EXPERIMENTAL ANALYSIS OF CYBER PHYSICAL SYSTEM PERSPECTIVE FOR WATER MONITORING AND DISTRIBUTION

Thesis

Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by
ABHISHEK M B

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING,
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,
SURATHKAL, MANGALORE -575025

JANUARY, 2020

DECLARATION

I hereby ***declare*** that the research Thesis entitled **Experimental Analysis of Cyber Physical System Perspective for Water Monitoring and Distribution** which is being submitted to the ***National Institute of Technology Karnataka, Surathkal*** in partial fulfillment of the requirement for the award of the Degree of ***Doctor of Philosophy*** in **Department of Electronics and Communication Engineering** is a ***bonafide report of the research work carried out by me***. The material contained in this research Thesis has not been submitted to any University or Institution for the award of any degree.

Abhishek M B,
145082EC14F01,
Department of Electronics and
Communication Engineering,

Place: NITK-Surathkal.
Date:

CERTIFICATE

This is to certify that the Research Thesis entitled **Experimental Analysis of Cyber Physical System Perspective for Water Monitoring and Distribution** submitted by **Abhishek M B** (Register Number: 145082EC14F01) as the record of the research work carried out by him, is accepted as the *Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy**.

Research Guide
Dr N. Shekar V. Shet,
Professor, Department of ECE

Chairman-DRPC
(Signature with Date and Seal)

Acknowledgements

I want to sincerely express my gratitude to all the staffs members of ECE department for their continued support in my research work as well as in my personality development. Successful completion of my research work is attributed to the dedication and coordination by a number of people. It is really a very proud moment for me now to acknowledge and thank the people who helped me every way.

I feel lucky to have Dr N. Shekar V. Shet, Professor, Department of ECE, as my research guide. Without his guidance, this research would not be possible at all. I am deeply indebted to him and I give my sincere respect for his encouragement throughout the research. His constant appraisal and suggestions have made my work better. His parental care and guidance has been a driving force. I'm very much inspired by his nature, the knowledge he possesses and the way he serves the knowledge to his students. He will always stay in my life as a great role model.

I want to offer my special and heart-filled thanks to Prof. Muralidhar Kulkarni, Professor, Department of ECE, NITK for his technical and philosophical guidance. I am very honored to get his encouragement and full support. He is one of the inspirational person in my life and also a well-wisher.

I extend my special and heartfulness thanks to Prof. U. Shripathi Acharya, Professor, Department of ECE, NITK. He is one of my most favorite professor and the way he supported me in my personal life to reach my carrier goal is tremendously appreciable. His encouragement and positive words given a lot of impact on my research work and also on my personal life.

I also wish to thank Professor Annapa, and Professor Shyamlal, my RPAC member. Professor Annappa and Shyamlal have been a part of great advisors in my research work. Their enthusiasm and vision on my research work will be an inspiration. It was a learning experience and also very inspiring to watch, build and work for a successful research program.I want to thank my RPAC members, whose feedback, technical suggestions and conversations, helped me in many ways. I would like to thank Dr Mohit P. Tahiliani, Department of CSE, NITK who helped me to develop a keen interest in my work and supported me throughout the process.

Professor M.S.Bhat has provided me support, encouragement and his feedback has enhanced my work in many ways. I thank him for everything he has invested in my research especially his time. . I would like to thank Professor John D'Souza and Professor Sumam David S. department of ECE, NITK for lending their helping hands throughout my research work. Professor and Head, Department of ECE, NITK has been providing me with the resources and support for carrying out my research work and I am extremely grateful for the same. I have been enlightened a lot by his extensive knowledge.

I am indeed grateful to all Faculties of Department of ECE, NITK for there guidance and motivation. There knowledge has always inspired me to do my best for this research work and I am very grateful.

It is my duty to thank all supporting staff of the department, especially, Sri Subramanya Karanth, Smt. Amitha P. Amin, Sri Sanjeeva Poojary, Sri Tilak, Sri Ratheesh, Smt. Prabha for their support throughout my research work. The staff members of the department of ECE, NITK have been very supportive in the entire process of carrying out my research and I am very grateful.

My colleagues Nagaraj, Shreyas, Vasudev Reddy , Jayaram, Deepa, Ragesh, Raghu, Yajunath Kaliyath, Chandrashaker Balure, Goutham Simha G.D, K Shareef Babu, Prasad, Ashish Patil, Jhnanesh , Sarvesh , Raghavendra MA N S have been very supportive and caring. These are all supportive pillars for me, in my entire research work. They made my workplace feel like home. Without them the entire process would have been so much harder.

Needless to say, the support and sacrifice of my Father M.S.Basavaraju and My Mother Savitha ML during the research period was inevitable for me. I want to thank Muniyappa Uncle, without whom I would have never come this far and Dr. Girish Chandra for his consistent help with the research. Finally, This would not have been possible without the love and affection of my sister Yashaswini.

Last but not the least all my friends and well wishers had been of great support to me in my research work and also most importantly in my complete life. Without my friends and well wishers, I'm nothing. Thank you and love you all my friends and well wishers from the bottom of my heart for your journey in my life.

Thank you each and everyone for making this possible in a great way!

Abstract

Water plays an imperative role in proper functioning of the Earth's ecosystems, and basically all individual activities, such as cultivation, manufacturing, transportation, and energy production.

Abdul Kalam quotes that Water will become a very coveted treasure, more precious than gold and diamonds by 2050. In todays scenario pure potable water resources are depleting quite fast and human beings are ready to go to war to acquire hold on this natural resource. Hence, safeguarding(conservation) of this natural resources is given prime importance in my research work. The main cause of this is the rate at which human population is growing. If the rate of a growing population is compared with that of the potable water available on the earth, it logically concludes that we have to protect and preserve the precious water resource. Water conservation refers to the control, protection and development of the water resources both on ground and surface in order to prevent pollution. As water is the primary resource for life sustainability, it's our responsibility to promote water management by conserving and using water efficiently. Additionally, it is our responsibility to save the water for our next generation and educate them to keep the water clean and safe for their upcoming usage.

Physical water monitoring and management is a tedious and time consuming job. Alternately water management can be done using the available technology.

Remarkable advancements in technology have so far addressed the water sustainability challenges using valuable tools. Key technologies include sensing technology, wireless communications and networking, hydrodynamic modeling, data analysis and control enabled intelligently wireless networked water cyber-physical systems (CPS) with embedded sensors, processors, and actuators that can sense and interact with the water environment. Unmanned water surveillance is possible with the help of communication technology. Real-time surveillance problem require intelligent organizations, planning, infrastructure design, water supply system and management. Smart water management is a good instance of real time CPS. For realizing this instantiation, calls a systematic framework together

with the actual implementation of the associated modules is to be devised. In Cyber physical system there are 4 general working blocks: Monitoring unit, Networking unit, Computation unit and Actuation unit.

In monitoring unit we have presented a systematic approach of customizing the existing water pipeline infrastructure using wireless technology. This is done to explore the possibility of wireless monitoring of the water distribution through a storage tank and later bring out the aspects of Cyber Physical system (CPS). To monitor the amount of water flows in storage tanks, we have carried out experiment at NITK, Surathkal, Mangalore, India. A set up of RF transceiver modules interfaced with water meter has been installed to get the water inflow and outflow details for each pipe. With respect to sampling time in the monitoring unit, a set up of another RF transceiver interfaced with aggregator and DTU will receive the water inflow and outflow details. This is then displayed on the web server.

The networking unit includes the operation of software requirement and communication between data transfer unit (DTU) and aggregator unit (AU). The DTU and AU are present in the monitor unit to analyze the amount of water flowing in and out of storage tanks. The resultant data is available on a web server. The data aims to adapt ways and methods to overcome the improper handling and care of water flow in and out of storage tanks.

In Computation unit, time series data of water flow rate has been used for forecasting. The water flow rate data is obtained at a sampling interval of 15 minutes and is classified into daily and monthly data. For time series data based water flow rate forecasting, we have considered and compared most prominent predictive models that are used in CPS. The forecasting models used are Simple Moving Average (SMA), Holt-Winters (HW), AutoRegressive Integrated Moving Average (ARIMA), Neural Networks (NN) and MultiLayer Perceptron (MLP). Performance metrics used are the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The prediction obtained for daily data is observed to be more precise as compared to that of monthly data. In the case of instantaneous and cross-validation (10 fold), HW model and ARIMA models are observed to predict with low RMSE, low MAE and low

MAPE, as compared to other models. In this work we have proposed a new hybrid model that is an ensemble of ARIMA and HW model. Simulation results indicate that the hybrid model is more accurate when compared with its basis time series model. This hybrid model can be used to conserve and distribute water when needed most.
Real-time monitoring of water flow rates information helps us to conserve water when needed most hence we consider the first impediment in forecasting that is; handling missing data in the real-time monitoring system using the different imputation techniques. The performance is evaluated using traditional methods like RMSE, MAE and MAPE. Here two types of data missing are considered: 1) missing at random and 2) missing at burst. The validation is carried out by computing for a range of 'data missing', varying from 0% to 60% using different imputation techniques. This suggests that the proposed technique helps us to take necessary action to complete the decision-making activities (Computation unit) in CPS.

Keywords: Cyber Physical System; Water management; Monitoring Unit; Networking Unit; Computation Unit; Forecasting; Imputation.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Expansion
CPS	Cyber Physical System
RF	Radio Frequency
DTU	Data Transfer Unit
MIU	Meter Interface Unit
ACE	Affordability, Convenience, and Efficiency
TCP/IP	Transmission Control Protocol/Internet Protocol
GPRS	General Packet Radio Services
GPS	Global Positioning System
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ARIMA	AutoRegressive Integrated Moving Average
SMA	Simple Moving Average
HW	Holt-Winter's
NN	Neural Network
MLP	Multi Layer perceptron
K-NN	K- Nearest Neighbour
EM	Expectation Maximization
MC	MAtrix Completion
IOT	Internet of Things
WSN	Wireless Sensor Network

Chapter 1

INTRODUCTION

Outline of the chapter

- **Section 1.1** runs through a complete framework of Cyber Physical Systems i.e., its definition, workflow, structure and applications.
- **Section 1.2** focuses on an all-inclusive comprehensive description regarding importance of water, causes of water scarcity, numerous methods in monitoring water wastage and role of CPS as a protagonist in water management. It also provides a wide ranging literature review of numerous studies associated with the domain.
- **Section 1.3** briefs up on the origin, whys and wherefores which lead to this research study.
- **Section 1.4** discusses the clear description of the problem(s), involves a vision, declaration of the issue, and technique used to fix the problem.
- **Section 1.5** delivers a manifesto of our research work used to solve the water scarcity issue.

1.1 CYBER PHYSICAL SYSTEM

1.1.1 Definition

Cyber Physical System is defined as an amalgamation of computer-aid and software components with both electronic and mechanical elements, which all together has access to each other by means of data infrastructures interacting through internet ([UC Berkeley]). It can also be defined as an incorporation of physical protocols with computation whose behavior is delineated by physical and cyber components equally. Cyber-Physical Systems are a mechanism intended to be foreseen and handled through computer-based algorithms. Entire system of CPS perceptibly is incorporated with by means of Internet; therefore it includes the term 'cyber' in it. In a CPS, both software and physical components function from side to side through diverse temporal and spatial scales. A CPS is meant to emphasize on essential problematic issues such as adjoining engineering ethnicities of both the physical and cyber domains.

