s^*$

  else

    $s^* = \underset{k \in N}{\operatorname{argmax}}(Z_1^*(\{k\} \cup S))$

    $S = S \cup \{s^*\}$

  end

end

and sensor detection redundancy. The first node is determined based on network coverage, followed by sensor detection redundancy, then again for network coverage, and so on till N_s sensors are selected.

The four algorithms mentioned above produce four initial solutions that will be added to the initial population in the NSGA-II algorithm. The NSGA-II algorithm, i.e. multi-objective genetic algorithm (MGA), is utilized to find the Pareto fronts. The parameters of MGA are: population size = 250; generations=100; crossover fraction=0.8; mutation= adaptive feasible; Pareto fraction: 0.35; number of runs = 100. The final non-dominating set is generated by compiling all the Pareto points and checking each Pareto point for non-dominance. A Pareto point is added to the final non-dominating set only if no other Pareto point solution existed that was better than it in both objectives.

4.3 Results and Discussion

The simulations and multi-objective optimization were performed in a system with configuration, processor: Intel i7-8565U CPU, 1.80GHz; RAM: 16 GB; Windows 64-bit OS. The contamination propagation matrix (CPM) developed for Guwahati WDS is of size 326x326. An initial assessment of the CPM reveals that 146 sensors will be required to achieve 100% network coverage, corresponding to 146 dead-end nodes of the network. However, the maximum number of sensors in this study for the WDN was set to 20. Above 20 sensors, the differential increase in network coverage (estimated as an increase in network coverage per new sensor added) was less than 0.61%, as each new sensor was associated with a dead-end node of the branches. The placement of WSN in WDN will have scenarios with sufficiency or insufficiency in monitoring and wireless connectivity arising due to the number of sensors and sensor transmission range, respectively. In this study, 4 different cases are defined to verify the

Table 4.1 Cases for WSN placement in WDN

Cases	Number of sensors	Transmission range	Sensors for WDN monitoring	Sensors for connectivity
Case 1	10	0.25 km	Insufficient	Insufficient
Case 2	10	1 km	Insufficient	Sufficient
Case 3	20	0.25 km	Sufficient	Insufficient
Case 4	20	1 km	Sufficient	Sufficient

model's applicability over 4 different possibilities, as shown in Table 4.1. It must be noted that when transmission range is 1 km, 10 sensors can form a connected WSN covering the complete WDN, whereas when transmission range = 0.25 km, even 20 sensors will be insufficient to cover the complete WDN [103].

4.3.1 Greedy search algorithms

The greedy algorithms developed in the study are first assessed by generating sensor solutions for placing 1 to 20 sensors.

Figure 4.2 gives the network coverage achieved with the increasing number of sensors at the transmission ranges: 0.25 km and 1 km. It can be observed that the proposed algorithms, $sCov_{max}$ and $sCov_{cen}$, generally outperform the Cov_{max} algorithm in terms of achieving higher network coverage values for the placement of 1 to 20 sensors.

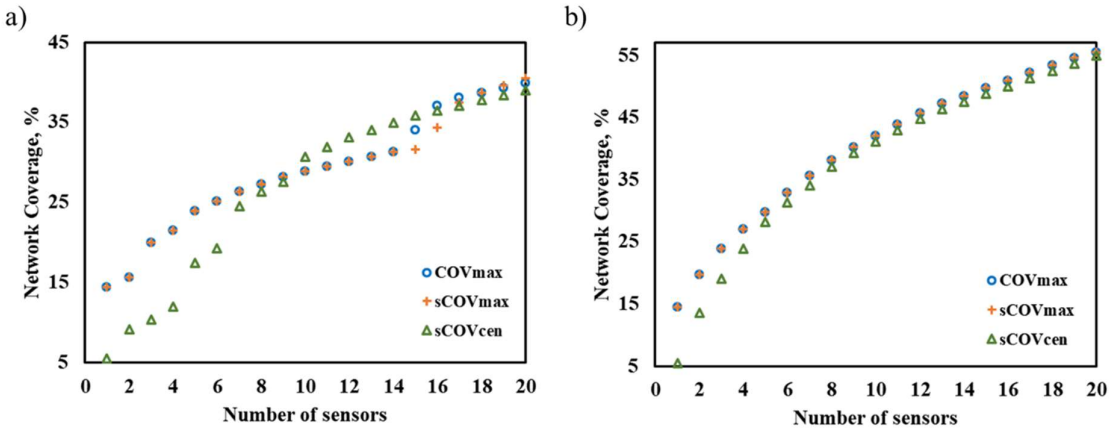


Figure 4.2 Comparison of algorithms for maximum Network coverage a) TR = 0.25 km b) TR = 1 km

At the transmission range of 0.25 km, Cov_{max} and $sCov_{max}$ initially yield higher network coverage values than $sCov_{cen}$ when N_s varies from 1 to 9. This is because $sCov_{cen}$ starts with the $N_{centralnode}$ that has a lower network coverage (only 5.52%) compared to the maximum network coverage node (14.41%) used in Cov_{max} and $sCov_{max}$. However, when the number of sensors is increased from 10 to 15, $sCov_{cen}$ performs better in terms of network coverage as it allows for the expansion of feasible communication range in all directions from the center of the water distribution network (WDN). In this specific range of N_s (10 to 15), the algorithms, Cov_{max} and $sCov_{max}$, have lower network coverage as previously selected nodes (1 to 9) are concentrated near one dead-end. The propagation to the other regions of the WDN requires the

selection of the intermediate nodes. Once these nodes are chosen, algorithms Cov_{max} and $sCov_{max}$, yield network coverage values comparable to those of $sCov_{cen}$. For placing 16 and 17 sensors, the Cov_{max} performs better than the proposed algorithms. In $sCov_{max}$ and algo $sCov_{cen}$, the prime criterion for selecting a node is network coverage; however, if two nodes yield equal network coverage, then the node with the maximum spread is selected. In doing so, the algorithms might overlook the closer nodes, which are closer to the other nodes of higher network coverage, as observed for placing the 16^{th} and 17^{th} sensors. For placing 18th and above sensors, sCovmax yields the highest network coverage value, up to 70 sensors.

At the transmission range of 1 km, Cov_{max} and $sCov_{max}$ algorithms resulted in equal and maximum network coverage for all 20 sensors. The increase in sensor transmission range helps the algorithms to reach from one dead-end to another and, thus, yield maximum network coverage. In contrast, algorithm $sCov_{cen}$ resulted in lower network coverage values for all 20 sensors due to the bias towards the central node, which may not result in the best combination of nodes within the larger transmission range. Overall, the proposed algorithms, $sCov_{max}$ and $sCov_{cen}$, perform better than the Cov_{max} algorithm for maximizing network coverage, with $sCov_{max}$ being more effective for larger transmission range and $sCov_{cen}$ being more effective for smaller transmission range.

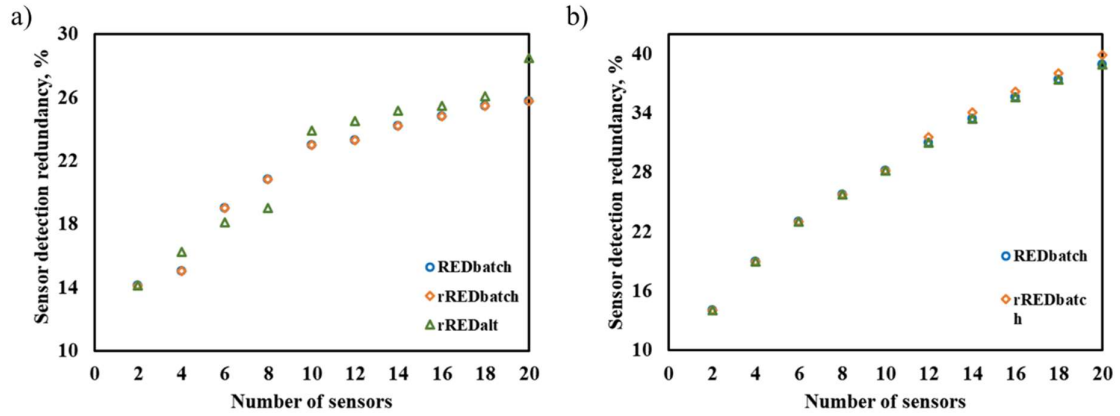


Figure 4.3 Comparison of algorithms for maximum Sensor detection redundancy a) TR = 0.25 km b) TR = 1 km

The sensor detection redundancy achieved by the proposed algorithms for N_s from 2 to 20 is shown in

Figure 4.3. Since the first sensor node is always the maximum network coverage node and the second node has to be selected within the transmission range of 0.25 km, all three algorithms choose the same node and, thus, yield the same sensor detection redundancy at N_s=2. It can be

observed that $r\text{Red}_{alt}$ yields a higher sensor detection redundancy value than the Red_{batch} and $r\text{Red}_{batch}$ for all N_s, except for N_s = 6 and 8. When the sensors are selected alternatively in $r\text{Red}_{alt}$, there is freedom for the sensors to be chosen for each objective individually, in contrast to the other two algorithms where nodes are chosen in batches. Alternative selection enables a larger feasible communication range and consequentially maximum extraction of the two objectives for every two nodes selected. Meanwhile, batch-wise selection of nodes is constrained by feasible regions available within transmission range. To further illustrate, consider 6 sensors {a, b, c, d, e and f} are to be placed in WDN. In batch-wise selection, sensors a, b and c will be placed for network coverage and d, e and f for sensor detection redundancy. 'a' is the maximum network coverage node, and 'b' has to be placed within the transmission range for 'a', and 'c' has to be placed within the range of 'a' and 'b'. The following 3 sensors are designated for sensor detection redundancy. However, it is evident from the intuition that they will be positioned close to the nodes previously selected for network coverage.

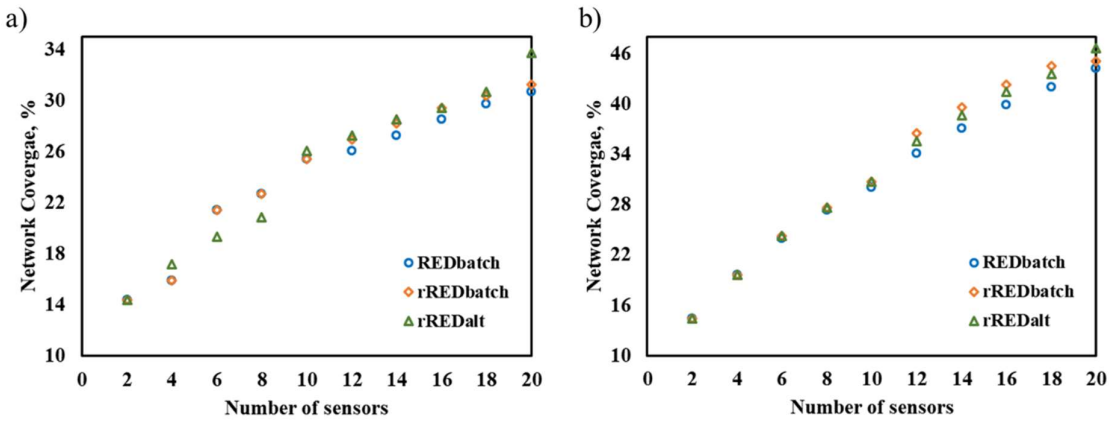


Figure 4.4 Network coverage observed for SDR greedy algorithms a) TR = 0.25km b) TR = 1km

In $r\text{Red}_{alt}$, the first and second sensors are placed for network coverage and sensor detection redundancy, respectively, i.e. a, c and e will be utilized for maximizing network coverage. In contrast, b, d, and f will be used for maximizing sensor detection redundancy. 'a' will be placed at the maximum network coverage node and 'b' will be positioned near 'a' on the node with maximum sensor detection redundancy. The next sensor 'c' (the second sensor designated for network coverage) will have a larger feasible region than the second sensor designated for network coverage in $r\text{Red}_{batch}$. This phenomenon is noticed for up to Ns = 42 (given in Appendix). However, when the transmission range is large (=1 km), the performance of $r\text{Red}_{batch}$ improves, as shown in Figure 4.3b. All three algorithms yield the exact value of sensor

redundancy for up to 10 sensors, and with increment in Ns, algorithm $r\text{Red}_{batch}$ outperforms Red_{batch} and algorithm $r\text{Red}_{alt}$. The proposed algorithms for sensor detection redundancy give precedence to network coverage in case two nodes of equal sensor detection redundancy have been realized.

Figure 4.4 indicates the advantage of the inclusion of network coverage in the sensor detection redundancy fitness function (at N_s>=10), with an average increment of 3.97% and 4.95% in network coverage values at transmission range = 0.25 km and transmission range = 1 km, respectively, in comparison to Red_{batch}. This is important because monitoring data-deficient WDNs requires consideration of both network coverage and sensor detection redundancy, which are critical objectives.

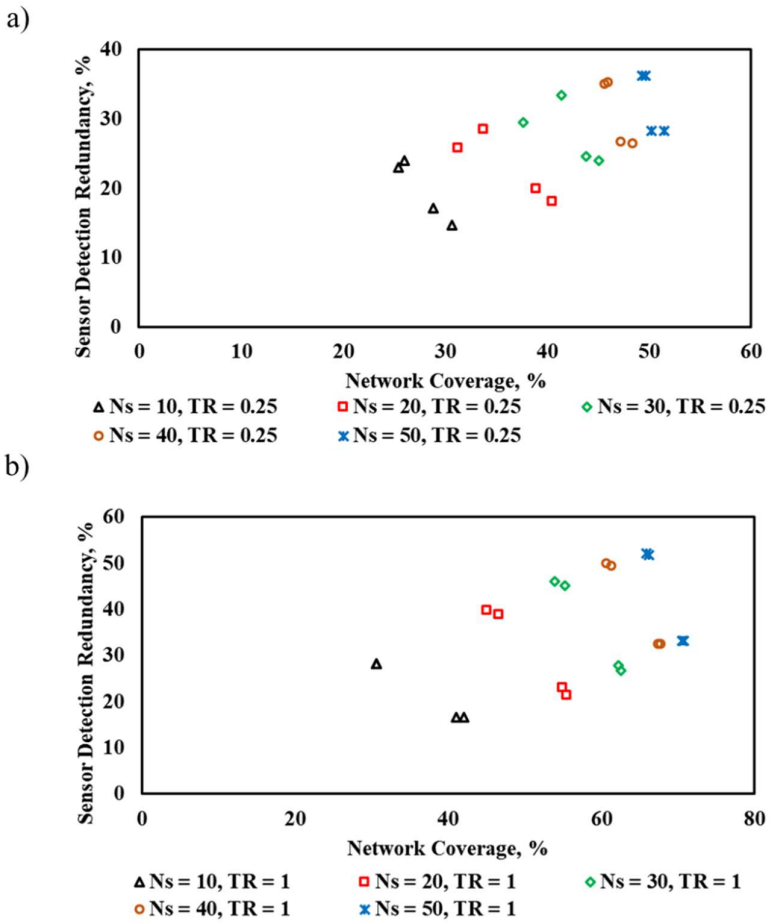


Figure 4.5 Greedy solutions from all the algorithms a) TR = 0.25 km b) TR = 1 km

The results of all 4 algorithms $s\text{Cov}_{max}$, $s\text{Cov}_{cen}$, $r\text{Red}_{batch}$ and $r\text{Red}_{alt}$ for N_s = 10, 20, 30, 40 and 50 and sensor transmission range = 0.25 km and 1 km are shown in Figure 4.5. The results show that as the number of sensors and the sensor range increase, both the network coverage

and sensor redundancy also increase. This is expected, as a larger number of sensors and a larger transmission range allow more nodes to be monitored.

For the given transmission range of 0.25 km and 1 km, the network coverage and sensor detection redundancy increase by an average of 35.55% and 27.48%, respectively, when the number of sensors increases from 10 to 20. However, the net increase in network coverage and sensor detection redundancy depreciates as the number of sensors is increased from 20 to 50. This is expected as, after 20 sensors, the contribution of each sensor towards network coverage is two nodes, which is further reduced to one node after 35 sensors (Figure 4.6a). This indicates that adding more sensors beyond a certain point does not provide significant additional benefits in network coverage and sensor detection redundancy.