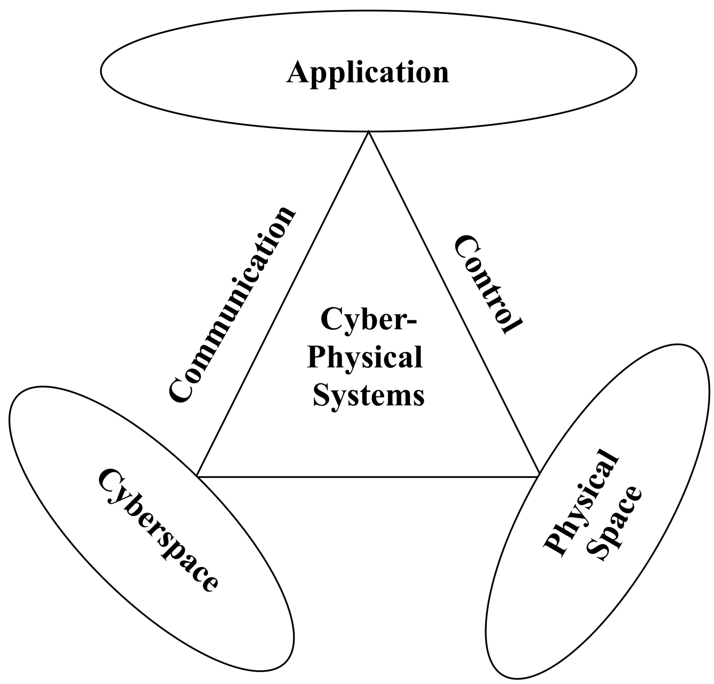


Figure 1.1: Definition of CPS (2b1 consulting (2018))

CPS's are adamant to design, model and analysis in comparison with homogeneous systems. Considering Cyber Physical Systems barely on a physical forefront is not adequate, its interaction sphere must also be inferred. CPS's by nature are heterogeneous. They unite communication, computation and physical dynamics too. Networks and Embedded systems examine and direct the physical protocols, typically through feedback systems wherein physical protocols affect computations or/and contrariwise. Henceforth, it is a fact that CPS is regarded as the coming together and not the amalgamation, of cyber and corporal zones. CPS segments eco-friendly facts and figures locally and progresses supply consumption.

Cyber Physical Systems are intensely related to present day prominent terms such as Industry 4.0, Internet of Things (IoT), Machine-to-Machine (M2M), the Internet of Everything, T Sensors (trillion sensors) etc.,

1.1.2 Cyberspace to Cybernetics and Cyber Physical System

Helen Gill from the National Science Foundation, United States introduced the word 'Cyber – Physical Systems' in 2006. Sir William Gibson attributed the term 'cyberspace' that was also used in his novel Neuromancer. Rootstocks of Cyber Physical Systems are profound and bygone. An American Mathematician named (Wiener (1965)) stated that 'Cybernetics' is the origin of in cooperation 'Cyber –Physical Systems' and 'Cyberspace'. Wiener formulated the term 'Cybernetics' from the Greek word kybernetes, connotation governor, pilot, helmsman or rudder. The metaphor of which is pertinent for control systems.

Norbert Wiener also introduced and launched technologies for automated aiming and firing of anti-aircraft guns and also vitally bestowed in the domain of control systems theory development. In spite of the mechanisms implied by him, it did not involve digital computers, nevertheless the principles implied in it are analogous to present day's computer-based feedback control systems. Wiener's control logic was in point of fact a computation, although one of them implied with mechanical parts and analog circuits. Henceforth, Cybernetics is the concurrence of communication, physical processes and computation.

By these advances, at present and in the coming years, there will be an augmented implication and usage of Cyber-Physical Systems. Cyber-Physical system is defined by Baheti and Gill as: "Systems that integrate cyber world with the physical world (van Lier (2013)). Computational and physical components of CPS are decidedly interconnected and synchronized to work together commendably. The amalgamation of cybernetics with physical objects in environment shows up as a fundamental role in the development of CPS. This amalgamation leads to a novel and improved physical system or whole as an object, in which the properties of the whole will be more than the sum of the discrete components or/and parts. Through CPS a new-fangled peer group of systems with assimilated computational and physical capabilities comes into existence, which can intermingle and act together with humans over numerous modalities. Capability to intermingle and increase competencies of the physical world by communication, computation, and control is a crucial enabler in forthcoming technology expansions.

1.1.3 Simple Structure of CPS

A simple assembly of Cyber Physical System is demonstrated in Figure 1.2. It comprises of three prominent sections i.e., Physical Plant, Network Fabric and computational platforms (Lee and Seshia (2016)).

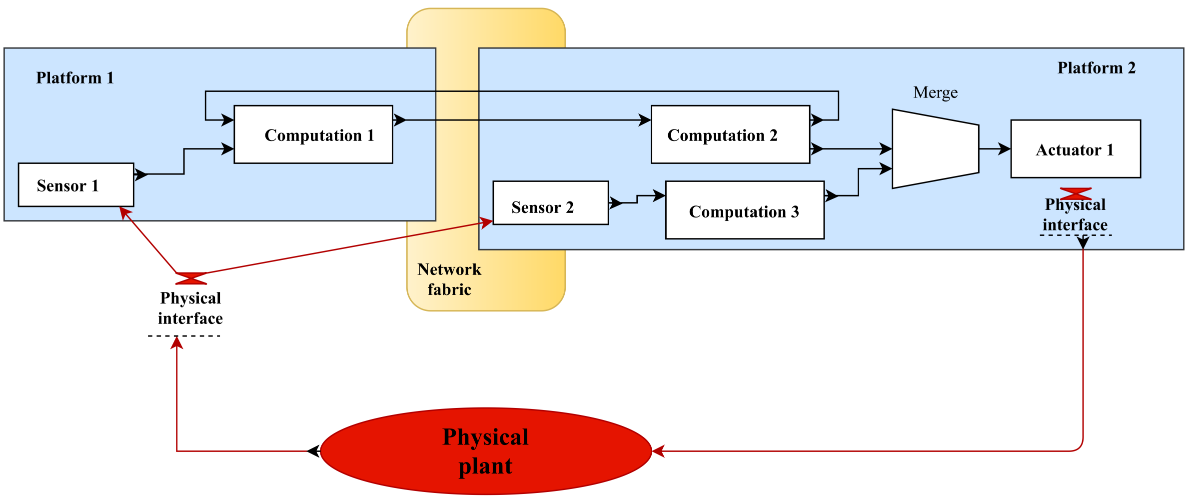


Figure 1.2: Simple Structure of Cyber Physical System (Lee and Seshia (2016))

- The Physical Plant involves human operators, chemical or biological protocols or mechanical components. This section does not include digital networks.
- Secondly, Network fabric bestows techniques and processes involved in communication protocols. Finally, computational platforms are entailed with numerous computers and operating systems, actuators, sensors etc.,
- Computational platforms and Network fabric altogether make up the “cyber” part of Cyber-Physical System.

Dual networked platforms individually with its own sensor and/or actuator is also depicted. An actuator’s activity habitually affects the sensors data delivered by means of a physical plant. As depicted, Platform 2 regulates the physical plant by way of Actuator 1. Physical plant progressions are measured via Sensor 2. The section Computation 2 implies a control law, which concludes and mainly centered on the sensor data whatever commands to be imposed and delivered to the actuator. This loop is termed as 'feedback control loop'. Platform 1 knocks off added surplus measurements through Sensor 1, and directs messages to Platform 2 using network fabric. Computation 3 comprehends an additional control law, amalgamated with that of Computation 2, conceivably prompting to pre-empty it.

1.1.4 CPS versus MANET, WSN, M2M

In (Wu *et al.* (2011)) a review learning is prepared on WSN and MANET and also the way they have driven in the direction of CPS's progress.(Wan *et al.* (2013)) explored features of M2M and CPS, and also indicated correlations between them. Characteristics of M2M architecture, design challenges and application of M2M have been evaluated. Table 1.1 portrays a qualitative assessment between CPS, WSN, M2M and MANET.

Table 1.1: Comparison between MANET, M2M, WSN and CPS

Networks	**Features**	MANET	**M2M**	**WSN**	**CPS**
Network formation	Random deployment	Yes	Yes	Yes	Yes
	Dynamic topology	Yes	Yes	No	Yes
	Time-vary deployment	No	Yes	No	Yes
	Interconnection among multiple network	No	Yes	No	Yes
	Internet supporting network	No	No	Yes	Yes
Communication pattern	Query Response Flows	Yes	Yes	Yes	Yes
	Arbitary Communication Flows	Yes	Yes	No	Yes
	Cross network Communication Flows	No	No	No	Yes
	Cross domain Communication Flows	No	No	No	Yes
	Deterministic Delay communication Flows	No	No	No	Yes
Power management network	Opportunistic sleep	Yes	Yes	Yes	Yes
	Multiple sleep modes of nodes	No	Yes	No	Yes
	Power management techniques for both sensors and central serves	No	Yes	No	Yes
Network connectivity and coverage	Connectivity	Yes	Yes	Yes	Yes
	Coverage	No	Yes	Yes	Yes
	Heterogeneous coverage	No	No	No	Yes
Knowledge mining	Data mining and data base management	No	No	Yes	Yes
	Multi-domain data sources	No	No	Yes	Yes
	Data privacy and security	No	No	No	Yes
Quality of services	Networking QOS	Yes	No	Yes	Yes
	Multiple data resolution	No	No	Yes	Yes

1.1.5 CPS Work-Flow

An overall workflow of Cyber Physical System is written off as four key phases and illustrated in Figure 1.3 (Wang *et al.* (2015), Wang *et al.* (2010)) :

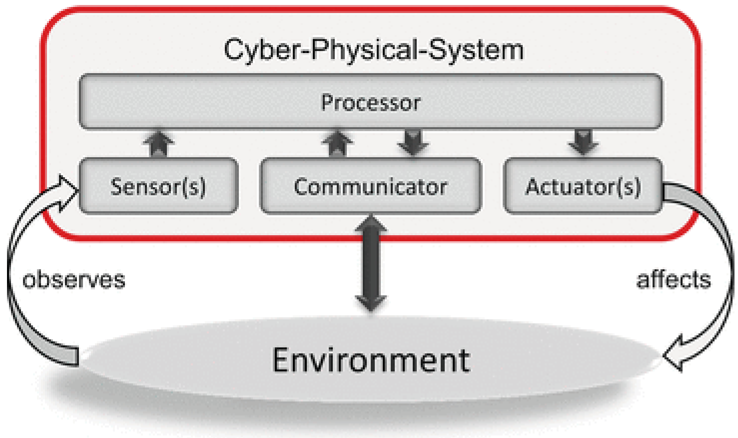


Figure 1.3: Workflow of Cyber Physical System (Wiesner and Thoben (2017))

- Monitoring: It is a primary phase in Cyber Physical Systems. In our research study, physical system is supervised from end to end by analyzing the historical data collected. This data will then be employed by the CPS so that every single procedure involved is guaranteed for future use. Any imperfection perceived in the physical system must be essentially rectified by Cyber Physical System.

- Networking: In this phase, data is to be collected. In a Cyber Physical System, numerous RF modules, sensors and/or wireless devices can be implied. Sensors and/or RF modules produce data in real time. Later the data produced must be accumulated altogether for additional processing. In the meantime, numerous applications also intermingle and communicate through the network.

- Computing: In this phase, data collected is synopsized as well as evaluated during supervising review and testing of physical system's conditions. Actions to be accomplished must be implemented running through the conditions of the physical system. For instance, in our research study we use data predictive model so as to forecast valve controller in storage tank as per the algorithm intended in lines with historic data for forthcoming operations.

- Actuation: In this phase, actions planned in the computing stage will be brought into effect. Actuation will actuate numerous actions akin as physical processes control; correction in cyber's behaviour etc., as an example to be considered, in our research study action is to regulate and govern the valve in storage tank as per the historic data.

1.1.6 Applications of CPS

Cyber Physical Systems has its applications in various arenas such as medical monitoring, sovereign automobile systems, process control systems, robotics systems, instinctive pilot avionics, smart grids, traffic logistics systems etc., Numerous applications of CPS are illustrated in Figure 1.4

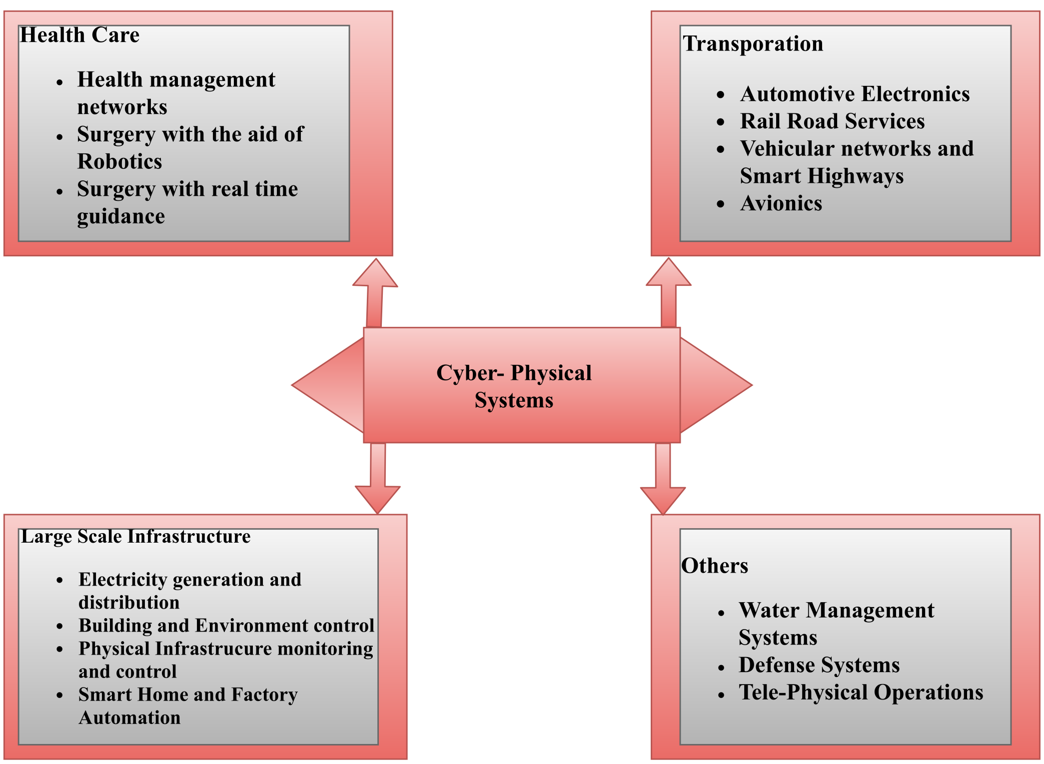


Figure 1.4: Applications of Cyber Physical Systems

1.2 WATER MANAGEMENT

1.2.1 Importance of Water

The surface of planet Earth is in great part composed of water, in the form of oceans and rivers, but it is also present in the cells of every single living organism. Less than 3% of the Earth's water is freshwater. Most of the freshwater on Earth can be found in icecaps and glaciers (69%) and in groundwater (30%). Rivers, lakes, and swamps only represent a small portion, i.e. only 0.3% of the planet's freshwater reserves (water science school). Interestingly, lake Baikal in Siberia, which is the world's largest and deepest lake, holds the largest solitary volume of freshwater in the world (UNESCO (2006)). Water plays a very substantial role in the world's economy. Approximately 70% of freshwater in convention is subsidized for agriculture (Baroni *et al.* (2007)) . In numerous parts of the globe, fishing in both fresh and salt water bodies is a notable source of income. A great number of long-distance trade merchandises - for instance oil and natural gas - are all transported using water. Water is also essential in several sports and recreational activities such as swimming, surfing, sport fishing, boat racing, diving etc. Our entire society is built around water and its cycles and yet, as we will see, it is becoming more and more scarce.

The requirement for Water conservation: According to the water research data reports Geographic, Igor (1993) as shown in Figure 1.5 here are some of the facts:

- 97% of the water on the earth is salty, which cannot be utilized for drinking.
- On Earth, 3% of water is only clean, and just 0.5% is accessible for drinking.
- The other 2.5% of clean water is secured ice tops, icy masses, the air, soil, or under the world's surface, or is unreasonably contaminated for utilization.

1.2.2 Water Scarcity

Water is a vital resource for the planet: without water, there would be no life on earth. While it used to be an abundant natural resource, it is quickly becoming a treasured commodity because of overuse, wastage and drought. Climate change is also playing a major role in making our water resources more scarce and unpredictable. So, today more than ever, water needs to be acknowledged as both a unique and valuable resource, especially when considering our rapidly increasing population, as well as the

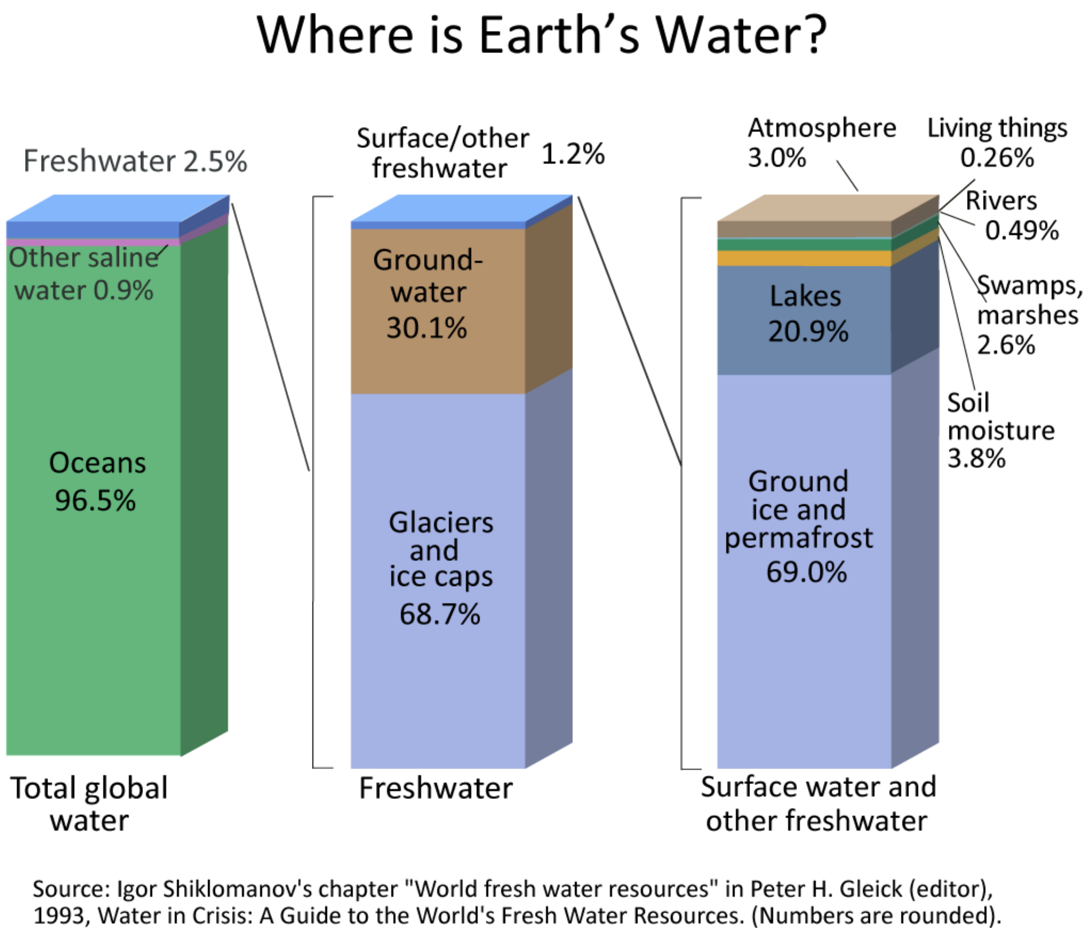


Figure 1.5: A Graphical distribution of water locations on earth (water science school, Bralower)

pollution associated with it. As depicted in Figure. 1.6 , there is massive water scarcity in India due to the difficulty in fulfilling basic water requirement (Institute and Centre (2016) ,Ramachandran) . Continuous water supply is a major challenge due to seasonal fluctuations at the source. There is almost no water service provider in India capable of delivering a continuous water supply. Why is that and how can water management be improved? This is what we will focus on next.

1.2.3 Water Management

As we have seen, the impeding threat of water scarcity makes it vital to optimize our utilization of water resources now - and for the future. We need to critically

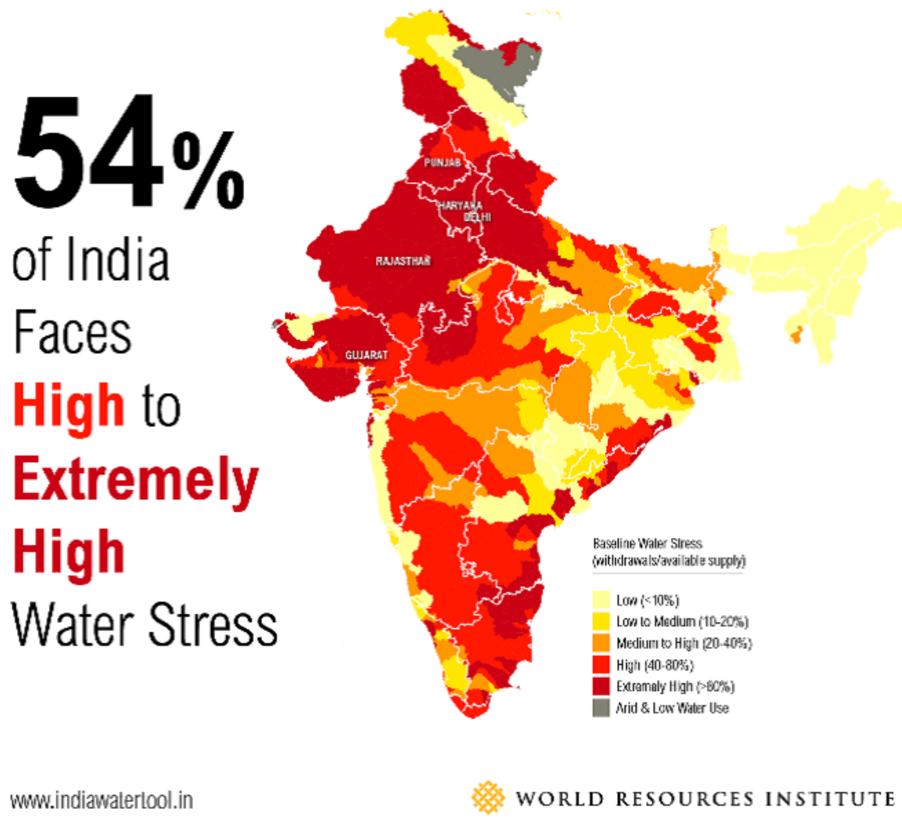


Figure 1.6: 54 % of India's entire expanse fronting extraordinary to tremendously high water pressure. Source: www.wri.org/ http://www.indiawatertool.in/

look at how we prepare, increase, dispense and handle it. Regulations are required to regulate the volume of water wastage in transfers between the water source and the end-consumer.