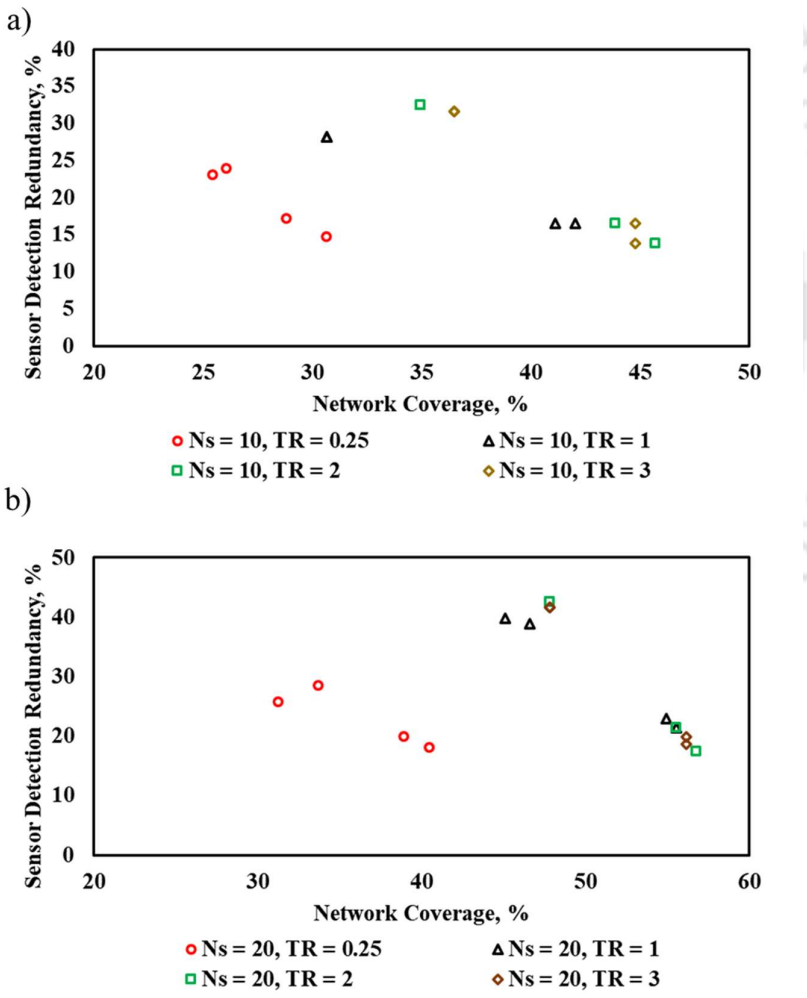


Figure 4.6 Greedy solutions from all the algorithms a) Ns = 10 b) Ns = 20

Similarly, when the transmission range is increased from 0.25 km to 1 km, the network coverage and sensor detection redundancy increase by an average of 34.78% and 21.86% for

Ns = 10 and 20, respectively (depicted in Figure 4.6b). Extending the transmission range from 1 to 2 km results in only a minor increase in network coverage and sensor detection redundancy. Further increases in transmission range do not significantly impact the objective values. Overall, these results indicate that there is an optimal number of sensors and sensor range that can achieve maximum network coverage and sensor detection redundancy in the given water distribution network. Adding more sensors or increasing the sensor range beyond this optimal point does not provide significant additional benefits.

Furthermore, in all 4 cases considered, the results exhibit that as the network coverage increases, sensor detection redundancy reduces and vice versa. The objectives, network coverage and sensor detection redundancy, are competitive and require multi-objective optimization. In the following sub-section, the results of multi-objective optimization are discussed.

4.3.2 Multi-objective optimization

Multi-objective optimization was performed using the multi-objective Genetic Algorithm tool 'gamultiobj' available in MATLAB2019a. The parameters were: Population size: 200; Iteration: 100; Crossover fraction: 0.80; Tolerance: 1e-04 number of runs: 100. The number of iterations for greedy solutions incorporated MGA was reduced to compensate for function evaluations performed during the greedy search. The legends 'MGA' and 'gMGA' refer to standalone MGA without using solutions from greedy search in the initial population and MGA using solutions from greedy search in the initial population, respectively.

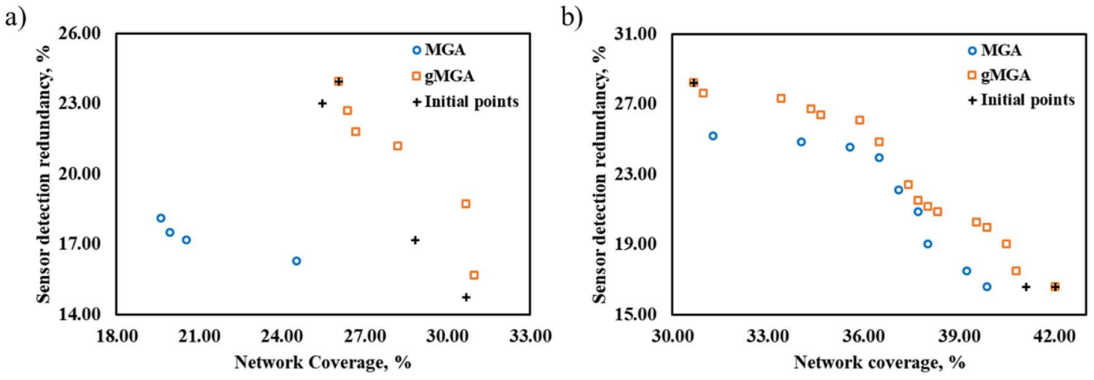


Figure 4.7 MOO results of MGA and gMGA approach for Ns = 10 a) TR = 0.25 km b) TR = 1km

The results of the greedy algorithms indicate that the performance of $s\text{Cov}_{max}$, $s\text{Cov}_{cen}$, $r\text{Red}_{batch,}$ and $r\text{Red}_{alt}$ depend on N_s and transmission range for a given WDN. There is no one

single algorithm that will work for all the cases. Thus, all the solutions generated by the proposed algorithms are incorporated into the initial population of NSGA-II. The Pareto points for N_s =10, along with the initial solutions, are shown in Figure 4.7. All the solutions generated by MGA were dominated by gMGA for both transmission ranges. The maximum network coverage from MGA at 0.25 km was 20.79% less than gMGA, while the maximum sensor detection redundancy was less by 24.36. Similarly, the maximum network coverage and sensor detection redundancy from MGA at transmission range =1 km were 5.11% and 10.87% lower than the gMGA approach. Thus, including a better initial solution led to better extreme points in the Pareto set.The Pareto solutions for N_s = 20 sensors are shown in Figure 4.8. As observed for N_s = 10, all the Pareto points from gMGA dominate the Pareto points from the MGA approach. The maximum network coverage and sensor detection redundancy in gMGA is higher by 17.89% and 27.09% (average) than observed in the MGA approach.

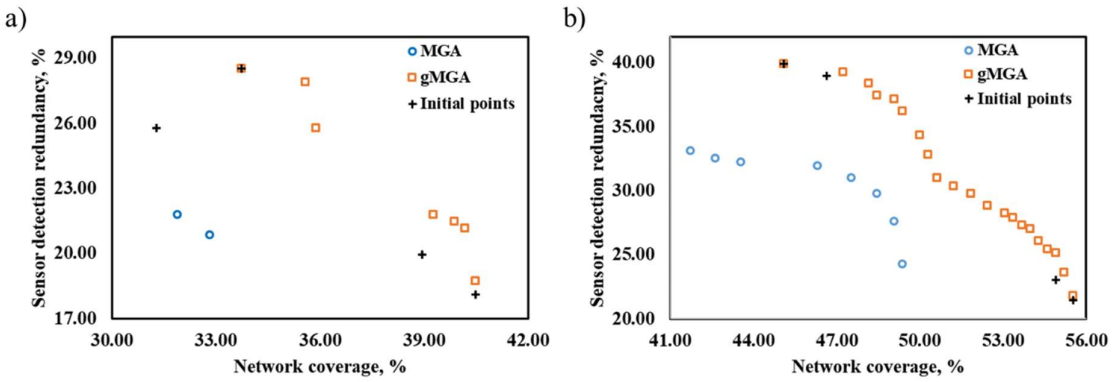


Figure 4.8 MOO results of MGA and gMGA approach for Ns = 20 a) TR = 0.25 km b) TR = 1 km

The number of Pareto points generated in MGA is less than the Pareto points generated in the gMGA approach for all the cases given in Table 4.2.

Table 4.2 Number of Pareto points generated

Cases	MGA	gMGA
N_s = 10, Range = 0.25 km	4	6
N_s = 10, Range = 1 km	9	16
N_s = 20, Range = 0.25 km	2	7
N_s = 20, Range = 1 km	8	21

Introducing greedy solutions might aid in generating sensor solutions with better objective values than yielded at the initial solutions. It is evident through the maximum network coverage achieved in gMGA at N_s=10, transmission range = 0.25 km, and N_s=20, transmission range = 1 km, higher than the network coverage realized in the initial solution. Collectively, it is

apparent that the greedy solutions in the population of MOO for WSN improve the network coverage and sensor detection redundancy, as well as the number of options (sensor placement solutions) available for the engineers to choose from.

The sensor locations of N_s=10 and transmission range = 1 km generated from gMGA are shown in Figure 4.9. For ease of explanation, only the maximum network coverage, maximum sensor detection redundancy, and intermediate solution are depicted. The nodes selected for maximum network coverage are far apart, as expected. Whereas maximum sensor detection redundancy has 5 nodes of maximum network coverage, the remaining 5 are placed near these nodes. In the absence of the 'no two sensors at same location' penalty, the sensor nodes for sensor detection redundancy would have been placed at the first 5 network coverage nodes. It can be observed that nodes 422 and 423 will yield the same sensor detection redundancy values, but the latter will result in a higher network coverage value. The proposed algorithm handles this by comparing network coverage values; thus, node 423 was selected for sensor placement. The sensor locations of the intermediate solution have the 5 maximum network coverage nodes as observed in the maximum sensor detection redundancy solution. However, 4 out of 5 nodes for sensor detection redundancy are placed away from the common nodes (encircled in Figure 9). The solutions corresponding to the maximum value of objectives depict that sensor nodes are either spread out or placed in close pairs in the WDN. The former results in poor reliability of the measured data, while the latter facilitates monitoring of only fewer nodes of the WDN. Thus, an intermediate solution would have to be assessed to better place WSN in WDN.

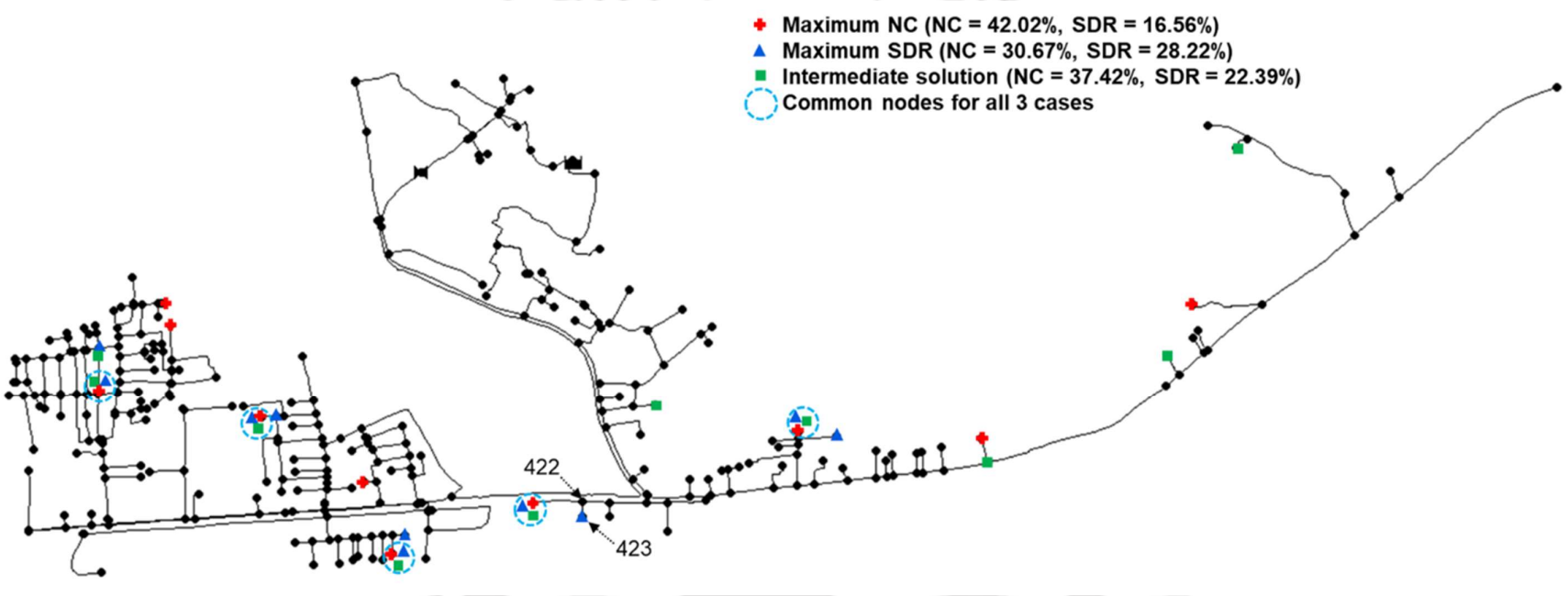


Figure 4.9 Sensor locations at Ns=10 and TR = 1 km at maximum NC, maximum sensor detection redundancy and intermediate solution

4.4 Summary

The implementation of WSN for monitoring WDNs with data deficiency is investigated in this chapter. The objectives for monitoring the WDN (network coverage and sensor detection redundancy) are defined first, followed by an efficient method to verify the connectivity among the wireless sensors. These are utilized to generate sensor placement solutions for each objective using greedy search algorithms. The effect of transmission range and number of sensors on monitoring objectives was examined further. Then, multi-objective optimization for maximizing network coverage and sensor detection redundancy is performed with and without greedy initialization. The Pareto front yielded in MOO with greedy initialization dominated the one obtained in MOO without greedy initialization. Finally, the necessity of choosing an appropriate solution from the Pareto front is demonstrated by presenting the wireless sensor placement solutions at maximum network coverage, maximum sensor detection redundancy and an intermediate point.

Chapter 5: MILP model for optimal placement of WSN in WDNs

5.1 Foreword

In water distribution networks (WDN) monitoring, the mobility of wireless sensors leverages its adaptability for real-time mitigation. The performance of a sensor-based monitoring system depends on its optimal placement derived based on the network characteristics. When there is a significant change in these characteristics, such as the extension of WDN or enhanced monitoring of critical areas, the existing sensor placement design may not be the best solution. Subsequently, a new sensor placement solution has to be implemented based on the latest characteristics of WDN. In these scenarios, wireless communication enables the reallocation of installed sensors within the transmission range. The time and effort required to reallocate wireless sensors is minute compared to reallocating wired sensors. The optimization problem formulation presented in Chapter 4 assumes that wireless sensors are fixed at specific nodes, and the number of sensors designated for WDN monitoring remains unchanged. Furthermore, the WSN placement is studied for a single base station architecture; the initial node for a given algorithm serves as the focal point for sensor placement. This requirement is satisfied by identifying either the node with maximum network connectivity or a central node. In one of these strategically selected nodes, a base station sensor is installed, while the remaining nodes are equipped with node stations. Furthermore, the number of nodes to be selected corresponds directly to the available sensor count, which facilitates a greedy selection process. Upon completion of the algorithm, a connected tree with Ns vertices emerges, represented by the chosen nodes.