Water resource management aims at regulating the usage of water via regulations and policies. It is an important lever for action considering issues such as cleaning of bodies of water, future developments, preserving current water resources from pollution, preventing water disputes etc. Also there is concern about the availability, quality and management of water. So, preferably, water resource management planning will respect to all challenging and conflicting aspects of water management to effectively distribute water from a good source to the end-users. In that regard, widespread hydrological studies are essential to improve the management and preservation of water. As we have mentioned, continuous water supply is hard to achieve due to seasonal fluctuations at the source. And yet continuous supply is one of the condition for

the supplier to be able to build a proper user demand. Indeed, there is a direct correlation between the quality of water management and demand (the number of end-consumers). Below a certain threshold of demand, the quality of water management drops significantly (Ramachandran). Management of water is done both on the supply side and the demand side. Supply management covers the management of water resources and water transit. Demand management is where the water service provider tries to reduce water consumption by increasing awareness of customers on their usage of water. Both sides are currently improperly handled in India.
We have identified several tools and ideas for proper water managements:

- Water conservation devices: Low-slung flow water conservation devices decrease water usage and expenses as well as sewer expenses.
- Weather-based irrigation control: This type of irrigation avoids water wastage, decreases usage and cost in the agriculture sector during the hot summer months.
- Concentrated water heaters: Low-flow water devices decrease the quantity of hot water used which in turn, decreases the quantity of energy used to heat the water.
- Upgrading piping: Replacing old, imprecise water meters and circulation piping can result in large water savings for the water service provider. This avoids waste during transit.
- 5. Upgrading water treatment facilities: Waste water treatment facilities are one of the principal energy users in a city. Upgrading plant and optimizing ventilation can radically decrease energy usage and cost.

This tools need to be implemented as part of a reasoned water management strategy. As we will see, focusing on reducing water wastage, is the easiest and most efficient way to improve water management.

1.2.4 Importance of Monitoring Water Wastage

The effective dispensing and management of water continuously and to every citizen is far from being a reality in India. And yet, this could be achieved through an effective water distribution system (Organization *et al.* (2015)). In our research work, currently

prevailing infrastructure of water pipelines will be custom-built so as to monitor water dispensed to every respective streamline, subsequently managing water demands. Water storage and avoiding uneven consumption are most important factors in water management. Operational steps towards regulating and decreasing water consumption by end consumers as well as from water resources must be undertaken.
Unnoticed water leakage in pipes hints to huge water loss (Mutikanga *et al.* (2012)). As a result, detecting leakage is a priority as well as stabilizing water utilization.. Water pipelines are installed and maintained by humans, which can introduce mistakes, which results in more waste. For instance rotating valves of water pipes off or/and on.
As a consequence, in this research study we have come up with an idea for a water management system which monitors water consumption in storing tanks. The experimentations are performed through existing water channel substructure. Through this water monitoring system, dispersal can be traced and the quantity of water distributed to end users can be documented. Now this experimental set up, customization of the prevailing water pipeline is made to cross over on both water outflow and inflow particularly in storage tanks.

1.2.5 CPS in Water Management

Earth's ecosphere is facing inconsiderate and harsh encounters in terms of water sustainability. Simultaneously, communication and information expertise is swiftly progressing and is also anticipated to seizure and evaluate data at a scale without precedent. With a prospective to mark aquatic environment's remote sensing allover, Cyber-Physical Systems can increase assessment making in lines with several extortions to water security, tenacious water quality glitches and contamination, flood forecasting and response, climate changes prediction and a host of others. Implementing Cyber Physical System's perception in water monitoring and dispersal chains in the undertaking to fund unceasing water 24/7 stockpiled through storing reservoirs to its one-to-one utility. Real time information through monitoring of water flow rates aids in conserving water when required the utmost. In future, Cyber Physical System will be vitally essential, as on-line mechanism is massive, improvident, involves enormous hard work and tough in sustaining too. In our research study, we are working in view of the networking dominion of CPS to monitor water distribution system in storing reservoirs and converse messaging amongst web server and DTU by means of nominal wireless tools.

CPS is an emergent exploration and research domain. Some key technologies and roles of the CPS in water management system Wang *et al.* (2015):

- The distributed sensing. Perception and communication enable the functionality of time critical and time aware.
- The predictive and adaptive control of the hierarchical hybrid achieves synchronized actions and tightly coordinated in water CPS to be intrinsically synchronous, noisy and distributed.
- The prognostic and diagnostic help in predicting, identifying and preventing faults.
- The human interaction facilitates the design-based model of reactive water.
- The verification, certification, and validation ensure system functionality and safety.
- The modularity composability and abstraction enable the element to be reused and combined while retaining security, safety, and reliability.
- The system-based engineering standards and engineering enable the development and design of the reliability system ensuring integration and interoperability with the legacy system.
- The cybersecurity provides safety by guarding against the malicious attacks.

Application for the sustainability of water for CPS:

- The increase in reliability efficiency public confidence and security increase water systems thus the path towards water sustainability has been set. The sensing of water quality advances in computing and communication in real time for monitoring and interaction. On a continuous basis, the remotely used in-situ data infrastructures are cost-effective to analyze visualize, communicate and store in real-time.
- In the CPS the important application in the water sector is the systematic monitoring of water distribution.

- For the emergency management and advance-warning, the quality monitoring of the source water can be critical. The water supplies need to be a safeguard in order to protect the ecosystem, public health and time-monitoring due to cost limitation.

In our research study, literature review and assessment is principally engrossed on monitoring and distribution of water by means of Cyber Physical System.

In Eswaran and Kumar (2012) research study, an intangible proposal of computerized water delivery and metering scheme for domestic edifice is prepared. Proteus 7.6 simulators were used for simulation in numerous real times input and output conditions. Dual channel padlocked loop computerised water and metering distribution scheme with GUI existed, established and employed efficaciously. The perception projected by them is appropriate for executing and to assimilate apartments or entire towns.

Suri *et al.* (2017) emphasizes on water related issues on the basis of underground water mugging, Leak Detection, and govern watching in Water Supply. This paper boons various data analysis in numerous domains, their strengths, limits, and relative study grounded on thorough literature assessment. In Reddy and Eswaran (2013) , metro water overhead tanks monitoring system-using ZigBee centred WSN is proposed. The system monitors filling of overhead tanks on the basis of available water level and also keeps checking on water distribution. Prototype hardware and GUI for controller node is also urbanized and verified.

Nicolini and Patriarca (2011) designate an innovative tactic for model calibration and mock-up of Water Distribution Networks, commencing from physical time monitoring of water pressure and movement all over the structure. This practise is implied in real water distribution network, because of which vigorous seepage regulation and system reintegration happened resulting in significant water savings and reduced energy savings.

In Suresh *et al.* (2014), a Cyber- Physical system (CPS) is projected for unceasing observing of a Water Distribution System. The research study largess chief design of a CPWDS, entailing mobile sensors moving over pipes assisted by water flow, static beacons are planted outside the pipes and they collect data from mobile sensors, and a control system exists to adjust flow course in the main pipes of the WDS. Stream reversal in numerous pipes of a WDS is also validated in this work.

In Wang *et al.* (2015) a gestalt of water CPS for enduring existence from four perilous aspects i.e., communications besides networking, sensing and instrumentation, and calculating and regulating is poised. In Imen and Chang (2016), a Cyber-Physical System (CPS) is established which responds to requirements of shrewd and maintainable drinking water organisation supervision. Combined 5-level CPS architecture stands projected.

In Lin *et al.* (2011) chapter converses copious challenges correlated with simulation and modelling of CPS. One of the foremost defies i.e., specifically and perfectly demonstrating operation and features of physical structure through implementing domain-specific software tool EPANET; a simulant for WDNs is also addressed in this research work. Integrated simulation technique accessible in this manuscript is an introductory footstep towards facilitating advanced exploration and research in the direction of CPS-based virtual reality and conservational choice sustenance.

1.3 MOTIVATION FOR THIS RESEARCH STUDY

Water management is of tremendous sanitary and monetary importance. Yet there is currently a widespread scarcity in water resources and distribution due to economic factors. This is also due to improper water management, which fails to extract, purify, and deliver hygienic water at reasonable prices. The International Water Management Institute [IWMI] has categorized the ecosphere into areas of physical and economic scarcity as presented in Figure 1.7 . While the former are the zones wherein people cannot pay for clean water, the latter are the zones with little to no natural water accessible to upkeep the population. At present around 1.2 billion individuals live in areas with physical scarcity, while 1.6 billion people in regions considered to have economically induced water scarcity. The remaining regions have no form of scarcity. For that reason, there is a great opportunity for enhancing public water delivery systems to guarantee proper and continuous water delivery to everyone. Efficient management of water is the key area of improvement and innovation to help these 2.8 billion people.

Public water delivery systems typically range hundreds of square miles, and inefficiencies can be found everywhere. It has been projected that public utilities can hypothetically lessen their energy costs by approximately 15-30%. So as to upsurge efficacy of these systems and recognize the would-be cutbacks in energy costs, numerous parameters through the distributed network must be traced and their respective

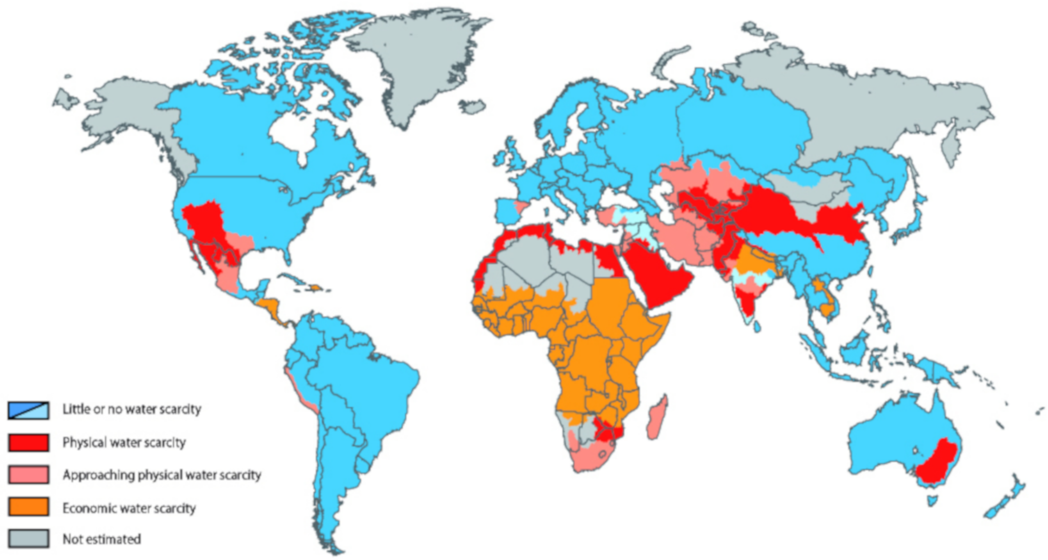


Figure 1.7: Regions of the world with physical and economic water scarcity

information and statistics must also be dealt with to take the best course of action. Manually executing such maneuvers is wearisome, error-prone, and costly.

Consequently, the easiest area of improvement is to intelligently manage the obtainable water resources. This should in turn show great ecological as well as economic benefits: improvement on water systems includes decrease wastage and decreases the overall cost of water management. It is a win-win solution.

The overarching aim of this thesis is to better understand Cyber-Physical Systems such as next generation water distribution networks and smart cities and to better control them through the development of integrated self-adaptive protocols that support distributed and collaborative analysis, prediction, and control using remote terminal units and servers and which provides guarantees pertaining to reliability, stability, convergence, and sustainability. Our research focus is intended to slog out a Water-Management CPS by means of small scale deployment and large-scale simulation determined by outcomes of the earlier one, in the direction of effectual deployment of minimum wireless devices at storage tanks in NITK campus. We are self-assured regarding the study outcomes of the PhD work which will deliver a worthy intermingling of small scale deployment and system-design related outcomes headed for proficient consumption and dissemination of water through CPS and accordingly dipping water wastage.

1.4 PROBLEM OVERVIEW

Water is one of the most important and precious resource on Earth. At no time has its value been more important and never has it been more scarce and unreliable. Three-quarters of Earth's surface contains water, however 98% of it is salt water. Less than one percent of water on Earth is freshwater accessible for human intake (WATER). Administrations in developed and developing countries guarantee and make sure that water is accessible to its citizens on a consistent basis to sustain existence. With the coming economic development, if current undeveloped nation-states advance in their development to a similar level as developed countries today, then water demand will more than triple he consumption which was of 3,350 cubic kilometers in 2000 (Experts).

In India, where water scarcity is perpetually increasing, an enormous margin of water delivered by metropolises is wasted during the transit and delivery process because of leakage or mixing with non-potable water. An average Indian consumes around 150-200 litres of water every day. Yet water pipes in the distribution network are not properly maintained, due to which most of the leakages are not discovered. Water passed on through trucks also trickle water all the way to fund merely reduced water quantity to individuals. On an average a family wastes around 800 litres of water per week, or 42,700 litres of water yearly, because of household leaks. It is equivalent to the quantity of water needed for over 300 loads of laundry. Domestic leaks waste approximately 1 trillion litres of water every year countrywide, which in turn is equivalent to the yearly domestic water use of closely 11 million households (GAO (2014)). Most of the water is wasted on an everyday basis without anyone even being conscious of it. As per a latest report of India's Silicon Valley which details this man-made water shortage, it is high time to take suitable measures to save water [31]. And this shows the timeliness of our research study and, in part, explains why we chose this area of research.

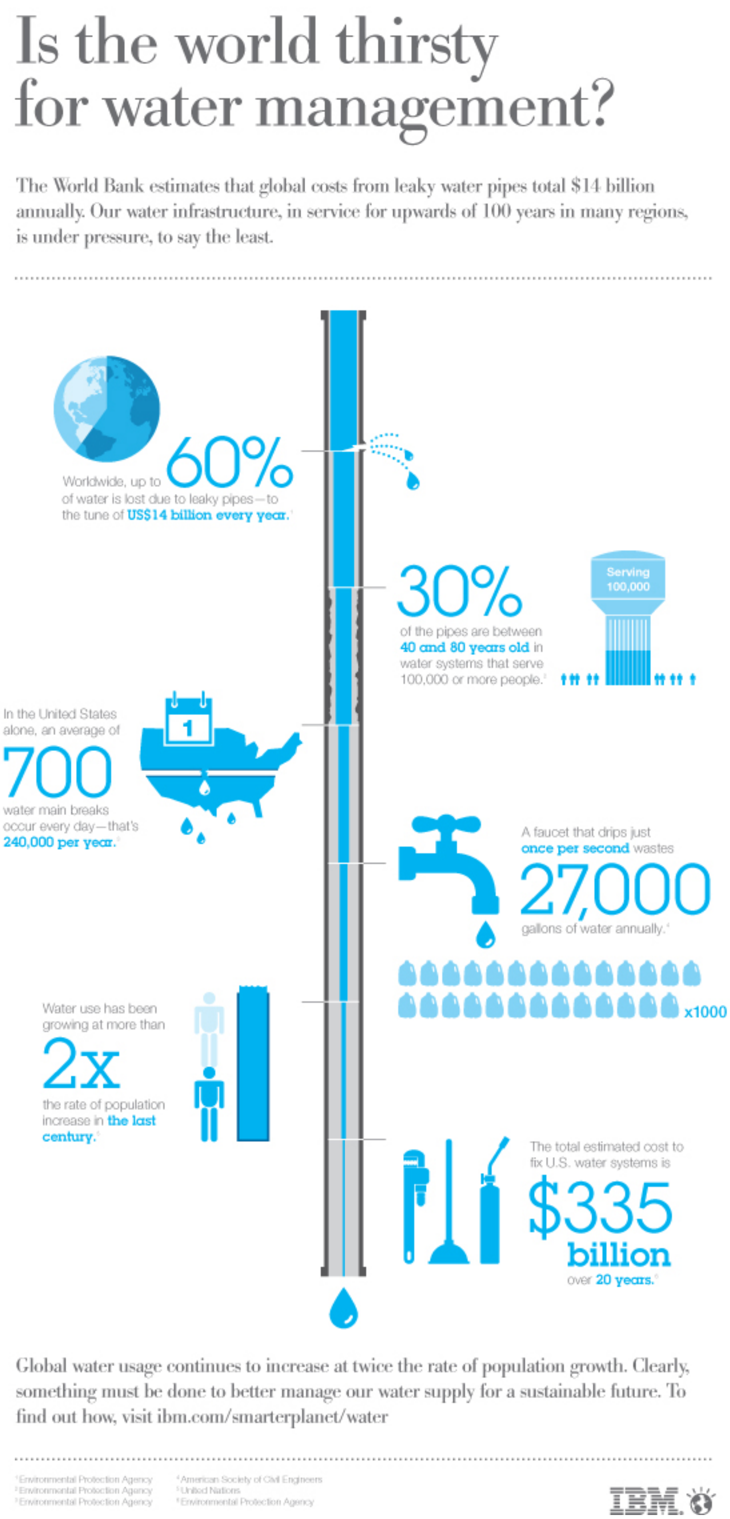


Figure 1.8: World Bank approximates global costs from leaky water pipes as 14 billion dollars per annum. Present water substructure, in service for in the air of 100 years in foremost provinces is underneath pressure. Global water usage remains to upturn at twice the rate of population growth (Maxwell-Gaines (2012), McNabb (2019))

On an average a family wastes of around 180 gallons of water per week, or 9,400 gallons of water yearly, because of household leaks. It is correspondent to the quantity of water essential to wash-down more than 300 loads of laundry. Domestic leaks waste approximately about 900 billion gallons of water every year countrywide, which in turn is equivalent to the yearly domestic water use of closely 11 million households (USEPA). Most of the water is wasted on an everyday basis without even becoming conscious of it. As per a latest report of India's Silicon Valley fronting a man-made water predicament, it is high time to take suitable measures to save water Experts. At present many researchers are working in cyber physical system technology for monitoring and control of several activities within a large area or campus. Since water will be used efficiently, the probable shortage of water to different areas could be mitigated. Our effort is to implement CPS to reduce water scarcity issue by using wireless sensor to monitor and control the water flow in a campus and extending to large area efficiently. So that, Water wastage can be reduced and utilization of water can be improved. After the implementation of this project we hope to reduce the wastage of water due to efficient utilization and distribution of water through CPS.

1.5 RESEARCH OBJECTIVES and APPROACHES

As mentioned earlier, our approach is to have a small scale-deployment at NITK and carry out the experimental studies at Robert Bosch Centre for Cyber Physical Systems(RBCCPS), Indian Institute of Science (IISC), Bengaluru, Karnataka. The data collected is first analyzed in off-line fashion to get useful insights. The Analytics also facilitate useful models which are absolutely essential to carry out large-scale simulation. We would be resorting to simulation since campus-wide deployment of the sensors, actuators and connectivity is infeasible. We believe, a small hardware based prototype, data analytics, CPS simulation and experimenting with a World Class Test Facility at IISc, will enable us to get useful outputs culminating towards the Campus Level Water Management CPS. The objectives are explicitly outlined below:

- Experimental analysis of Cyber Physical System for an Island (Campus Storage tank) water management.

- Simulation and selection of imputing data algorithms for handling of missing data

- To propose a model for forecasting water flow data.

For few tasks planned, there is enough clarity in terms of the approaches and methodologies. For others, there would be appropriate efforts/ studies in nailing down the exact strategies. Data Analysis would be carried out using R Software. Since the data available through our NITK small-scale deployment and tailored experiment at IISc are unique to our scenario, there is a good scope to get couple of publications.

- We envisage the following immediate tasks in Data analysis.
  - How to handle missing data
    * Planning to bring the aspect of comparing different types of imputation algorithm and choosing the best suited algorithm for our application.
  - Are there any periodicity.
    * Periodicities help us to work out better Water Management Strategies
  - Predictive Analytics
    * Predictive Analytics on the data is also useful in the Overall Water Management.For instance, if we could predict the consumption for the next day, only necessary amount of pumping can be carried out, which, not only results in efficient managing of water but also results in electrical energy savings. There is a good scope to arrive at novel algorithms, through right functional approximations.
  - Leakage detection
    * Leakage detection is carried out through flow balancing to start with. There is a further scope to arrive at methods based on statistical inference.

Chapter 2

MONITORING UNIT

Outline of the chapter

- **Section 2.1** gives the gist of the introduction about importance of monitoring water usage.
- **Section 2.2** discusses about projecting and anticipated customization of existing pipeline infrastructure structure to monitor water flow rate in storage tanks using minimum wireless technology and assimilate the same into CPS which provokes the controller on the basis of historic real-time water flow information.
- **Section 2.3** summarizes monitoring unit of CPS in this Chapter 2.

2.1 INTRODUCTION

The operation of water distribution service for continuous water supply is not effectively managed in India. Water supply provided is said to be continuous if water is delivered to every customer continuously every day throughout the year, using an effective water distribution system (UNESCO (2006)).So in our experiment, we are customizing existing water pipeline infrastructure, in order to monitor water distribution to practice demand management. Storing water and preventing unbalanced water consumption are the major challenge in water management research. Effective steps need to be taken to control and reduce the water wastage from end users as well as from water resources. Undetected water leak in a pipe leads to huge water loss (Mutikanga *et al.* (2012)). Thus, identifying the water leak in pipes in prior will help the water management system improving balanced water utilization. For several years, water pipelines are deployed and maintained by human interventions, such as turning off or turning on valves of water pipes. It is known that the chance of water wastage is more because humans are prone to make mistakes. Thus, we come up with a water management system that monitors the water utilization in storage tanks. Our experiment is done with available water pipeline infrastructure. By the help of this water monitoring system, water distribution can be tracked, and the amount of water consumed by the end user can be recorded. In our experimental set up, customization of the existing water pipeline has been done to check up the water inflow and outflow details in storage tanks.