Extending greedy algorithms to a multi-base station topology within wireless sensors encounters two primary limitations:

1. The number of prerequisite nodes is contingent upon the number of base stations that can be installed. For instance, if the budget permits the installation of three base station sensors, then three initial nodes are necessary to commence the greedy algorithms, potentially resulting in three separate connected trees. While selecting the first node (the maximum network coverage node) is straightforward, determining the subsequent two nodes introduces many possible combinations. Simplistically, any node beyond the '$TRxN_s$' range could be a candidate for a prerequisite node. Evaluating each potential node significantly increases computational demands.

2. After establishing the number of base stations, it is imperative to determine the corresponding number of nodes for each station. The complexity of potential combinations

escalates with the increase in Ns and the number of base stations. For example, if we consider five nodes (a, b, c, d, and e) with two sensors and nodes a and e designated as base stations, the combinations for node a could include: { {}, {b}, {c}, {d}, {b, c}, {c, d}, {b, d}, {b, c, d} }, with the complementary combinations for node e. Determining the optimal configuration necessitates evaluating the objective values for each scenario. This complexity is further compounded in real-world applications where selections are made from a larger pool of nodes.

These complexities render the application of greedy search methods for multi-base station configurations inefficient. While existing literature offers a Mixed-Integer Nonlinear Programming (MINLP) model, its complexity often requires the use of rapid heuristic methods, such as Genetic Algorithms (GA), for resolution. This study introduces a Mixed-Integer Linear Programming (MILP) model to address the multi-base station WSN placement challenge within WDNs. This model has been validated against the scenarios examined in previous studies. Subsequently, a multi-objective optimization analysis is conducted using Augmented ε-constraint programming, followed by a discussion on the benefits of implementing multi-base station sensor placement. Finally, the developed model is applied to reallocation issues that commonly arise during the operation of actual WDNs

5.2 Methodology – WSN with mobility for WDN

The problem statement of this chapter is derived from Chapter 4, defined as choosing N_s nodes from the N_{node} of a WDN, such that the base stations and their corresponding node stations form a wireless network and yield maximum network coverage and sensor data redundancy.

The proposed approach for wireless water quality sensor placement within Water Distribution Networks (WDN) employs a Mixed-Integer Linear Programming (MILP) framework. This methodology is designed to tackle the intricate challenge of optimally positioning water quality sensors to ensure effective monitoring and contamination detection within the network.

The objective functions within this model are crafted to minimize critical metrics for data-deficient WDNs, such as maximizing network coverage and sensor detection redundancy. These objectives are subject to a series of constraints that govern the hydraulics of the network, sensor placement limitations, and connectivity requirements.

To solve this complex MILP problem, the study leverages flow cuts and secondary objectives capable of handling the combinatorial nature of the problem. The solutions derived from this approach aim to provide a balanced trade-off between competing objectives: network coverage

for detecting most of the contamination events and ensuring detection redundancy for reliability.

The versatility of the proposed MILP-based approach is further demonstrated through its application to various monitoring scenarios, including the repositioning of sensors due to network maintenance, extension, or sensor failure. This flexibility is crucial for adapting to the dynamic nature of WDNs and operational challenges.

In essence, the MILP-based approach for wireless water quality SPP represents a comprehensive and robust framework that can significantly enhance the decision-making process for sensor placement in WDNs, leading to improved water quality monitoring and public health protection.

5.2.1 Objectives

The WDN monitoring objectives from Chapter 4, network coverage and sensor detection redundancy, for monitoring data-deficient networks are carried forward in this chapter. Similarly, the CPM required for estimating the objectives is also adapted from the previous chapter. The mixed integer linear programming model for maximizing the network coverage and sensor detection redundancy is presented below. The number of nodes is estimated, where det_{ij} is the contamination propagation matrix. x_i indicates whether node 'i' is selected for placing node station sensor (if yes =1 else =0). And y_i indicates whether the node 'i' is selected for placing the base station node. t_j indicates the number of sensors covering each node, while d_j indicates whether a node is covered. The sum of d_j then results in the number of nodes covered and the percentage of nodes covered is estimated by (5.4).

$$t_j = \sum_i x_i \times \det_{ij} \quad \forall j \tag{5.1}$$

$$d_j \le t_j \quad \forall j \tag{5.2}$$

$$t_j \le M \times d_j \quad \forall j \tag{5.3}$$

$$NC = \frac{100}{N_{node}} \times \sum_j d_j \tag{5.4}$$

The sensor redundancy objective is estimated by estimating the number of contamination events detected by more than one sensor. Here, based on t_j, h_j indicates whether a node is covered at least 'Nr' times. The sum of h_j gives the number of nodes covered at least 'Nr'

times and the average value calculated in the SDR equation. Here '∂' is small positive number less than 1.

$$h_j \le \frac{t_j}{Nr} \qquad \forall j \tag{5.5}$$

$$t_j - Nr + \partial \le M \times h_j \qquad \forall j \tag{5.6}$$

$$SDR = \frac{100}{N_{node}} \times \sum_j h_j \tag{5.7}$$

5.2.2 Constraints

The constraint on the number of sensors to be deployed is below.

$$\sum_i x_i + \sum_i y_i \le N_s \tag{5.8}$$

$$y_i + x_i \le 1 \qquad \forall i$$

The connectivity matrix denotes whether two nodes are within the transmission range. This matrix can be generated with a simple if-else operation. First, a matrix is defined as 'C' with rows and columns equal to the number of nodes in the matrix. Then, the values in each cell are filled with 0 or 1 if the corresponding cell in the distance matrix has a value less than the transmission range.

The connectivity between the selected nodes is implemented by considering phantom flow between the selected nodes. The base station nodes are the source nodes. The source node supplies a maximum of N_s-1 flow units to node station nodes. The node station connected to the base station node consumes 1 unit of flow. A continuous positive variable 'f_{ij}' defines flow from node i to node j.

$$f_{ij} \ge 0 \qquad \forall i, j \tag{5.9}$$

The flow can happen only between the nodes selected for sensor placement. This is ensured by the constraints,

$$f_{ij} \le M(x_i + y_i) \qquad \forall i \tag{5.10}$$

$$f_{ij} \le M(x_j + y_j) \qquad \forall j \tag{5.11}$$

The flow between two nodes is only possible if they are within the transmission range. This constraint is given by,

$$f_{ij} \le MC_{ij} \qquad \forall i, j \tag{5.12}$$

The below constraint ensures that any node does not supply to itself.

$$f_{ij} = 0, \; if \; i = j \qquad \forall i, j \tag{5.13}$$

The total flow to the source node is 0. This is incorporated by,

$$\sum_j f_{ji} \le (N_s - 1) \times (1 - y_i) \qquad \forall i \tag{5.14}$$

In the above constraint, if y_i is 1, then the term $\sum_j f_{ji}$ has a value of 0. That is, no flow enters the base station node. And when y_i=0, the $\sum_j f_{ji}$ has the range [0 N_s -1]. The selected nodes, except the source node, can supply any flow between 0 to N_s -1 units to other selected nodes. Once the source nodes are selected, the flow balance between the nodes ensures the connectivity between the nodes. This is included as,

$$\sum_j f_{ij} - \sum_j f_{ji} \le y_i (N_s - 1) - x_i \quad \forall i \tag{5.15}$$

The term $\sum_j f_{ij}$ denotes the total flow units entering node 'i' and $\sum_j f_{ji}$ denotes the flow units entering node 'i'. The difference between these two gives the net flow through a node. It is at least '1' for all the selected nodes except the source nodes. In the case of the source node, the term $\sum_j f_{ji}$ is equal to 0, consequently, $\sum_j f_{ij}$ will have a value between [0 N_s -1]. For all the unselected nodes, the net flow is zero. It must be noted that the constraints are active only when a node is selected for placing sensors. When a node is selected, then the flow consumption at every selected node is '1', but the flow emanating from the base node can be higher than the total flow consumed by the connected nodes. The equations (5.1) to (5.15) define the MILP model for solving the wireless sensor placement problem.

5.2.3 Modifications for improving the MILP model

In further improving the model, the following cuts were introduced.

A) Maximum outflow from any node:

$$\sum_j f_{ij} \le (N_s - 1) \times (x_i + y_i) \qquad \forall i \tag{5.16}$$

This cut is an extension of the maximum flow cut (for maximum capacity of a link) applied to the total flow emanating from a node. As each node consumes only one unit of flow and the flow leaving the source node is N_s-1, it can be inferred that the flow leaving any node cannot be more than N_s minus 1. The second part on the right-hand side of (5.16) ensures that flow emanates from a node if and only if it is selected for sensor placement.

B) Maximum inflow to any node

$$\sum_i f_{ij} \le \left(N_s - 1\right) \times \left(x_j + y_j\right) \qquad \forall j \tag{5.17}$$

This cut is similar to the above-described cut but focuses on the total amount of flow entering any node.

C) Total maximum flow:

$$\sum_i \sum_j f_{ij} \le \frac{N_s(N_s - 1)}{2} \tag{5.18}$$

This constraint is a combination of constraints A and B. It can be inferred that the total flow between the selected nodes cannot be more than $\frac{N_s(N_s-1)}{2}$. This value is realized by assuming that each selected unit is connected to only one other selected node and estimate the total flow among the selected nodes.

D) Incorporating the total flow within the network in the objective function

In this approach, the objective function was modified to select the flow pattern with the lowest total flow among the selected nodes. The denominator on the right-hand side term is used for scaling the secondary objective. It ensures that an insignificant increase in the primary objective is prioritized over a significant decrease in the flow term. This modification aims to aid the MILP solver in choosing a solution when multiple solutions of the same network coverage values but different flows are generated.

$$Objective: NC - eps \times \frac{\sum_i \sum_j f_{ij}}{\left(N_s \times N\right)^2} \tag{5.19}$$

$$Objective: SDR - eps \times \frac{\sum_i \sum_j f_{ij}}{\left(N_s \times N\right)^2} \tag{5.20}$$

Another constraint was added in combination with the above constraints that fixes the total number of node stations based on the number of base stations and is given by

$$-\sum_i x_i \leq \sum_i y_i - N_s \tag{5.21}$$

5.2.4 Multi-objective optimization

The objectives, network coverage and sensor detection redundancy, are competing, and thus, the engineer will have to choose from the trade-off solutions. The previous work performed for MOO on these objectives proves that there are sensor placement designs that provide the same network coverage but yield different sensor detection redundancy values. This scenario can also be reversed, that there might be solutions with network coverage values even though they have the same sensor detection redundancy value. The best solution among all the solutions is the one with the maximum sensor detection redundancy value at a given network coverage value or vice versa. To overcome this issue, the Augmented ε-constraint (AUGMECON) [104] method is used for generating the Pareto solutions. AUGMECON maximizes the objectives function lexicographically to overcome inefficient solutions. First, the range of sensor detection redundancy is determined by the following procedure:

Step1: Maximize SDR, yields SDR_{max}

Step2: Maximize NC, yields NC_{max}

Step3: Maximize SDR s.t. $NC \geq NC_{max}$, yields SDR_{min}

The range of the sensor detection redundancy is [SDR_{min} SDR_{max}]. This range is cut into a number of intervals, resulting in 'intervals+1' grid points. These grid points are 'ε' values used in the ε-constraint method for generating Pareto solutions. The formulation now becomes,

$$Maximize \;\; NC + eps \times s \tag{5.22}$$

$$s.t \qquad SDR - s \geq \varepsilon \tag{5.23}$$

Here, 's' is a surplus or slack variable. It is the difference between the 'ε' and maximum 'SDR' value that can be obtained at the given NC value. The value of eps should be such that the product 'eps x s' does not shift the value of maximum value of network coverage but improves the sensor detection redundancy. If the maximum value possible for sensor detection redundancy, at maximum network coverage, is 'ε', then the value of 's' is zero. Any positive

value of ‘s’ denotes that a higher value sensor detection redundancy exists at the maximum network coverage than ‘ε’.

5.2.5 Specific applications of the developed MILP model:

5.2.5.1 Mandatory monitoring stations for supply management:

The water supply board or management segregates the WDN into multiple district metered areas (DMAs) for ease of operation and maintenance. Each DMA is connected with the source supply line, and then water entering the DMA propagates to all the demand points within the DMA. Thus, during the maintenance of one of the DMAs, only the entrance valve of that particular DMA has to be closed while all other DMAs continue receiving water. Placing water quality sensors at the entrance of the DMAs ensures that water entering the DMAs is safe. Similarly, a sensor at the exit of the water sources, such as WTP, reservoirs or water tanks, ensures that the water entering the WDN is safe. The wireless SPP for this use-case can be translated as ‘given the location of few wireless sensors, find the optimal location of other sensors for maximizing WDN monitoring while maintaining connectivity’.

The heuristically important nodes can either have node station or base station sensor. Let S_M be the set of indices corresponding to nodes where sensors have to be placed. The following constraint ensures that these nodes are always selected.

$$y_i + x_i = 1 \qquad \forall i \in S_M \tag{5.24}$$

The above problem can be extended to scenarios where WDN undergoes extension. WDNs often go through extensions to cater to new regions. These extensions are generally a year-long process. Sensors are placed in the existing WDN to monitor the water quality in the current system. Once the extension is completed, new sensors can be added to monitor the extended portion of WDN. This will require updating the number of nodes, CPM, and connectivity matrix followed by solving the MILP with the above constraint given in (5.24).

5.2.5.2 Repositioning of sensors as few nodes have been prioritized or designated as critical.

In this case, some existing sensors must be shifted to the critical nodes. This situation might arise after flushing or maintenance of a section of the network. The sensor solution must ensure that only a few sensors are repositioned while maintaining connectivity.

Let the positions of critical nodes for monitoring be given by $S^{\#}$. The first constraint dictates that the sensors are placed at the critical nodes. However, the sensors can be either base station sensors or node station sensors.

$$y_i + x_i = 1 \qquad \forall i \in S^{\#} \tag{5.25}$$

Let the location of base stations and node stations of existing sensors be given by y^* and x^*. The number of changes made in the new sensor placement solution will be,

$$xy_i^* = x_i^* + y_i^* \qquad \forall i = 1...N \tag{5.26}$$

$$xy_i = x_i + y_i \qquad \forall i = 1...N \tag{5.27}$$

$$xy_i^* - xy_i \le xy_i^{df} \qquad \forall i = 1...N \tag{5.28}$$

$$-\left(xy_i^* - xy_i\right) \le xy_i^{df} \qquad \forall i = 1...N \tag{5.29}$$

The variables xy_i^* and xy_i denote the overall nodes designated with sensors (node station and base station) in the existing solution and new solution, respectively. The variables xy_i^{df} indicate whether there is any modification in the sensor designation at node 'i'. xy_i^{df} is '1' when node 'i' does not have a sensor in the initial set up and but a sensor is designated to it or node 'i' had a sensor in the initial set up and but later the sensor is removed. The swapping of sensors between node locations does not incur any cost in this formulation because when sensors have to be swapped, the sensor installation setup of the previous sensors can be used. This is achieved by combining the sensor locations of node station and base station in the variable xy_i^{df}.