2.2 EXPERIMENTAL SET UP FOR MONITORING UNIT

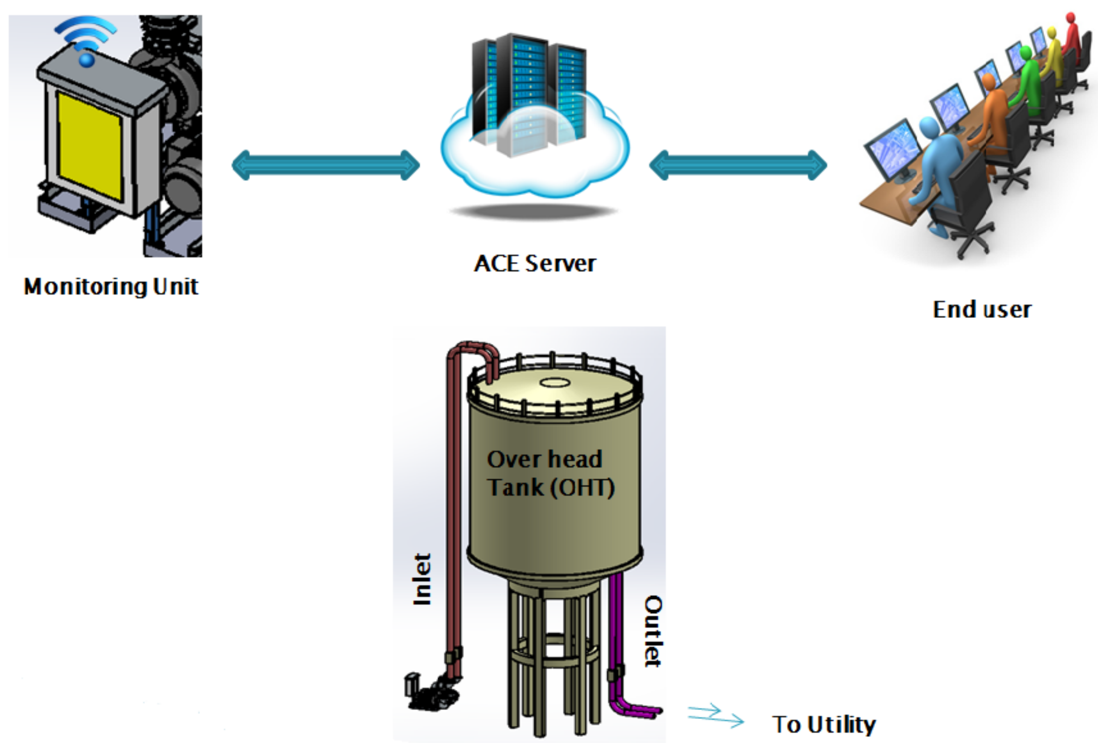
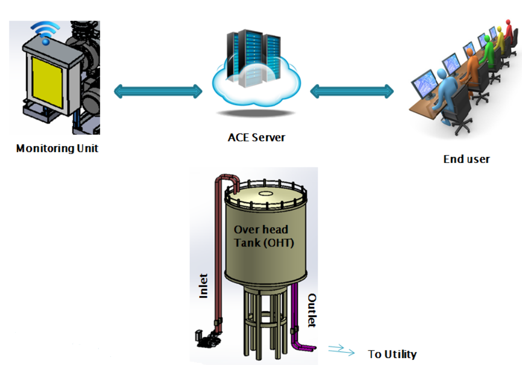


Figure 2.1: System Tank Overview

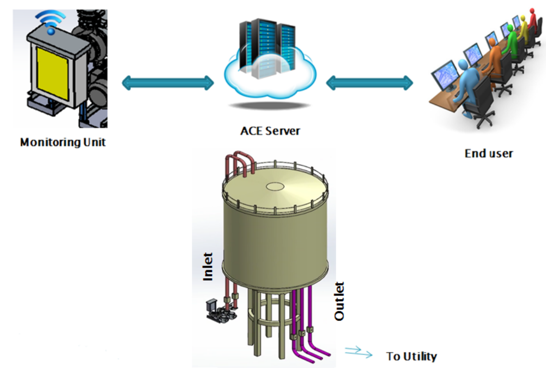
Each storage water tank consists of water monitoring system see Figure 2.1 which shall include following components:

- METER INTERFACE UNIT : MIU includes water meter, reed switch and RF module.
- AGGREGATOR : To aggregate data from different MIUs and communicate with DTU
- DTU(Data transmission Unit): To process water flow level related data and upload data to server.

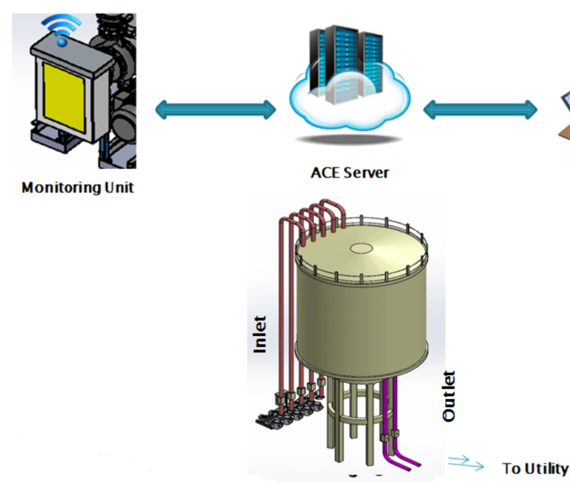
The pictorial representation of three Storage tanks installed at different places in NITK, Surathkal campus is as shown in Figure 2.2.



(a) Storage tank 1.



(b) Storage tank 2.



(c) Storage tank 3.

Figure 2.2: Pictorial representation of three Storage tanks respectively

The storage tank shown in Figure 2.2(a) consists of two inlets and three outlets. Similarly Figure 2.2 (b) and Figure 2.2 (c) consists of one inlet and three outlets,five inlets and two outlets respectively. Here each inlet and outlet pipe has RF module mounted on water meter. The RF module is used to count the number of pulses with the help of water meter and stores the counted pulses into flash memory. Then aggregator communicates with all RF modules of respective tanks and receives the counted pulses and shares it with DTU.

2.2.1 Meter Interface Unit

Each inlet and outlet pipe of a storage tank is connected to MIU (Meter Interface Unit) as shown in Figure 2.3.A water meter dial is connected to reed switch and then

RF module is interfaced with the reed switch. Whenever the rotating needle in the dial inside the water meter contacts the magnet of the reed switch, then reed switch will discharge the RF module pin. This is read out as count one and it is stored in non-volatile memory (RF flash memory). It keeps on storing the discharge count until aggregator sends the request query to the RF module for the inlets and outlets data. Once the aggregator receives all details from the RF module, then aggregator sends reset command to RF module to reset the count pulse to zero.

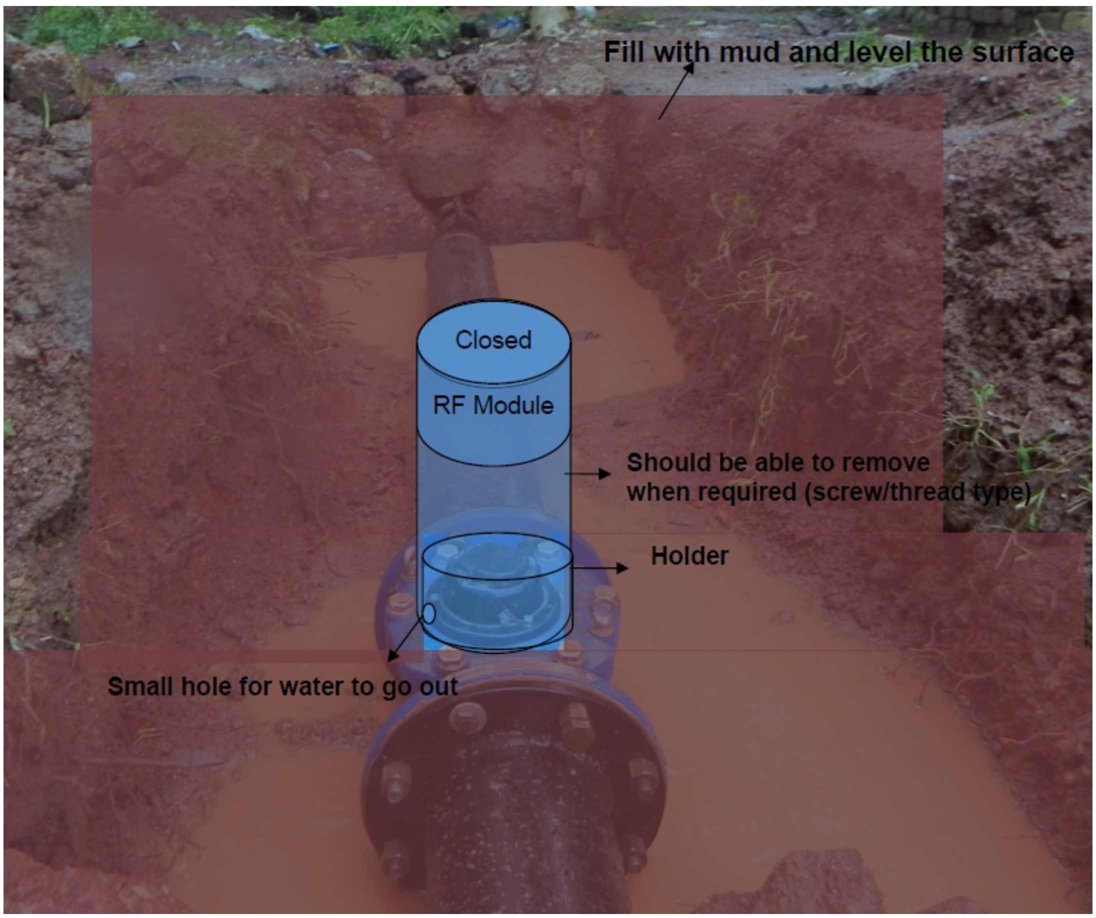


Figure 2.3: Meter Interface Unit

2.2.2 865-867Mhz RF module

In this work, we have used (TM (2010)) 865-867 MHz RF module which provides robust wireless communication for all conditions. The main advantage of this module is that with the help of serial data interface the module is suitable for adding wireless capability to any embedded devices. Other advantages include minimal power and provide reliability of data deliveries between devices. The main features of this device are (TM (2010)):

- Supports point to point, point to multipoint and mesh topologies. (Here, star network is established).
- Support 2 FSK, 2-GFSK.
- Configurable transmit power up to 10dBm.
- Maximum RF data rate is 500kbps.
- Data reliability, acknowledgment mode communication, is used
- Source/Destination addressing.
- Unicast and Broadcast communication.
- Analog to Digital conversion and Digital I/O line support.
- Power saving modes.

Specification of 865-867 MHz RF module is as shown in the Table 2.1.

Table 2.1: Specification of 865-867 MHz RF module (TM (2010))
Source: www.melangesystems.com.