The total number of changes is then estimated by the sum of xy_i^{df} over all the nodes. It must be noted that allowing swapping might lead to swapping node station and base station locations even when unnecessary. This is constrained by estimating the changes in base station location, which can be calculated as,

$$y_i^* - y_i \le y_i^{df} \qquad \forall i = 1...N \tag{5.30}$$

$$-\left(y_i^* - y_i\right) \le y_i^{df} \qquad \forall i = 1...N \tag{5.31}$$

The objective of the final sensor placement design is the monitoring of WDN while a solution with the least change in the existing positions

$$Objective = NC - \frac{eps}{N_s}\left(\sum_i xy_i^{df} + \frac{1}{N_s^2}\sum_i y_i^{df}\right) \tag{5.32}$$

5.2.5.3 Repositioning of the sensors when sensors are damaged.

In this case, the main issue arises from the loss of connectivity. Suppose the damaged sensor was relaying measurement data. In that case, a new route or position of the existing sensors must be modified to reinstate connectivity while ensuring there is less compromise on the observability and sensor detection redundancy. The connectivity constraints will ensure a path between sensors, whereas the objective is to reinstate connectivity with minimum changes. The equations (5.26) through (5.31) estimate the total changes in the sensor placement design. Also, as the existing sensors are damaged, the number of available sensors is given by,

$$\sum_i x_i \le \sum_i x_i^* \tag{5.33}$$

$$\sum_i y_i \le \sum_i y_i^* \tag{5.34}$$

The objective for reinstating the connectivity with minimum change in sensor placement design is given by,

$$Objective = -\left(\sum_i xy_i^{df} + \frac{1}{N_s^2}\sum_i y_i^{df}\right) + \frac{eps}{N} \times NC \tag{5.35}$$

5.3 Results and Discussion

5.3.1 Analysis of MILP model and cuts

The simulations and multi-objective optimization were carried out in a system with configuration, processor: Intel i7-8565U CPU, 1.80GHz; RAM: 16 GB; Windows 64-bit OS. The contamination propagation matrix (CPM) developed for Guwahati WDS was obtained from the previous section. All the models were run in IBM ILOG CPLEX Optimization Studio, version: 22.1.0.0. Firstly, the model is solved for maximizing each of the objectives (network coverage and sensor detection redundancy) for the 4 cases of varying N_s (10 and 20) and transmission range (0.25 km and 1 km). This results in 8 cases to be analyzed. The maximum number of base station sensor nodes was fixed to 3 during optimization of various cases, and

the rest of the sensors were node station sensors. Besides the additional constraints given by (5.16) to (5.21) , the combinations of these constraints were also tested.

The performance of the model and cuts is assessed by the number of cases reaching optimality and the time needed to solve them. Table 5.1 summarizes the number of constraints, variables, and the number of cases reaching optimality for the model in 300 sec and 900 sec, and the naming convention of the cuts and combinations. No extra variables were included for the cuts; thus, the number of variables is equal in all the models. All the cuts, A, B, and C, and their combinations increase the number of constraints in the model. The option D is the modification in the objective function and thus has a similar number of constraints. The original model could solve only 6 of the 8 cases to optimality in 300 secs, whereas cuts A and B and its combinations were able to solve all 7 cases. However, the cuts C and D solved 5 of 8 cases, fewer than the original model. In 900 secs, only the set AB solved all 8 cases. Further assessment of the models and cuts is described in the following paragraphs.

Table 5.1 Results of MILP and corresponding modifications

Method	Designation	Optimality in 300 secs	Optimality in 900 secs	Variables	Constraints
Original Model	MILP	6	7	107906	321765
Original Model + A	A	7	7	107906	322091
Original Model + B	B	7	7	107906	322091
Original Model + C	C	5	6	107906	322176
Original Model + D	D	5	5	107906	321765
Original Model + A + B	AB	7	8	107906	322417
Original Model + A + B + C	ABC	7	7	107906	322418
Original Model + A + B + D	ABD	7	7	107906	322417
Original Model + A + B + E	AB′	7	7	107906	322418

The computation time for solving represents the time taken to solve the problem to optimality. The computation time to reach the optimal solution for maximizing network coverage at transmission range = 0.25 km is shown in Figure 5.1. For placing N_s = 10, all variations except cuts C and D, lowered the computation time by an average of 94.05%. The MILP took 34360.53 sec to reach a 2% optimality gap for N_s=20, depicting the complexity of the formulation with the increase in the number of sensors. In this case, cuts A and B and their combination with other cuts minimized the solving time. However, the cuts C and D were slower, and thus, the

computation time for 2% optimality was provided. This was also one of the reasons for combining these cuts with A and B to observe if the combination has an impact on the computation time. Though the computation time for transmission range = 1 km was lower than TR = 0.25 km, the trend in the solving time of the MILP and its modification was similar to the results for transmission range = 0.25 km (Figure 5.2).

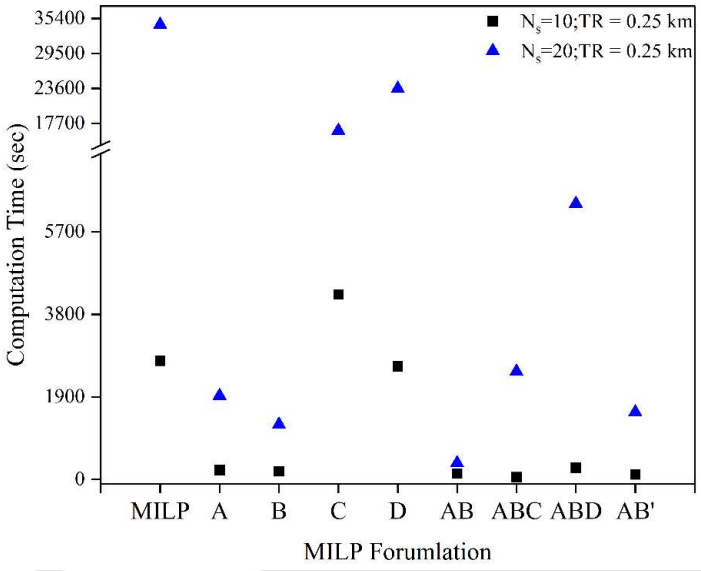


Figure 5.1 Comparison of Computation time for maximizing NC at TR = 0.25 km

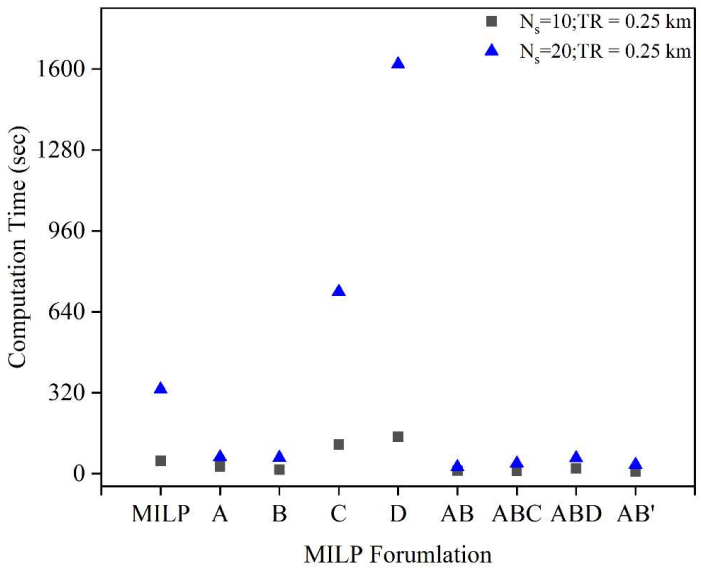


Figure 5.2 Comparison of Computation time for maximizing NC at TR = 1 km

The computation time to reach the optimal solution for maximizing sensor detection redundancy is shown in Figure 5.3 and Figure 5.4. All the formulations yielded the maximum

sensor detection redundancy value within 100 secs at transmission range = 0.25 km, except MILP without any modifications, formulations C and D. The latter two took more time than

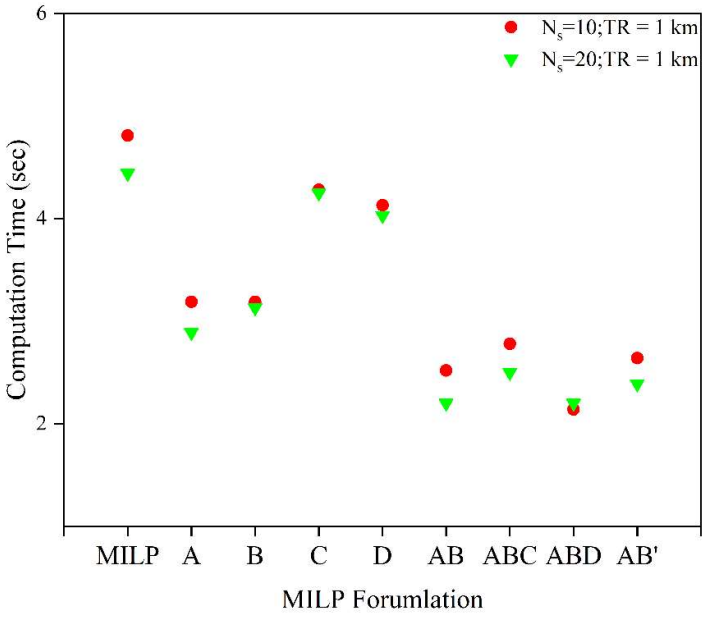


Figure 5.3 Comparison of Computation time for maximizing SDR at TR = 0.25 km

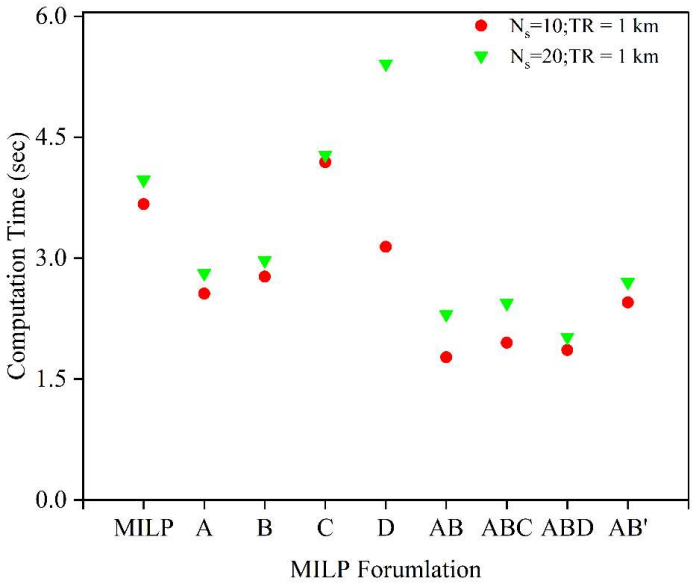


Figure 5.4 Comparison of Computation time for maximizing SDR at TR = 1 km

the actual MILP formulation. The computation time was less than 5 sec when transmission range was increased to 1 km.

5.3.2 Augmented ε-constraint for MOO study

The developed MILP model was further considered for solving the MOO problem of placing wireless sensors in WDN to maximize network coverage and sensor redundancy. In the ε-constrained method for MOO study, one of the objectives is set at a limit while the other objective is maximized. In this study, the sensor redundancy is considered as the limiting objective, and network coverage is maximized. The limits on sensor redundancy are generated by dividing its range into required intervals yielding grid points. The comparison of ranges observed for the four cases (N_s = 10 and 20, R = 0.25km and 1 km) in ε-constrained and AUGMECON is shown in the following Table. The MILP was modeled in IBM ILOG CPLEX optimization studio version 22.1.0.0 academic version. The values of ε were set as follows (Table 5.2):

Table 5.2 ε-value for MOO

Setting	1	2	3	4	5	6	7	8	9	10	11
ε value	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

The range for sensor detection redundancy obtained in ECON and AUGMECON is provided in Table 5.3. The lower limits for cases 3 and 4 in ε-constrained are lower by 6.78% and 1.72% than those obtained in AUGMECON. When the sensor detection redundancy value is limited between the lower limit of ε-constrained and the lower limit of AUGMECON, maximizing coverage will always result in the same value. However, to generate Pareto, a non-dominated set is needed. As all the sensor detection redundancy values in between this range will be equal, the non-dominance criteria will be decided by the maximum value of sensor detection redundancy. In this case, it is evident that the sensor detection redundancy value from AUGMECON is maximum and, thus, will be added to the Pareto set. Furthermore, it can be understood that solving the MILP for the limit of sensor detection redundancy below the AUGMECON lower limit only leads to unnecessary use of computation effort and time. Thus, AUGMECON helps eliminate unnecessary computation required for generating the Pareto by constricting the range of sensor detection redundancy effectively.

Table 5.3 Range of SDR achieved in ECON and AUGMECON.

#	Case	ε-constrained		AUMENCON	
		Lower limit(%)	Upper limit(%)	Lower limit(%)	Upper limit(%)
1	N_s = 10, TR = 0.25	12.27	32.822	12.27	32.822
2	N_s = 10, TR = 1	13.804	34.049	13.804	34.049
3	N_s = 20, TR = 0.25	**16.871**	41.411	**18.098**	41.411
4	N_s = 20, TR = 1	**17.485**	42.638	**17.791**	42.638

The solutions of ε-constrained and AUGMECON after performing MOO at transmission range = 0.25 km and transmission range = 1 km are shown in Figure 5.5 and Figure 5.6, respectively. The points that are dominated are encircled in brown. At transmission range = 0.25 km, only 8 Pareto points are generated as settings 2, 3 and 4 yield the same solution, whereas 6 non-dominated solutions are generated in ECON, as 3 out of 9 solutions are dominated. At N_s=20, 8 Pareto points generated by ECON are dominated by AUGMECON.

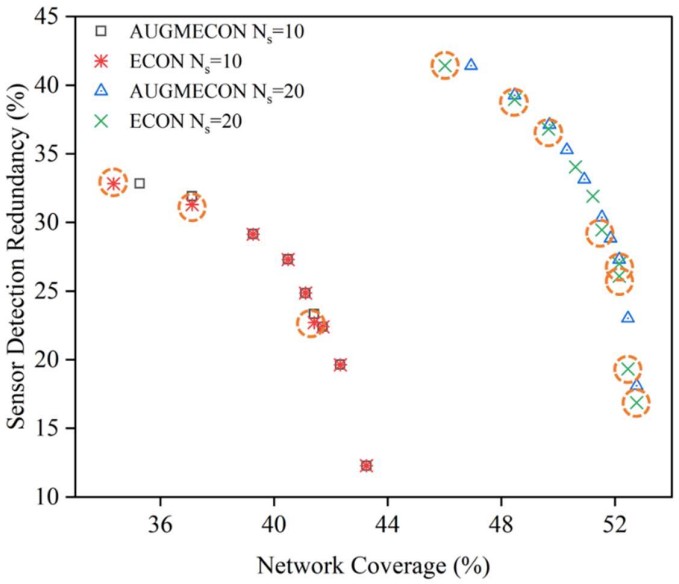


Figure 5.5 Pareto points comparison for AUGMECON and ECON at TR = 0.25 km

At transmission range = 1 km, only one solution (maximum sensor detection redundancy) from ECON is dominated by AUGMECON when N_s=10. On increasing N_s to 20, the number of dominated solutions in ECON also increased to 7. This trend was also observed when 20 sensors were placed with transmission range = 0.25 km. This indicates that when N_s is large, the number of sensor placement designs at a given network coverage value is also large. Each

of these designs will have a different sensor detection redundancy value. This makes it essential to include a strategy that inherently chooses the design with maximum sensor detection redundancy value. Overall, AUGMECON not only improves the range of the objectives but also determines Pareto points efficiently.

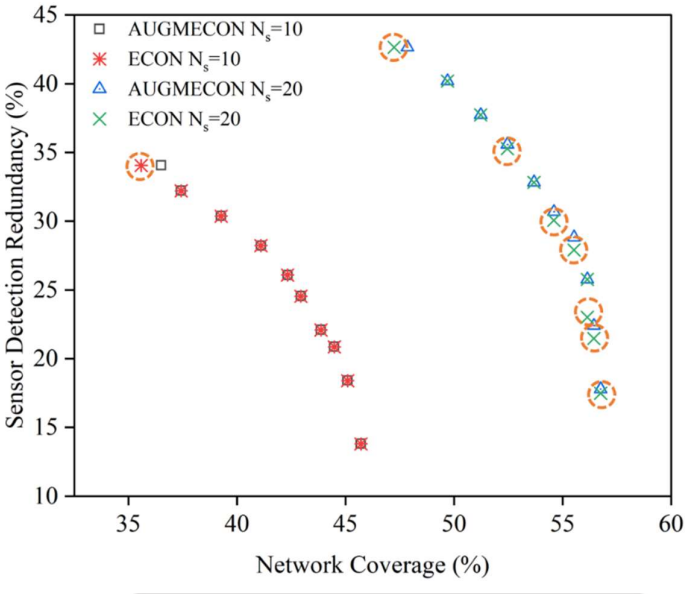


Figure 5.6 Pareto points comparison for AUGMECON and ECON at TR = 1 km

5.3.3 Effect of multiple base station architecture

The advantages of implementing multiple base stations can be demonstrated by comparing the solutions with the single base station architecture. Figure 5.7 depicts the Pareto points obtained with single and multi-base station architecture for the transmission range of 0.25 km. On average, network coverage and sensor detection redundancy increase by about 31.31% and 24.55% when the multiple base sensor stations are incorporated. Due to the single base station architecture, the farthest any two sensors can be is N_sxTR, provided there is a path for the flow of information between the sensors. This constraint is lifted once the multiple base station architecture is implemented.

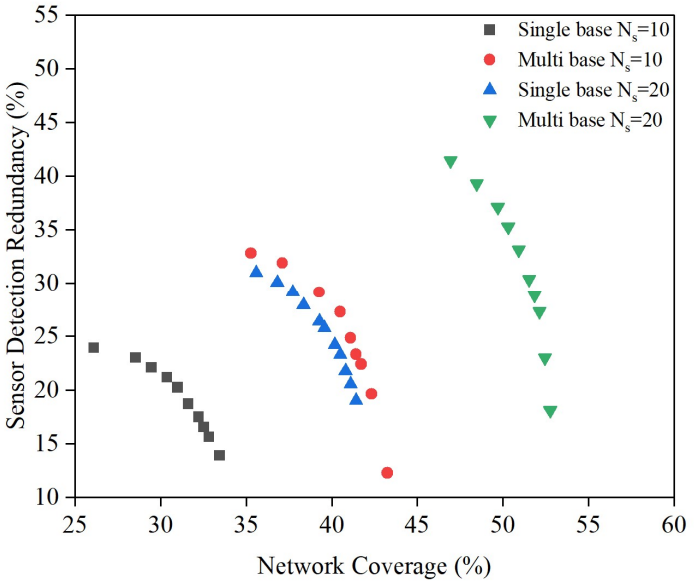


Figure 5.7 Comparison of Single and Multiple base station architecture at TR = 0.25 km

This effect is also illustrated in

Figure 5.8, where the Pareto front of multiple base stations dominates the front observed for single base station allocation.

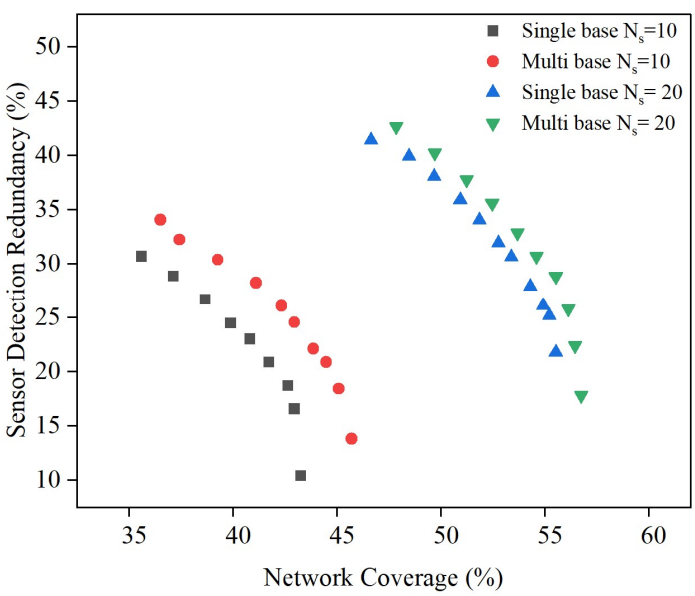


Figure 5.8 Comparison of Single and Multiple base station architecture at TR = 1 km

5.3.4 Sensor solutions for special cases

5.3.4.1 Mandatory locations

The mandatory locations for sensor placement were the reservoir outlet, DMA1 entrance, DMA2 entrance, DMA 3 entrance and DMA4 entrance. The sum of corresponding values of x and y were set to 1 as per (5.25). The maximum network coverage and sensor detection redundancy values observed with these locations are shown in Figure 5.9. The values of network coverage and sensor detection redundancy realized for all four cases are lower than those without mandatory monitoring restrictions.

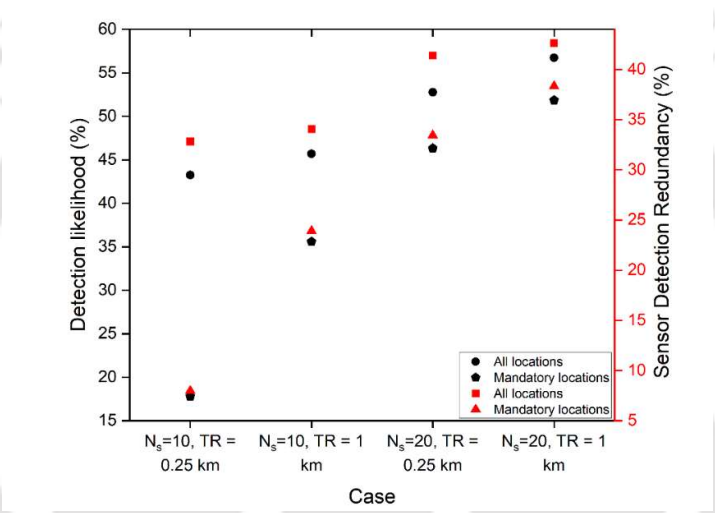


Figure 5.9 Performance of sensor placement solutions when mandatory nodes are monitored

Further investigation of the constraint levied by mandatory locations is performed by solving the MILP at ε=0.5 and plotting the WSN nodes, as shown in Figure 5.10. The mandatory locations are marked with green dotted circles, and the transmission range is marked with blue dotted circles. The sensor placement design utilizes three base stations, described in Figure 5.10. Two of the three base stations are placed in DMA1 with no other sensor connected to them, while the third base station has to be chosen from one of the 8 connected sensors from DMA2, DMA3 and DMA4. The identification of optimal base stations is beyond the scope of this study. However, many algorithms have been discussed in the literature and can be implemented to select the optimal base station node.

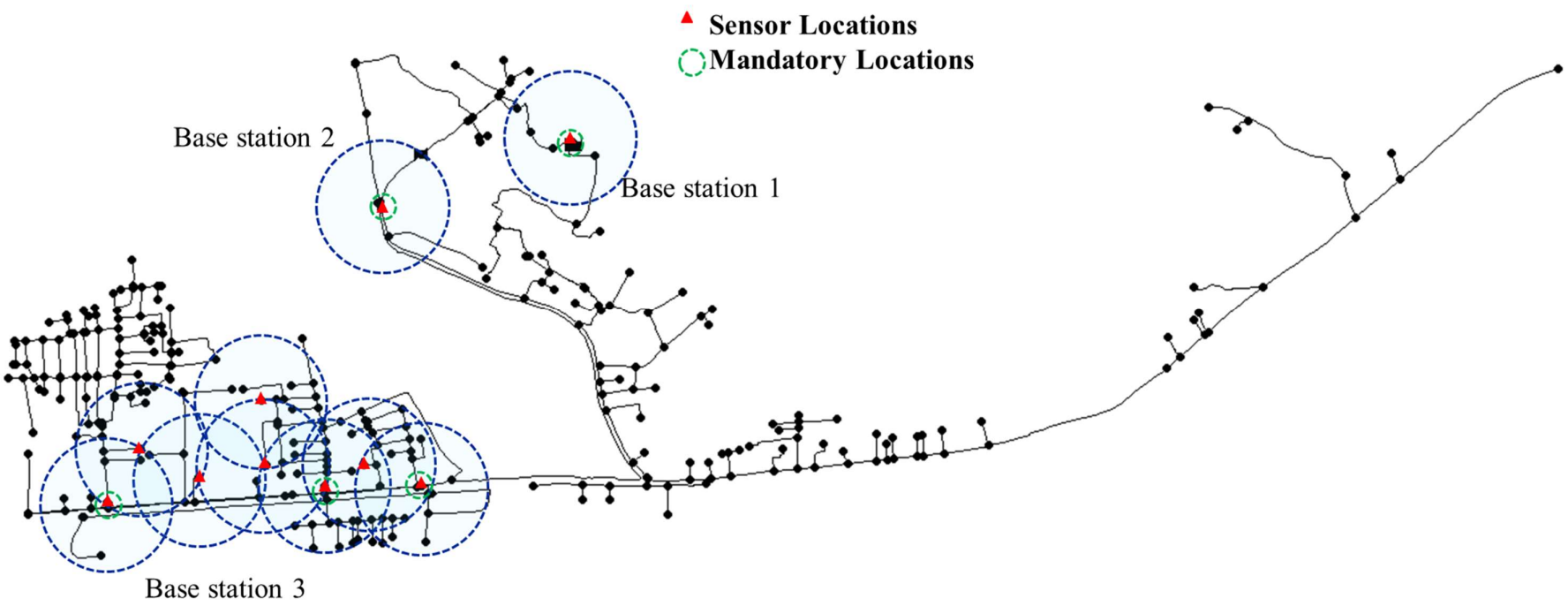


Figure 5.10 Sensor placement design for monitoring mandatory locations

5.3.4.2 Placement at nodes of high priority

The sensor placement design at maximum network coverage observed at transmission range = 0.25 km and N_s = 10 will be considered in this case for shifting the sensors. The new critical nodes selected are shown in Figure 5.11. The list of nodes prior to and after the placement of the sensor on node GJ71 is described in Table 5.4 below.

Table 5.4 Comparison of sensor placement design for critical nodes at TR = 0.25 km

Base station grouping	Sensor locations prior to monitoring GJ71	Sensor locations after monitoring GJ71
Base station 1	GJ452	GJ452
Base station 2	GJ104, **GJ148**, **GJ151**, GJ350, **GJ5**, GJ319	**GJ1**, **GJ71**, **GJ93**, GJ104, **GJ147**, GJ319, GJ350
Base station 3	GJ393, GJ333, **GJ348**	GJ393, GJ333

Note: The node locations that have been either removed or added are written in boldface

The old locations GJ5, GJ148, GJ151 and GJ348 are replaced by locations GJ1, GJ71, GJ93 and GJ147. Among these, the node GJ71 has been deemed critical. The changes in the location of the sensor are mainly because of constraints from the transmission range.

A similar study with transmission range = 1 km (depicted in Figure 5.12) reveals that only one location of the sensor had to be changed (Table 5. 5). This was expected as a wider transmission range facilitates further mobility of sensors without loss of connectivity.

Table 5.5 Comparison of sensor placement design for critical nodes at TR = 1 km (the locations that have been removed and added are shown in boldface)

Base station grouping	Sensor locations prior to monitoring GJ71	Sensor locations after monitoring GJ71
Base station 1	GJ451	GJ451
Base station 2	GJ513, **GJ531, GJ458**	GJ513
Base station 3	GJ350, GJ151, GJ329, GJ104, GJ393, GJ5	GJ350, GJ151, GJ329, GJ104, GJ393, GJ5, **GJ7, GJ71**

Note: The node locations that have been either removed or added are written in boldface

When the penalty with respect to a number of changing locations was included, it was observed that to implement new solutions, 5 sensor locations have to be changed for the cases discussed above. The model does not differentiate between the placement of the node station or base station at the critical location because when only one sensor type is replaced with another, the same installation setup can be re-utilized.

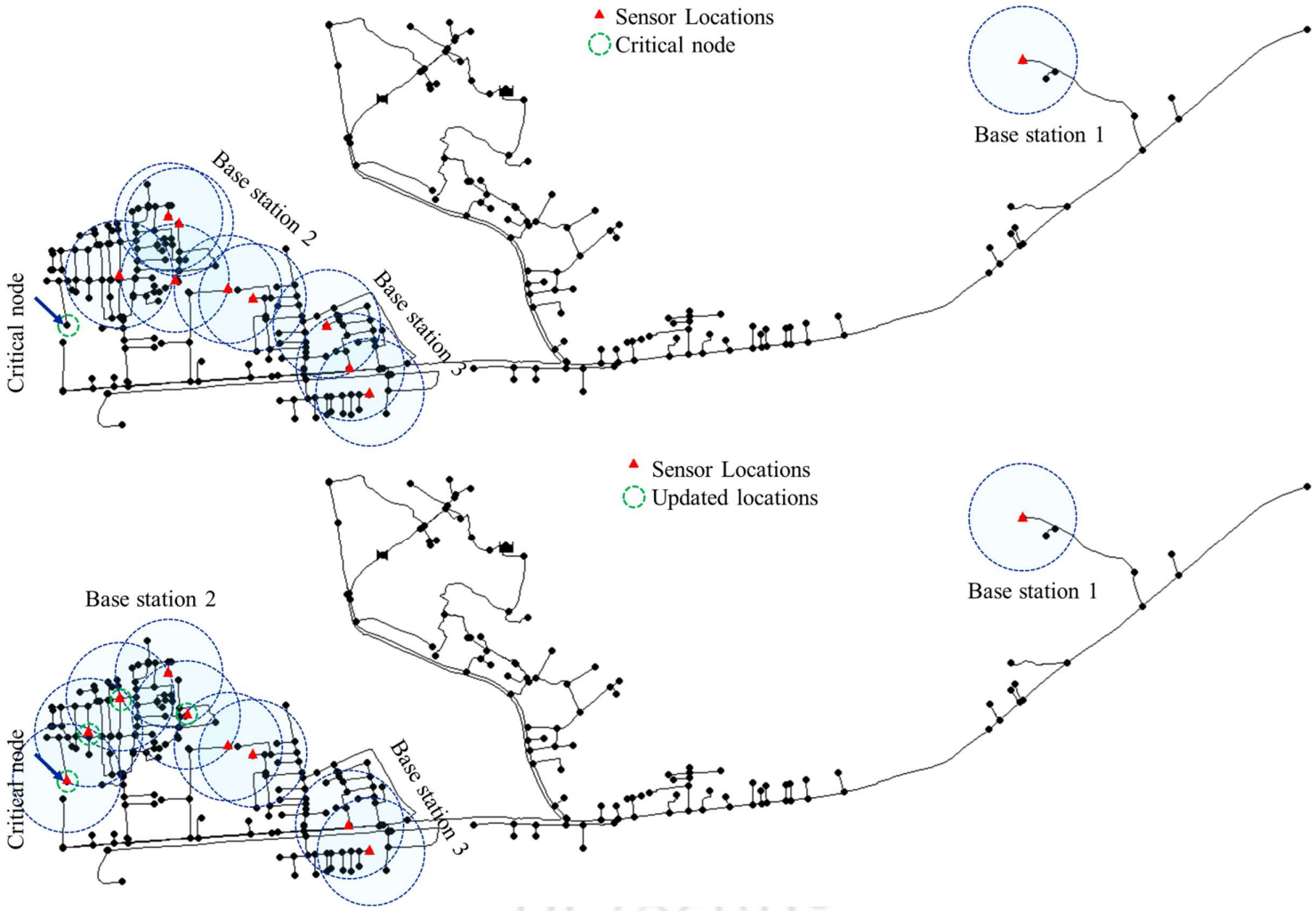


Figure 5.11 Sensor placement design for monitoring high priority node at TR = 0.25 km