SL Num	PARAMETER	DETAILS
1	**Power**	
i)	Supply Voltage	2.4 to 3.6 V.
ii)	Transmit Current	20mA@0dBm, 35mA@10dBm.
iii)	Idle/Receive Current	20mA.
iv)	Power-down Current	<10micro A.
2	**General**	
i)	Frequency	Quality of sensing data
ii)	Nominal Transmit Power	10dBm.
iii)	RF Data Rate	2.4kbps to 500kbps(Max).
iv)	Receiver Sensitivity	-106dBm(at 9600 baud).
v)	Serial Data Rate	Up to 115200 baud.
vi)	Operating temperature	-40 to 85ˆ0C.
vii)	Antenna Connector options	MMCX.
viii)	On-board Antenna	Spring Antenna.
3)	**Network**	
i)	Supported network topologies	Point to point, point to multipoint, mesh topologies.
ii)	Addressing options	PAN ID and addresses.
4)	**Mechanical**	
i)	Dimension	36mm * 26mm.
ii)	Interface connector	2 * 10 pin berg stick, 2.00 mm pitch.

2.2.3 Pictures of Installed Meter Interface Unit including water meter, reed switch and RF module.

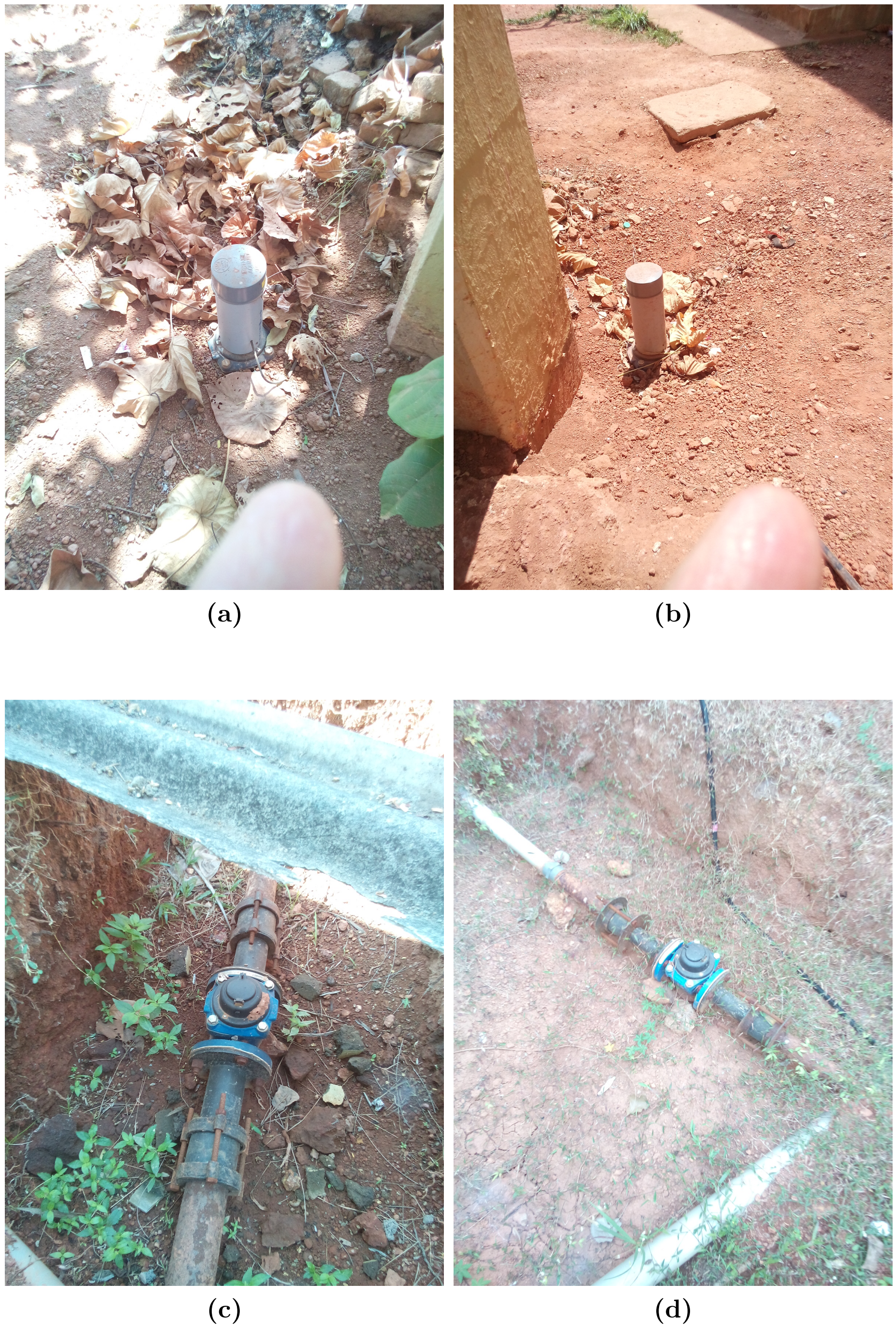


Figure 2.4: Pictures of Installed MIU including water meter, reed switch and RF module.

2.3 SUMMARY

In this work we have addressed the systematic approach of modifying the existing water pipeline infrastructure using minimum wireless technology shown in the Figure 2.4 to explore the possibility to monitor the water distribution through a storage tank and later bring out the aspects of Cyber Physical system (CPS). Our experiment was carried out at NITK, Surathkal, Mangalore, India to monitor the amount of water flows in storage tanks. Customization includes a set up of RF transceiver modules interfaced with water meter which has been installed to get the water inflow and outflow details for each pipe. With respect to sampling time in the monitoring unit, a set up of another RF transceiver interfaced with aggregator and DTU will receive the water inflow and outflow details and displays in the web server. Presently, we are wirelessly monitoring three storage tanks acquiring inlet and outlet data from each storage tank in the campus. Our experiment aims to adapt ways and proposes to use CPS technology to overcome the improper handling and care of water supply infrastructure. Our effort is to study Water distribution CPS, starting with small scale experimental set up at NITK campus and then extending that understanding to large area. From our result it is noticed that the data analytics facilitates useful models which are absolutely essential to carry out large-scale simulation which improves the water utilization in an efficient way.

Chapter 3

NETWORKING UNIT

Outline of the Chapter

- **Section 3.1** explains about the networking part of CPS which includes the software requirement and communication requirements used in the monitor unit to analyze the amount of water flowing in and out of storage tanks.
- **Section 3.2** describes the communication protocol used.
- **Section 3.3** elaborates the working of software and communication requirements between RF module and aggregator.
- **Section 3.4** focuses on details about communication between DTU and aggregator.
- **Section 3.5** concentrates on operation requirements for interaction between DTU and ACE web server.The resultant data is available in a web server.
- **Section 3.6** summarizes the networking unit of CPS in this Chapter 3.

3.1 INTRODUCTION

The pictorial representation of monitoring unit which includes aggregator, DTU and web server which forms the networking unit of CPS is as shown in the Figure 3.1. In the monitoring unit as shown in Figure 3.2 is used to develop an automation solution that reads the water inflow and outflow parameters from each storage tank. Aggregator receives the water inflow and outflow parameters from each storage tank.

An aggregator is interfaced to DTU with RS 232 port as shown in Figure 3.2. DTU is interfaced with an aggregator to read water inflow and outflow details and transmit the respective information to web server. Each storage tank will have one DTU, and DTU will collect information about water flow details from both inlet and outlet of the storage tank. For every 15 minutes, based on the water inflow and outflow details, DTU shall calculate the Water Pulse count for each Inlet ,Water Pulse count for each Outlet ,Total Pulse count for Inlet, Total Pulse count for Outlet, Net Volume of the storage tank. Finally, in ACE web, all the resulting information collected in DTU will be displayed.

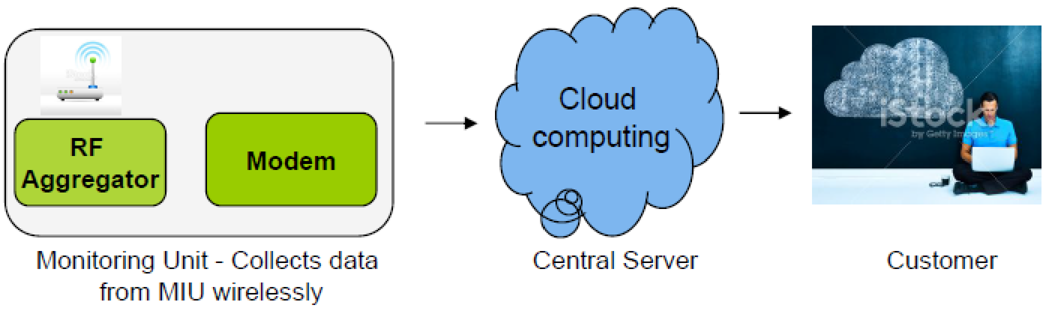


Figure 3.1: Networking unit

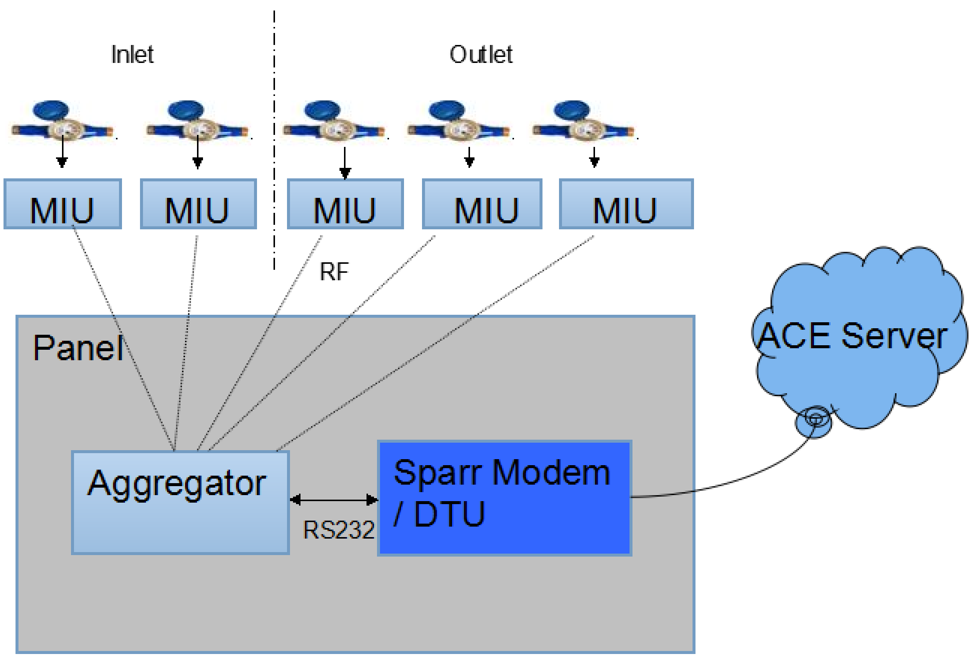


Figure 3.2: Communication between MIU, Aggregator, DTU and Web server.

3.2 COMMUNICATION PROTOCOL

The water meter is interfaced with RF module (865-867MHz) is a low power RF module which provides immune condition for wireless communication (TM (2010)). The main influence for using this RF module is that, with the help of serial data interface we can add wireless capability to any product. The use of wireless communication on water meter hugely cut down the cost of maintenance and manual cabling. The gateway is used to communicate with the ACE web server over GPRS and establish HTTP POST connection to send data. The communication between the DTU and the web server is device initiated communication and hence the DTU has to be pre-programmed with the server domain information. The web application is available for enterprise users through both Intranet and Internet as per the policy.