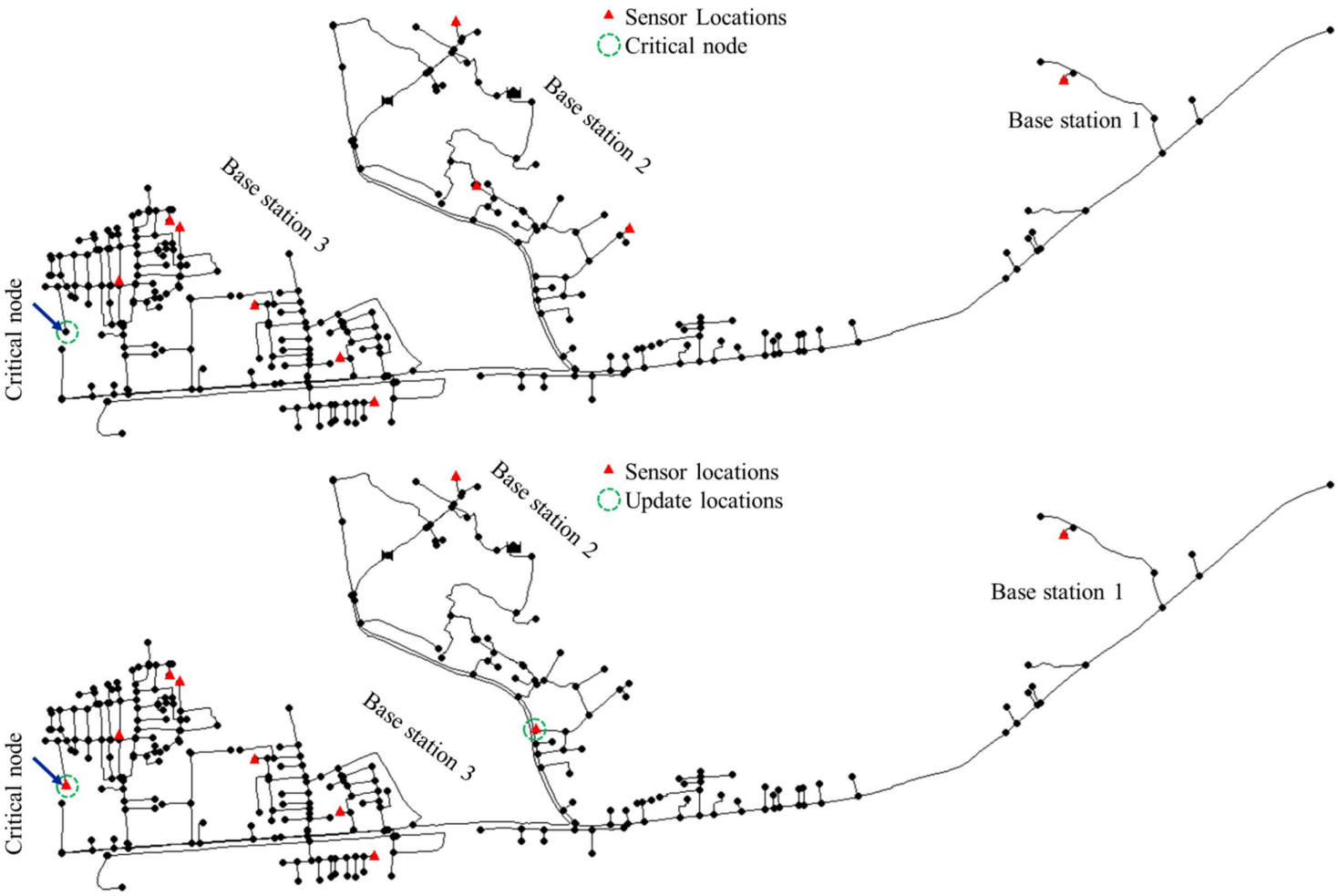


Figure 5.12 Sensor placement design for monitoring high priority node at TR = 1 km

5.3.4.3 Reallocation to ensure connectivity

The analysis of this model was performed in the case of N_s=10 and transmission range = 0.25 km. The base sensor solution was generated for maximum network coverage. With these base results, 10 runs were performed. In each run, one sensor was deemed non-working and corresponding solutions are presented below in Table 5.6.

Table 5.6 Sensor placement design modifications when a sensor is damaged

Damaged sensor location	**Sensor removed from nodes**	**Sensor placed at nodes**	**Sensor nodes (#sensors)**	**Maximum NC (%)**
None	NA	NA	GJ5, GJ393, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ348, GJ148, GJ104 (10)	43.25
GJ5	GJ5	None	GJ393, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ348, GJ148, GJ104 (9)	37.73
GJ393	GJ393	None	GJ5, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ348, GJ148, GJ104 (9)	40.18
GJ333	**GJ333**, GJ393	GJ354	GJ5, GJ354, GJ319, **GJ452**, GJ151, **GJ350**, GJ348, GJ148, GJ104 (9)	37.73
GJ319	GJ319, ***GJ350***	None	GJ5, GJ393, **GJ333**, **GJ452**, GJ151, GJ348, GJ148, **GJ104** (8)	39.57
GJ452	***GJ452***	None	GJ5, GJ393, **GJ333**, GJ319, GJ151, **GJ350**, GJ348, GJ148, GJ104 (9)	33.13
GJ151	GJ151, GJ104	None	GJ5, GJ393, **GJ333**, GJ319, **GJ452**, **GJ350**, GJ348, GJ148 (8)	39.26
GJ350	***GJ350, GJ452***	None	GJ5, GJ393, **GJ333**, **GJ319**, GJ79, GJ348, GJ148, GJ104 (8)	30.67
GJ348	GJ348	None	GJ5, GJ393, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ148, GJ104 (9)	42.02
GJ148	GJ348	GJ148	GJ5, GJ393, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ155, GJ104 (9)	42.02
GJ104	GJ104	None	GJ5, GJ393, **GJ333**, GJ319, **GJ452**, GJ151, **GJ350**, GJ348, GJ148 (9)	40.79

Note: The node locations that have been either removed or added are shown in boldface

The column damaged sensor location denotes the node of the damaged sensor. The minimum reallocation strategy required to reinstate connectivity is given by columns 'sensor removed from nodes' and 'sensor placed at nodes'. The former lists the nodes from existing locations where sensors will not be placed, while the latter lists nodes where sensors will be placed. The sensors are either placed at a new node (not present in the previous solution) or at the node where the sensor was damaged. The next column enumerates nodes for sensor placement with the base station sensor nodes boldened. The maximum network coverage achieved with the sensor placement design is provided in the maximum network coverage column. The first row gives the existing sensor placement design, which is depicted in Figure 5.13.

Damage of sensors at nodes GJ5, GJ393, GJ452, GJ348 and GJ104 require no shifting of the existing sensors to any new nodes. The sensors at these locations are directly connected to base station sensors, or they are at the end of the WSN, due to which their absence does not affect the connectivity within the network. On the other hand, when a node, acting as a bridge between two sensor nodes, is damaged, the node connected to this node is also removed. This is observed for damaged sensors at node GJ151 and GJ333 (base station node), where nodes adjacent to GJ104 and GJ393 are removed, respectively. The number of sensors in the former case is reduced to 8. However, when the base station sensor at node GJ333 is damaged, it is compensated by shifting the existing base station GJ350 to GJ148 and the node station at GJ148 to the new location GJ354. Thus, the final design utilises all the 9 available sensors. Similarly, when the base station GJ350 is damaged, the base station sensor node from GJ452 is shifted to GJ319 to restore connectivity. Overall, the maximum coverage is affected more by the damage of the base station sensor, reduced by an average of 21.75% compared to 6.99% when a node station is damaged.

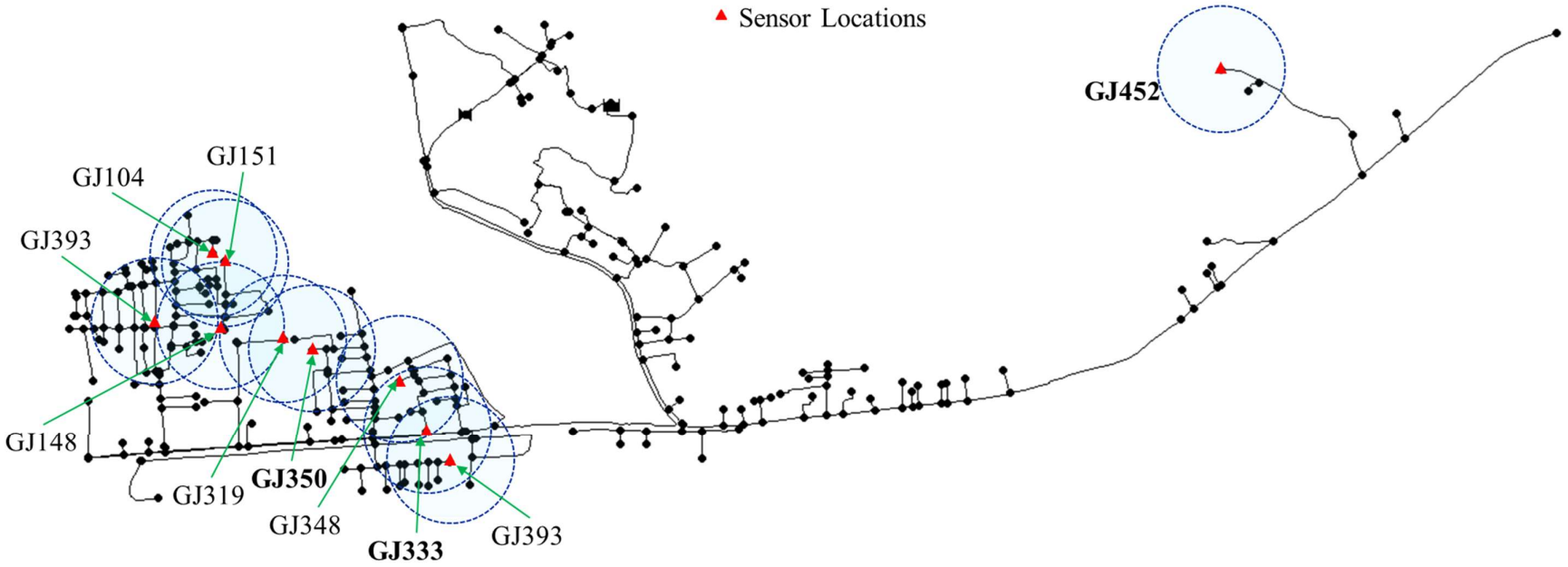


Figure 5.13 Sensor placement design at N_s = 10 and TR = 0.25 km depicting the base station node

5.4 Summary

This chapter discusses the implementation of MILP formulation for optimal placement of wireless water quality sensors. A MILP model is first presented with phantom flows between sensor nodes to ensure connectivity. Modifications are then presented to reduce the computational effort required for solving the problem. The resulting MOO is solved using an augmented 'ε' constraint to aid the solver in choosing only the non-dominating solutions. Then, the effect of using multiple base station architecture is demonstrated by comparing it with single base station results. Finally, the developed MILP is adapted for reallocation scenarios observed in WDNs. The presented sensor placement solution is robust for implementing wireless monitoring of WDN during regular monitoring as well as during anomalies.

Chapter 6: Conclusions

6.1 Foreword

This chapter details the significant contribution of this thesis to develop sensor placement solutions to determine optimal locations for placing water quality sensors in WDN. The study started by improvising the generic sensor placement that has been studied widely in the literature. The initial phase of the study focused on pre-processing the WDN hydraulic and quality simulation data, which will be used in optimization. Moving ahead, the SPP with wireless sensor networks was studied, with objectives aligned with real-case scenarios of data deficiency. The effect of the number of sensors, along with the sensor transmission range, was analysed. In the last chapter, the WSN SPP problem was attempted with multiple data aggregators (base stations) to improve monitoring.

6.2 Sensor placement in WDN

Chapter 3 developed a strategy for the pre-selection of nodes to reduce the decision variable space for optimal sensor placement in a WDN. The network was split into clusters based on a novel heuristics-based approach that incorporates nodal disparity in detecting contamination events. The selection procedure was two-phased; first, a set of nodes is selected based on their ability to detect events, followed by another set of nodes that reduced the detection time or affected population in combination with the previously selected nodes. The proposed algorithm was tested on BWSN Network1 and C-town network and compared based on the quality of the Pareto front produced after MOO to optimize detection likelihood and expected time of detection (or expected affected population).

- The search space was successfully reduced by 45% and 34% for the BWSN and C-town network, respectively, and the Pareto front obtained from the optimal subset of nodes was better in terms of the CR indicator than the complete set of nodes.
- The time taken to generate the final Pareto Front was better than the traditional method in 6 of the 8 MOO studies that were performed. These cases are BWSN Network: 5SPP and 20 SPP for both Z_1 vs Z_2 and Z_1 vs Z_3 studies, and C-town network: 5SPP and 20 SPP for Z_1 vs Z_3 study.
- The proposed method is robust and can be applied to any network irrespective of its complexity, different designs of contamination events, and different objectives with appropriate modifications. The only prerequisites required are the pollution matrices corresponding to the objectives considered for placing sensors.

6.3 WSN placement for data-deficient WDNs

In chapter 4, the wireless sensor network (WSN) placement for data-deficient networks for maximum network coverage and sensor detection redundancy is discussed and implemented on the newly commissioned South Guwahati WDN. Two improvised greedy search algorithms for each objective were presented, and their performance was analyzed. In the second phase, a multi-objective formulation for WSN placement is proposed with an efficient way to evaluate the wireless connectivity of the proposed sensor nodes. The above-discussed solutions were introduced in the initial population of NSGA-II (gMGA) and compared with the results of NSGA-II without including the initial solutions in the initial population (MGA).

- The proposed algorithms yielded sensor placement solutions with higher network coverage and sensor detection redundancy for most scenarios considered (N_s = 1 to 20, transmission range = 0.25km and 1 km) than the generic greedy search algorithms. It was also observed that each algorithm performed best at certain N_s and transmission range, indicating that the solutions of all algorithms are significant.
- The Pareto front yielded in gMGA was superior to the Pareto front from MGA for all the cases considered. The value of the CR indicator is 1 for all the 4 cases considered. Also, gMGA generated at least 50% more sensor placement solutions than MGA. The results also indicate that sensor solutions with objective values higher than the greedy solutions can be realized through optimization.
- A comparison of SPP solutions at three different trade-offs between network coverage and sensor detection redundancy revealed that inclining towards maximum network coverage position, the nodes are far apart, whereas focusing on sensor detection redundancy results in pair-wise placement of sensors. An evenly spread-out sensor placement solution near the centre of the Pareto front will result in optimal monitoring of the WDN.

6.4 Development of MILP model for placement of WSN in WDN

In Chapter 5, the placement of wireless sensor networks consisting of multiple base station sensor problems is attempted with MILP. The MILP is further enhanced using cuts and secondary objectives to improve the computation time. The MILP study focused on maximizing two objectives: network coverage and sensor detection redundancy, which are competing in the previous chapter. The resulting MOO problem is solved using augmented e-constrained, and its proficiency in overcoming inferior solutions is demonstrated.

- The average computation time taken for maximizing network coverage was 93.78% more than the time required for maximizing sensor detection redundancy. The improvements in the MILP model reduced the computation time by a maximum of 98.9%, observed at N_s=20 and transmission range = 0.25 km.
- Implementation of multiple base station architecture is advantageous when the transmission range of the sensors is small compared to the WDN area. In cases where transmission range and WDN areas are comparable, the WSN placement is leveraged by a multi-hopping data transmission system.
- The sensor placement solutions for reallocation case studies were generated within 180 sec. This is essential, as critical node monitoring and sensor failures would require immediate response, followed by executing mitigation strategies.

6.5 Future Scope

This thesis tries to offer sensor placement strategies for water quality monitoring in WDN, focusing on data transmission mode and WDN data availability. Based on the findings and observations made in this study, the following future scopes are identified:

- The clustering and subsequent node selection depend on the number of objectives considered. This study considered only two objectives in MOO; however, based on the monitoring requirements, more objectives will have to be incorporated. The reduction in decision variable search space can be extended to multiple WDN monitoring objectives such as the amount of contaminated volume consumed, level of service (LOS) based objectives, sensor detection redundancy, etc.
- The wireless sensor placement study performed in this thesis is for data-deficient WDNs. These methodologies will have to be extended to WDNs with temporal flow data.
- Apart from the connectivity, the WSN attributes like energy for transmission, data packet, budget constraints, and corresponding WSN reliability studies will have to be performed.
- The data generated from the pollution matrices can be utilized to develop machine learning models to localize the contamination source. Integrating this with real-time sensor solutions will help in swift mitigation in the event of contamination.
- Overall, the strategies developed in this study can be described as optimal monitoring of flow-based systems. Thus, the proposed methods can be extended to other fields

where flow and monitoring are essential, such as logistics, traffic control, and process industries.

References

[1] WHO/UNICEF, Progress on household drinking water, sanitation and hygiene 2000-2017, 2019. https://washdata.org/sites/default/files/documents/reports/2019-07/jmp-2019-wash-households.pdf.

[2] Government of India 2018 NSS 76th Round: Drinking Water, Sanitation, Hygiene and Housing Condition in India., New Delhi, India, 2008. https://mospi.gov.in/sites/default/files/NSS7612dws/Report_584_final.pdf.

[3] BIS, Indian standards drinking water specifications IS 10500:2012, 2012. http://cgwb.gov.in/Documents/WQ-standards.pdf.

[4] H. Mala-Jetmarova, A. Barton, A. Bagirov, A history of Water distribution systems and their optimisation, Water Sci. Technol. Water Supply. 15 (2015) 224–235. https://doi.org/10.2166/ws.2014.115.

[5] D. Savic, J.K. Banyard, Water Distribution Systems, ICE Pub., 2011. https://books.google.co.in/books?id=I9l0zQEACAAJ.

[6] B. Kowalska, P. Suchorab, D. Kowalski, Division of district metered areas (DMAs) in a part of water supply network using WaterGEMS (Bentley) software: a case study, Appl. Water Sci. 12 (2022) 1–10. https://doi.org/10.1007/s13201-022-01688-2.

[7] P.F. Boulos, B.W. Karney, D.J. Wood, S. Lingireddy, Hydraulic transient guidelines for protecting water distribution systems, J. / Am. Water Work. Assoc. 97 (2005) 111–124. https://doi.org/10.1002/j.1551-8833.2005.tb10892.x.

[8] F. Dong, J. Chen, C. Li, X. Ma, J. Jiang, Q. Lin, C. Lin, H. Diao, Evidence-based analysis on the toxicity of disinfection byproducts in vivo and in vitro for disinfection selection, Water Res. 165 (2019) 114976. https://doi.org/10.1016/j.watres.2019.114976.

[9] X.F. Li, W.A. Mitch, Drinking Water Disinfection Byproducts (DBPs) and Human Health Effects: Multidisciplinary Challenges and Opportunities, Environ. Sci. Technol. 52 (2018) 1681–1689. https://doi.org/10.1021/acs.est.7b05440.

[10] B. Kowalska, D. Kowalski, A. Musz, Chlorine decay and disinfection by-products in water distribution systems, Environ. Eng. - Proc. 2nd Natl. Congr. Environ. Eng. (2007) 191–199.

[11] J.J. Erickson, K.L. Nelson, D.D.J. Meyer, Does intermittent supply result in hydraulic transients? Mixed evidence from two systems, Aqua Water Infrastructure, Ecosyst. Soc. 71 (2022) 1251–1262. https://doi.org/10.2166/aqua.2022.206.

[12] S. Satpathy, R. Jha, Intermittent water supply in Indian cities: considering the intermittency beyond demand and supply, Aqua Water Infrastructure, Ecosyst. Soc. 71 (2022) 1395–1407. https://doi.org/10.2166/aqua.2022.149.

[13] I. Andrić, A. Vrsalović, T. Perković, M.A. Čuvić, P. Šolić, IoT approach towards smart water usage, J. Clean. Prod. 367 (2022) 132792. https://doi.org/10.1016/j.jclepro.2022.132792.

[14] M. Pule, A. Yahya, J. Chuma, Wireless sensor networks: A survey on monitoring water quality, J. Appl. Res. Technol. 15 (2017) 562–570.

https://doi.org/10.1016/j.jart.2017.07.004.

[15] T.P. Lambrou, C.C. Anastasiou, C.G. Panayiotou, M.M. Polycarpou, A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems, IEEE Sens. J. 14 (2014) 2765–2772. https://doi.org/10.1109/JSEN.2014.2316414.

[16] Á. Milánkovich, K. Klincsek, Wireless Sensor Network for Water Quality Monitoring, in: Eur. Proj. Sp. Inf. Commun. Syst., SCITEPRESS - Science and and Technology Publications, 2015: pp. 28–47. https://doi.org/10.5220/0006164600280047.

[17] E. Creaco, A. Campisano, N. Fontana, G. Marini, P.R. Page, T. Walski, Real time control of water distribution networks : A state-of-the-art review, Water Res. 161 (2019) 517–530. https://doi.org/10.1016/j.watres.2019.06.025.

[18] S. Nurani, H. Abdul, W. Lau, Sensors and Actuators B : Chemical Detection of contaminants in water supply : A review on state-of-the-art monitoring technologies and their applications, Sensors Actuators B. Chem. 255 (2018) 2657–2689. https://doi.org/10.1016/j.snb.2017.09.078.

[19] M. V Storey, B. Van Der Gaag, B.P. Burns, Advances in on-line drinking water quality monitoring and early warning systems, Water Res. 45 (2010) 741–747. https://doi.org/10.1016/j.watres.2010.08.049.

[20] P. Kruse, Review on water quality sensors, J. Phys. D. Appl. Phys. 51 (2018). https://doi.org/10.1088/1361-6463/aabb93.

[21] J. Bhardwaj, K.K. Gupta, R. Gupta, A review of emerging trends on water quality measurement sensors, Proc. - Int. Conf. Technol. Sustain. Dev. ICTSD 2015. (2015) 1–6. https://doi.org/10.1109/ICTSD.2015.7095919.

[22] M. Odhiambo, V. Viñas, E. Sokolova, T.J.R. Pettersson, Health risks due to intrusion into the drinking water distribution network: hydraulic modelling and quantitative microbial risk assessment, Environ. Sci. Water Res. Technol. 9 (2023) 1701–1716. https://doi.org/10.1039/d2ew00720g.

[23] I. Yaroshenko, D. Kirsanov, M. Marjanovic, P.A. Lieberzeit, O. Korostynska, A. Mason, I. Frau, A. Legin, Real-time water quality monitoring with chemical sensors, Sensors (Switzerland). 20 (2020) 1–22. https://doi.org/10.3390/s20123432.

[24] B. Lebental, S. Bila, E. Cloutet, C. Dejous, H. Hallil, S. Laporte, B.B. Ngoune, G. Perrin, Y. Ulanowski, Water and air quality monitoring with multiparameter chemical sensors Managing non-idealities from lab to field, Proc. IEEE Sensors. 2022-Octob (2022) 1–4. https://doi.org/10.1109/SENSORS52175.2022.9967256.

[25] J. Monroe, E. Ramsey, E. Berglund, Allocating countermeasures to defend water distribution systems against terrorist attack, Reliab. Eng. Syst. Saf. 179 (2018) 37–51. https://doi.org/10.1016/j.ress.2018.02.014.

[26] US EPA, Online Water Quality Monitoring Resources, (2022). https://www.epa.gov/waterqualitysurveillance/online-water-quality-monitoring-resources (accessed October 28, 2023).

[27] M. del Valle, Sensor Arrays and Electronic Tongue Systems, Int. J. Electrochem. 2012 (2012) 1–11. https://doi.org/10.1155/2012/986025.

[28] W.E. Hart, R. Murray, Review of sensor placement strategies for contamination warning systems in drinking water distribution systems, J. Water Resour. Plan. Manag. 136 (2010) 611–619. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000081.

[29] A. Agathokleous, S. Xanthos, S.E. Christodoulou, Real-time monitoring of water distribution networks, Water Util. J. 10 (2015) 15–24.

[30] L.S. Vamvakeridou-Lyroudia, J. Bicik, M. Morley, D. Savic, Z. Kapelan, A real-time intervention management model for reducing impacts due to pipe isolation in water distribution systems, in: Water Distrib. Syst. Anal. 2010 - Proc. 12th Int. Conf. WDSA 2010, 2012: pp. 209–221. https://doi.org/10.1061/41203(425)21.

[31] Z. Kapelan, D. Savić, H. Mahmoud, A response methodology for reducing impacts of failure events in water distribution networks, Procedia Eng. 186 (2017) 218–227. https://doi.org/10.1016/j.proeng.2017.03.231.

[32] H.A. Mahmoud, Z. Kapelan, D. Savić, Real-time operational response methodology for reducing failure impacts in water distribution systems, J. Water Resour. Plan. Manag. 144 (2018) 1–14. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000956.

[33] A. Maheshwari, A.A. Abokifa, R.D. Gudi, P. Biswas, Coordinated Decentralization-Based Optimization of Disinfectant Dosing in Large-Scale Water Distribution Networks, J. Water Resour. Plan. Manag. 144 (2018) 04018066. https://doi.org/10.1061/(asce)wr.1943-5452.0000979.

[34] A.A. Abokifa, A. Maheshwari, R.D. Gudi, P. Biswas, Influence of Dead-End Sections of Drinking Water Distribution Networks on Optimization of Booster Chlorination Systems, J. Water Resour. Plan. Manag. 145 (2019) 04019053. https://doi.org/10.1061/(asce)wr.1943-5452.0001125.

[35] D. Zeng, S. Zhang, L. Gu, S. Yu, Z. Fu, Quality-of-sensing aware budget constrained contaminant detection sensor deployment in water distribution system, J. Netw. Comput. Appl. 103 (2018) 274–279. https://doi.org/10.1016/j.jnca.2017.10.018.

[36] M. Mastaller, P. Klingel, Application of a water balance adapted to intermittent water supply and flat-rate tariffs without customer metering in Tiruvannamalai, India, Water Sci. Technol. Water Supply. 18 (2018) 347–356. https://doi.org/10.2166/ws.2017.121.

[37] V. Dutta, S. Chander, L. Srivastava, Public support for water supply improvements: Empirical evidence from unplanned settlements of Delhi, India, J. Environ. Dev. 14 (2005) 439–462. https://doi.org/10.1177/1070496505281841.

[38] M. Mukherjee, N. Chindarkar, J. Grönwall, Non-revenue water and cost recovery in urban India: The case of Bangalore, Water Policy. 17 (2015) 484–501. https://doi.org/10.2166/wp.2014.304.

[39] O. Bello, A.M. Abu-Mahfouz, Y. Hamam, P.R. Page, K.B. Adedeji, O. Piller, Solving management problems in water distribution networks: A survey of approaches and mathematical models, Water (Switzerland). 11 (2019). https://doi.org/10.3390/w11030562.

[40] M.S. Kadu, R.R. Dighade, Infrastructure Leakage Index and Challenges in Water Loss Management in Developing Countries, in: World Environ. Water Resour. Congr. 2015, American Society of Civil Engineers, Reston, VA, 2015: pp. 1322–1331. https://doi.org/10.1061/9780784479162.130.

[41] G. Mackintosh, C. Colvin, Failure of rural schemes in South Africa to provide potable water, Environ. Geol. 44 (2003) 101–105. https://doi.org/10.1007/s00254-002-0704-y.

[42] Y. Lalle, M. Fourati, L.C. Fourati, J.P. Barraca, Communication technologies for Smart Water Grid applications: Overview, opportunities, and research directions, Comput. Networks. 190 (2021) 107940. https://doi.org/10.1016/j.comnet.2021.107940.

[43] J.D. Miller, M. Hutchins, The impacts of urbanisation and climate change on urban flooding and urban water quality: A review of the evidence concerning the United Kingdom, J. Hydrol. Reg. Stud. 12 (2017) 345–362. https://doi.org/10.1016/j.ejrh.2017.06.006.

[44] C. Arrighi, F. Tarani, E. Vicario, F. Castelli, Flood impacts on a water distribution network, Nat. Hazards Earth Syst. Sci. 17 (2017) 2109–2123. https://doi.org/10.5194/nhess-17-2109-2017.

[45] A.M. Michalak, Study role of climate change in extreme threats to water quality, Nature. 535 (2016) 349–350. https://doi.org/10.1038/535349a.

[46] C. Hu, M. Li, D. Zeng, S. Guo, A survey on sensor placement for contamination detection in water distribution systems, Wirel. Networks. 24 (2018) 647–661. https://doi.org/10.1007/s11276-016-1358-0.

[47] S. Rathi, R. Gupta, Sensor placement methods for contamination detection in water distribution networks: A review, Procedia Eng. 89 (2014) 181–188. https://doi.org/10.1016/j.proeng.2014.11.175.

[48] N.P. Sonaje, M.G. Joshi, A review of modeling and application of water distribution networks, Int. J. Tech. Res. Appl. (2015) 1–6. https://doi.org/e-ISSN: 2320-8163.

[49] O.M. Awe, S.T.A. Okolie, O.S.I. Fayomi, Review of water distribution systems modelling and performance analysis softwares, in: J. Phys. Conf. Ser., 2019. https://doi.org/10.1088/1742-6596/1378/2/022067.

[50] L. Rossman, Epanet 2 users manual, in: Cincinnati US Environ. Prot. Agency Natl. Risk Manag. Res. Lab., 2000.

[51] C. Hu, L. Dai, X. Yan, W. Gong, X. Liu, L. Wang, Modified NSGA-III for sensor placement in water distribution system, Inf. Sci. (Ny). 509 (2020) 488–500. https://doi.org/10.1016/j.ins.2018.06.055.

[52] X. Xu, Y. Lu, S. Huang, Y. Xiao, W. Wang, Incremental sensor placement optimization on water network, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 8190 LNAI (2013) 467–482. https://doi.org/10.1007/978-3-642-40994-3_30.

[53] W. Hart, J. Berry, C. Phillips, J.-P. Watson, O. Avi, J. Uber, E. Salomons, The Battle of the Water Sensor Networks (BWSN): A design challenge for engineers and algorithms., J. Water Resour. Plan. Manag. 134 (2007) 556–568.

[54] A. Kumar, M.L. Kansal, A. Geeta, Identification of monitoring stations in water distribution system, J. Environ. Eng. (1997) 746–752.

[55] B.H. Lee, R.A. Deninger, Optimial locations of monitoring stations in water distribution system, J. Environ. Eng. 118 (1992) 4–16.

[56] J.W. Berry, W.E. Hart, C.A. Phillips, J. Watson, A Facility Location Approach to Sensor Placement Optimization, in: Water Distrib. Syst. Anal. Symp. 2006, American Society of Civil Engineers, Reston, VA, 2006: pp. 1–4. https://doi.org/10.1061/40941(247)111.

[57] A. Kumar, M.L. Kansal, G. Arora, A. Ostfeld, A. Kessler, Detecting Accidental Contaminations in Municipal Water Networks, J. Water Resour. Plan. Manag. 125 (1999) 308–310. https://doi.org/10.1061/(ASCE)0733-9496(1999)125:5(308).

[58] S. Rathi, R. Gupta, Monitoring stations in water distribution systems to detect contamination events, ISH J. Hydraul. Eng. 20 (2014) 142–150. https://doi.org/10.1080/09715010.2013.857470.

[59] A. Ostfeld, E. Salomons, Optimal layout of early warning detection stations for water distribution systems security, J. Water Resour. Plan. Manag. 130 (2004) 377–385. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:5(377).

[60] J.W. Berry, L. Fleischer, W.E. Hart, C.A. Phillips, J.P. Watson, Sensor placement in municipal water networks, J. Water Resour. Plan. Manag. 131 (2005) 237–243. https://doi.org/10.1061/(ASCE)0733-9496(2005)131:3(237).

[61] US Environmental Protection Agency, Threat Ensemble Vulnerability Assessment-Sensor Placement Optimization- Sensor Placement Optimization Tool (TEVA-SPOT) Graphical User Interface Threat Ensemble Vulnerability Assessment-Sensor Placement Optimization-Sensor Placement Optimization Tool, (2012). https://doi.org/10.13140/RG.2.2.18849.71521.

[62] O.S. Adedoja, Y. Hamam, B. Khalaf, R. Sadiku, A state-of-the-art review of an optimal sensor placement for contaminant warning system in a water distribution network, Urban Water J. 15 (2019) 985–1000. https://doi.org/10.1080/1573062X.2019.1597378.

[63] J.-P. Watson, W.E. Hart, H.J. Greenberg, C.A. Phillips, An Analysis of Multiple Contaminant Warning System Design Objectives for Sensor Placement Optimization in Water Distribution Networks, in: Harvey J. Greenb. A Leg. Bridg. Oper. Res. Comput., 2021: pp. 125–145. https://doi.org/10.1007/978-3-030-56429-2_7.

[64] L. Rosmman, US Environment Protection Agency, No Title, (2000).

[65] J.R. Chastain, Methodology for Locating Monitoring Stations to Detect Contamination in Potable Water Distribution Systems, J. Infrastruct. Syst. 12 (2006) 252–259. https://doi.org/10.1061/(ASCE)1076-0342(2006)12:4(252).

[66] J. Berry, W.E. Hart, C.A. Philips, J.G. Uber, J.P. Watson, Sensor placement in municipal water networks with temporal integer programming models, J. Water Resour. Plan. Manag. 132 (2006) 218–224. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:4(218).

[67] D. Eliades, M. Polycarpou, Iterative deepening of Pareto solutions in water sensor networks, in: 8th Annu. Water Distrib. Syst. Anal. Symp., 2006: pp. 1–19.

[68] A. Preis, A. Ostfeld, Multiobjective contaminant response modeling for water distribution systems security, J. Hydroinformatics. 10 (2008) 267–274. https://doi.org/10.2166/hydro.2008.061.

[69] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, C. Faloutsos, Efficient sensor placement optimization for securing large water distribution networks, J. Water Resour. Plan. Manag. 134 (2008) 516–526. https://doi.org/10.1061/(ASCE)0733-

9496(2008)134:6(516).

[70] M.M. Aral, J. Guan, M.L. Maslia, Optimal design of sensor placement in water distribution networks, J. Water Resour. Plan. Manag. 136 (2010) 5–18. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000001.

[71] J. Guan, M.M. Aral, M.L. Maslia, W.M. Grayman, Optimization model and algorithms for design of water sensor placement in water distribution dystems, J. Water Resour. Plan. Manag. (2006) 1–16.

[72] C.Y. Hu, D.J. Tian, C. Liu, X. Yan, Sensors placement in water distribution systems based on co-evolutionary optimization algorithm, Proc. 2015 1st Int. Conf. Ind. Networks Intell. Syst. INISCom 2015. (2015) 7–11. https://doi.org/10.4108/icst.iniscom.2015.258402.

[73] M.S. Marlim, D. Kang, Optimal water quality sensor placement by accounting for possible contamination events in water distribution networks, Water (Switzerland). 13 (2021). https://doi.org/10.3390/w13151999.

[74] J. Uber, R. Janke, R. Murray, P. Meyer, Greedy heuristic methods for locating water quality sensors in distribution systems, Proc. 2004 World Water Environmetal Resour. Congr. Crit. Transitions Water Environmetal Resour. Manag. (2004) 4859–4867. https://doi.org/10.1061/40737(2004)481.

[75] L. Sela, S. Amin, Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization, Adv. Eng. Informatics. 36 (2018) 55–63. https://doi.org/10.1016/j.aei.2018.02.004.

[76] H. Blockeel, K. Kersting, S. Nijssen, F. Zelezny, A Revised Publication Model for ECML PKDD, in: Eur. Conf. ECML PKDD 2013 Prague, Czech Republic, Sept. 23-27, 2013 Proceedings, Part III, 2013: pp. 467–482. http://arxiv.org/abs/1207.6324.

[77] K.A. Klise, C.A. Phillips, R.J. Janke, Two-Tiered Sensor Placement for Large Water Distribution Network Models, J. Infrastruct. Syst. 19 (2013) 465–473. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000156.

[78] J. Xu, P.S. Fischbeck, M.J. Small, J.M. VanBriesen, E. Casman, Identifying Sets of Key Nodes for Placing Sensors in Dynamic Water Distribution Networks, J. Water Resour. Plan. Manag. 134 (2008) 378–385. https://doi.org/10.1061/(asce)0733-9496(2008)134:4(378).

[79] A. Di Nardo, C. Giudicianni, R. Greco, M. Herrera, G.F. Santonastaso, A. Scala, Sensor Placement in Water Distribution Networks based on Spectral Algorithms, in: HIC 2018. 13th Int. Conf. Hydroinformatics, 2018: pp. 593–584. https://doi.org/10.29007/whzr.

[80] J. Sun, R. Wang, Sensors Layout Optimization Based on Cluster Analysis in Water Supply Network, DEStech Trans. Eng. Technol. Res. (2017) 1011–1015. https://doi.org/10.12783/dtetr/iceta2016/7125.

[81] K. Diao, W. Rauch, Controllability analysis as a pre-selection method for sensor placement in water distribution systems, Water Res. 47 (2013) 6097–6108. https://doi.org/10.1016/j.watres.2013.07.026.

[82] M.S. Khorshidi, M.R. Nikoo, M. Sadegh, Optimal and objective placement of sensors in water distribution systems using information theory, Water Res. 143 (2018) 218–228. https://doi.org/10.1016/j.watres.2018.06.050.

[83] P. Mandel, M. Maurel, D. Chenu, Better understanding of water quality evolution in water distribution networks using data clustering, Water Res. 87 (2015) 69–78. https://doi.org/10.1016/j.watres.2015.08.061.

[84] United Nations, Progress on Ambient Water Quality, 2021. https://www.unwater.org/sites/default/files/app/uploads/2021/09/SDG6_Indicator_Report_632_Progress-on-Ambient-Water-Quality_2021_EN.pdf.

[85] S. Rathi, R. Gupta, A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks, KSCE J. Civ. Eng. 20 (2016) 597–608. https://doi.org/10.1007/s12205-015-0024-x.

[86] S. Rathi, R. Gupta, Optimal sensor locations for contamination detection in pressure-deficient water distribution networks using genetic algorithm, Urban Water J. 14 (2017) 160–172. https://doi.org/10.1080/1573062X.2015.1080736.

[87] S. Rathi, R. Gupta, S. Kamble, A. Sargaonkar, Risk based analysis for contamination event selection and optimal sensor placement for intermittent water distribution network security, Water Resour. Manag. 30 (2016) 2671–2685. https://doi.org/10.1007/s11269-016-1309-7.

[88] M.A. Khaksar Fasaee, S. Monghasemi, M.R. Nikoo, M.E. Shafiee, E.Z. Berglund, P.H. Bakhtiari, A K-Sensor correlation-based evolutionary optimization algorithm to cluster contamination events and place sensors in water distribution systems, J. Clean. Prod. 319 (2021) 128763. https://doi.org/10.1016/j.jclepro.2021.128763.

[89] H. Shen, E. McBean, Pareto optimality for sensor placements in a water distribution system, J. Water Resour. Plan. Manag. 137 (2011) 243–248. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000111.

[90] M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: A survey, Ad Hoc Networks. 6 (2008) 621–655. https://doi.org/10.1016/j.adhoc.2007.05.003.

[91] N. Sankary, A. Ostfeld, F. Asce, Inline Mobile Sensors for Contaminant Early Warning Enhancement in Water Distribution Systems, 143 (2017) 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000732.

[92] A. Shahmirnoori, M. Saadatpour, A. Rasekh, Using mobile and fixed sensors for optimal monitoring of water distribution network under dynamic water quality simulations, Sustain. Cities Soc. 82 (2022) 103875. https://doi.org/10.1016/j.scs.2022.103875.

[93] T.S. Fu, A. Ghosh, E.A. Johnson, B. Krishnamachari, Energy-efficient deployment strategies in structural health monitoring using wireless sensor networks, Struct. Control Heal. Monit. 20 (2013) 971–986. https://doi.org/10.1002/stc.1510.

[94] M. Abdulkarem, K. Samsudin, F.Z. Rokhani, M.F. A Rasid, Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction, Struct. Heal. Monit. 19 (2020) 693–735. https://doi.org/10.1177/1475921719854528.

[95] M. Iqbal, M. Naeem, A. Anpalagan, N.N. Qadri, M. Imran, Multi-objective optimization in sensor networks: Optimization classification, applications and solution approaches, Comput. Networks. 99 (2016) 134–161. https://doi.org/10.1016/j.comnet.2016.01.015.

[96] M.A. Benatia, M. Sahnoun, D. Baudry, A. Louis, A. El-Hami, B. Mazari, Multi-

Objective WSN Deployment Using Genetic Algorithms Under Cost, Coverage, and Connectivity Constraints, Wirel. Pers. Commun. 94 (2017) 2739–2768. https://doi.org/10.1007/s11277-017-3974-0.

[97] A.J. Perez, M-SPOT : A Hybrid Multiobjective Evolutionary Algorithm for Node Placement in Wireless Sensor Networks, 2018 32nd Int. Conf. Adv. Inf. Netw. Appl. Work. (2018) 264–269. https://doi.org/10.1109/WAINA.2018.00096.

[98] X. Li, H. Cai, G. Liu, K. Lu, Base Station Positioning in Single-Tiered Wireless Sensor Networks, (2019) 7–12. https://doi.org/10.1109/PDCAT.2019.00013.

[99] G. Zhou, T. Yi, M. Xie, H. Li, J. Xu, Optimal Wireless Sensor Placement in Structural Health Monitoring Emphasizing Information Effectiveness and Network Performance, J. Aerosp. Eng. 34 (2021) 1–13. https://doi.org/10.1061/(ASCE)AS.1943-5525.0001226.

[100] E.F. Flushing, G.A. Di Caro, A flow-based optimization model for throughput-oriented relay node placement in wireless sensor networks, in: Proc. 28th Annu. ACM Symp. Appl. Comput., ACM, New York, NY, USA, 2013: pp. 632–639. https://doi.org/10.1145/2480362.2480482.

[101] A.L. Custódio, J.F.A. Madeira, A.I.F. Vaz, L.N. Vicente, Direct multisearch for multiobjective optimization, SIAM J. Optim. 21 (2011) 1109–1140. https://doi.org/10.1137/10079731X.

[102] Y. Cao, B.J. Smucker, T.J. Robinson, On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design, J. Stat. Plan. Inference. 160 (2015) 60–74. https://doi.org/10.1016/j.jspi.2014.12.004.

[103] Y. Wang, C. Hu, Y. Tseng, S. Member, Efficient Placement and Dispatch of Sensors in a Wireless Sensor Network, 7 (2008) 262–274.

[104] G. Mavrotas, Effective implementation of the e -constraint method in Multi-Objective Mathematical Programming problems, Appl. Math. Comput. 213 (2009) 455–465. https://doi.org/10.1016/j.amc.2009.03.037.

Appendix

A1. Design of contamination events

A contamination event is defined by the "intrusion of a toxic substance into the network in a certain quantity at a certain location in the network". This leads to three characteristic features which in various combinations generate the complete set of contamination events. These features are:

i. Contamination nodes: The EPANET or any other hydraulic simulation tool allows the addition of contaminants at the nodes. But, given a network, contaminant intrusion can happen at any location in the network. And to cover the entire network, nodes will have to be placed at each point of the network. This increases the computation time drastically for both hydraulic simulations and optimization. Thus, for simplification, the nodes can be reduced down such that the performance of the reduced network is similar to the exhaustively described network.

ii. Contamination time and duration: This characteristic refers to the time of intrusion and the duration for which intrusion takes place. The time of intrusion can have any value from 0000 hours to 2359 hours, and a full-scale analysis with increment in seconds leads to about 24*60*60 = 86400 such scenarios. This can be simplified by increasing the time gap to 5 minutes resulting in 24*12 = 288 scenarios. This assumption is reasonable as it has been observed that the hydraulics of a WDN do not vary drastically in such short time intervals.

iii. Contamination concentration: The concentration defines the amount of contaminant introduced into the distribution system. The added contaminant is assumed to be non-interacting.

Based on the various combinations of the above-given set of characteristics, a set of contamination events are designed. The total set of contamination events can be calculated as,

$$No.of\ contamination\ events = no.\ of\ nodes \times no.\ of\ times\ intrusion\ occuring\ in\ a\ day \times no.of\ contamination\ concentrations\ considered \quad \text{(A1.1)}$$

Once the contamination events are defined, for each event the WDN is simulated in EPANET (or any other hydraulic software) and the effects of the contamination are observed at each of the nodes. The effects at each of the nodes could be as simple as a binary variable, denoting 0

for non-detection and 1 for detection. Detection refers to the concentration of water leaving that particular node is greater than zero. Similarly, the time from the intrusion of contamination to the time when concentration is more than zero is registered, yielding the detection time for each event at each node. All these nodal observations are then provided as a matrix that encompasses the full set of contamination events and its effects, termed as pollution matrix. The pollution matrix is then used to estimate the effective of the sensor placement designs.

A2. Comparison of Pareto fronts from metaheuristic method (Chapter 4) and MILP formulation (Chapter 5)

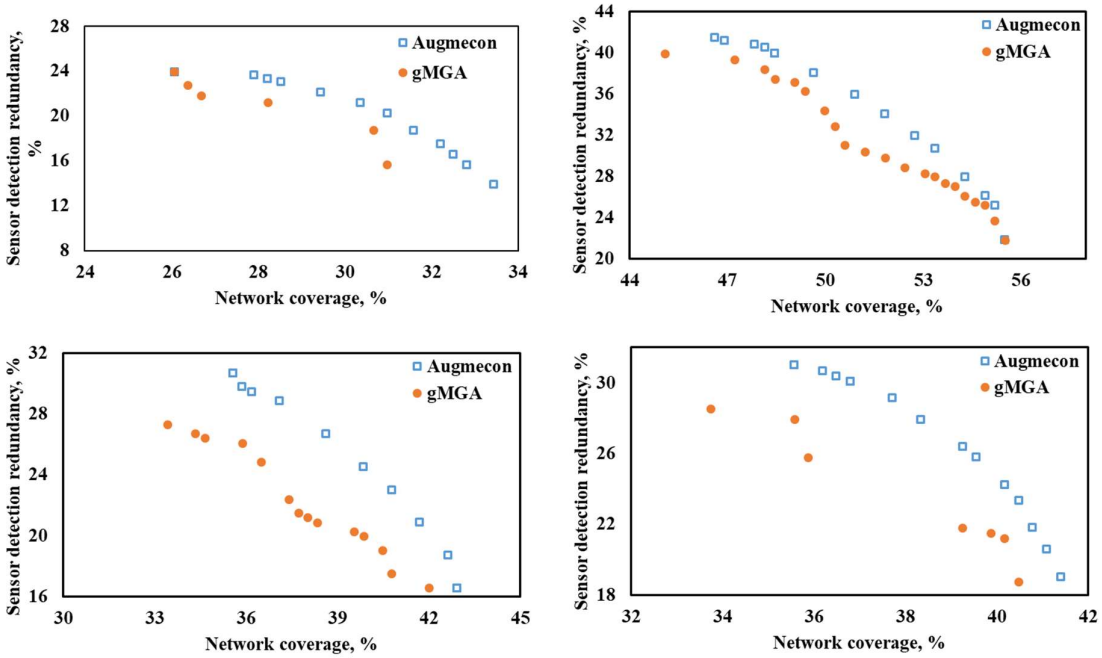


Figure A. 1Comparison of results from metaheuristics (chapter 4) and MILP (chapter 5) for a) Ns=10 TR = 0.25 km; a) Ns=10 TR = 1 km; c) Ns=20 TR = 0.25 km; d)a) Ns=20 TR = 1 km

Research output

- ***Journal publications***

  D.K. Gautam, P. Kotecha, S. Subbiah, Efficient k-means clustering and greedy selection-based reduction of nodal search space for optimization of sensor placement in the water distribution networks., Water Res. 220 (2022) 118666. https://doi.org/10.1016/j.watres.2022.118666.

- ***Conferences Presentation***

  D.K. Gautam, P. Kotecha, S. Subbiah, Simple and efficient placement of wireless sensors in data-deficient water distribution networks, Ninth International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 2022), Mykonos,2022.

- ***Journal article under review:***

  Multi-objective optimization enhanced with greedy initialization for placement of wireless sensor networks in water distribution networks.

- ***Journal article in writing phase:***

  MILP formulation for optimal placement of wireless water quality sensors in water distribution networks.