3.2.1 ACE(Affordable, Convenient and Efficient)

ACE, is a hardware, communication, software and analytics technology platform developed by a team at "Thought Focus", Bangalore that helps easy data collection and analysis. ACE convolutes various networking technologies to collect data from the field and transmit to a central server. The last mile networking technology, which is used to monitor the measurement points can have traditional wired bus topology, Infrared, Optical Interfaces, low range RF, wireless LAN to push the data to the central server using the widely available long haul TCP/IP, GPRS, GSM networks. Ubiquity, Agnostic, Affordability, Convenience, and Efficiency is the primary characteristics of ACE. Its cloud presence allows centralized monitoring and analysis. Today the challenge for people willing to start measuring is where to save these measurements and how to access the historical measurements for meaningful analysis. Every hardware manufacturer tries providing software that works with its products only and ends up having multiple software's only to be disillusioned. ACE solves that problem. It is capable of reading data from any device and brings all data to one common place.

Features of ACE

- ACE is Thought Focus proprietary smart measurement framework that allows companies monitor their operations and assets, without boundaries of geography, location and time.

- Powerful web-based network management, data analytics platform that offers a range of features to deduce meaningful inferences from the data that is collected.

- Models Physical Sensor conforming to IEEE 802.15.4 Application Profiles. This makes it highly flexible to measure Utility Consumption (Water, Energy, and Fuel), Quality Measurement (Water, Air, and Milk), and Asset Location.
- Developed as Service Oriented Architecture using latest technology platforms, it allows data collection from ACE and NON-ACE networks.
- The range of sensor network technologies available that include Ghz and Sub-GHz RF networks, Optical Fiber Back Bones, Standard Industrial Bus Topology, GSM / GPRS.
- Works equally effective in Metropolitan networks (¿ 1000 sensors), Industrial Networks (¿100 sensors), Private networks (¡100 sensors).

3.3 COMMUNICATION BETWEEN RF MODULE and AGGREGATOR

The following are the primary functionalities of the 865-867 MHz RF module are

- Water meter output is connected with RF module input.
- Increment pulse count when a pulse output from water meter.
- Store the incremented count into non-volatile memory.
- Send stored count value to Aggregator, when Aggregate read command interrupt.
- If Reset pulse count command, Reset the Pulse count in RF module

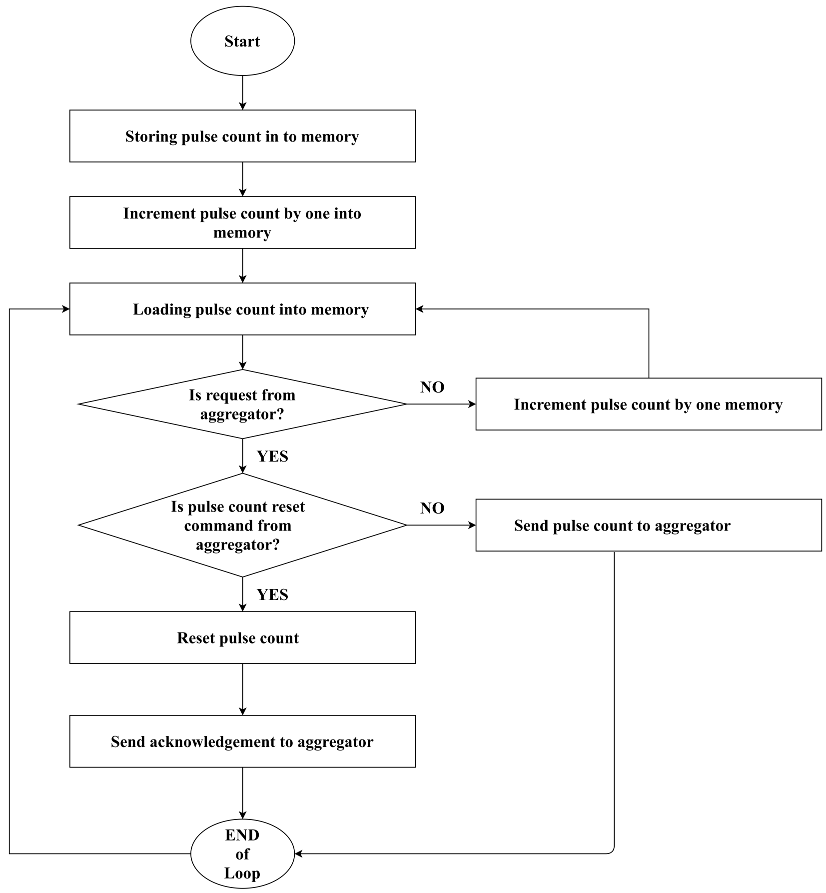


Figure 3.3: Flowchart for Communication between RF module and Aggregator

Algorithm for communication between RF module and aggregator as shown in Figure 3.3 is listed below:

- Step 1: Storing the pulse count into non-volatile memory.
- Step 2: Increment the counter value by one whenever reed switch discharges the RF module and store into non-volatile memory and go to Step 3.
- Step 3: Loading pulse count into non-volatile memory and go to Step 4.
- Step 4: If aggregator read command is received, go to step 5, else Increment pulse count and go to Step 3.
- Step 5: If aggregator Pulse count reset command is received, reset the pulse count and go to Step 6, else send pulse count to aggregator.
- Step 6: Send acknowledgement to aggregator and go to Step 3.

3.3.1 Structure of Pulse count query command

Table 3.1: Pulse count Query command

COMMAND START BYTE	LENGTH	COMMAND TYPE	DESTINATION ADDRESS	READ / WRITE	CHECK SUM	END BYTE
0X2B (1Byte)	0X08 (1Byte)	0X01 (1Byte)	0X00000002 (4Bytes)	0X00 (1Byte)	0x62 (1Byte)	0X2C (1Byte)

Length (8) == Length (1) + Command Type (1) + Destination address (4) + Read/write (1) + Checksum (1)
Checksum = Sum of all bytes in command byte stream except check sum(1 byte addition)

3.3.2 Response structure of Pulse count query command

Table 3.2: Response structure of Pulse count Query command

RESPONSE START BYTE	LENGTH	COMMAND TYPE	SOURCE ADDRESS	PULSE COUNT	CHECK SUM	END BYTE
0X2D (1 Byte)	0X0B (1 Byte)	0X01 (1 Byte)	0X00000002 (4 Bytes)	0X00000005 (4 Bytes)	0X6A (1 Byte)	0X2C (1 Byte)

Length (11) == Length (1) + Command Type (1) + Destination address (4) + Pulse count (4) + Checksum (1)
Checksum = Sum of all bytes in command byte stream except check sum(1 byte addition)

3.3.3 Structure of pulse count Reset command

Table 3.3: Pulse count Reset command

COMMAND START BYTE	LENGTH	COMMAND TYPE	DESTINATION ADDRESS	READ / WRITE	CHECK SUM	END BYTE
0X2B (1Byte)	0X08 (1 Byte)	0X02 (1 Byte)	0X00000002 (4 Bytes)	0X00 (1 Byte)	0X63 (1 Byte)	0X2C (1 Byte)

Length (8) == Length (1) + Command Type (1) + Destination address (4) + Read/write (1) + Checksum (1)
Checksum = Sum of all bytes in command byte stream except check sum(1 byte addition)

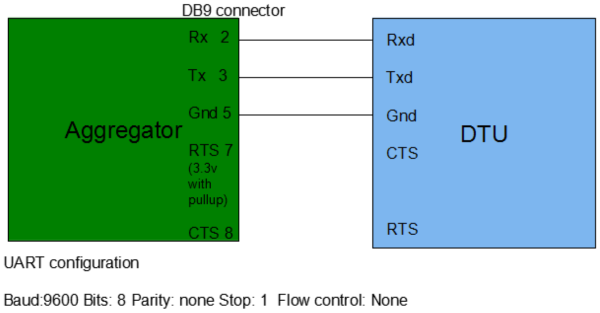
3.3.4 Response structure of Pulse count reset command

Table 3.4: Response structure of Pulse count Reset command

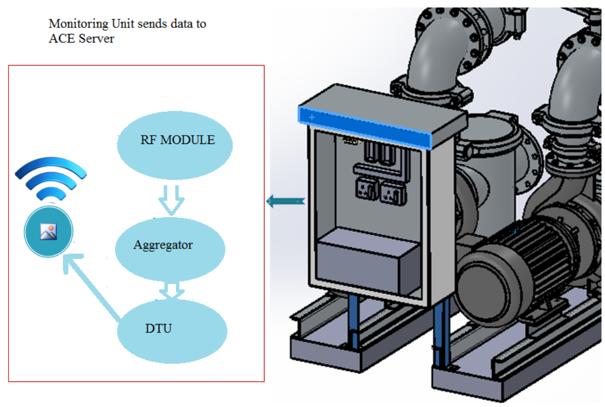
RESPONSE START BYTE	LENGTH	COMMAND TYPE	SOURCE ADDRESS	PULSE COUNT	CHECK SUM	END BYTE
0X2D (1 Byte)	0X0B (1 Byte)	0X02 (1 Byte)	0X00000002 (4 Bytes)	0X00000000 (4 Bytes)	0X68 (1 Byte)	0X2C (1 Byte)

Length (11) == Length (1) + Command Type (1) + Destination address (4) + Pulse count (4) + Checksum (1)
Checksum = Sum of all bytes in command byte stream except check sum(1 byte addition)

3.4 COMMUNICATION BETWEEN DTU AND AGGREGATOR



(a) Interface Diagram



(b) Monitoring Unit with DTU

(c) Photo representation of Hardware

Figure 3.4: Communication between DTU and Aggregator

Each storage tank will have one DTU, where DTU is used to collect the inlet and outlet water flow details of each pipe in a storage tank from the aggregator with the help of RS232 port through UART configuration as shown in the Figure 3.4. We have designed DTU to request the data from different MIUs for every sampling time period of 15mins. So, DTU after every 15minutes will query it through the aggregator. Then, aggregator communicates with different MIUs and collects the data from MIUs. Once the aggregator receives the information it starts transmitting the information to DTU through RS232 port. When the DTU receives the information based on inlet and outlet details of each tank, it will calculate the total volume of inlet, outlet and finally the net volume in the tank. Inlet volume of water flow in a storage tank is calculated as

$$ICj = Ij * Dj \tag{3.1}$$

where IC= Inlet count, Ij= Count in the water meter dial, Dj= Discharge count and j= represent the tank number ie., 1,2,3...
Outlet volume of water flow in a storage tank is calculated as

$$OCj = Oj * Dj \tag{3.2}$$

where OC= Outlet Count, Oj= Count in the water meter dial, Dj= Discharge count and j= represent the tank number ie., 1,2,3...
Total Inlet volume of water flow in a storage tank is calculated as

$$Inletj = [(PresentReading) - (PreviousReading)] * ICj \tag{3.3}$$

$$TotalInlet = Inletj + Inletj + 1 + Inletj + 2 + ... \tag{3.4}$$

Total Outlet volume of water flow in a storage tank is calculated as

$$Outletj = [(PresentReading) - (PreviousReading)] * OCj \tag{3.5}$$

$$TotalOutlet = Outletj + Outletj + 1 + Outletj + 2 + ... \tag{3.6}$$

Total Volume of water flow in a storage tank is calculated as

$$PresentWaterReading = PreviousWaterReading - TotalOutlet + TotalInlet \tag{3.7}$$

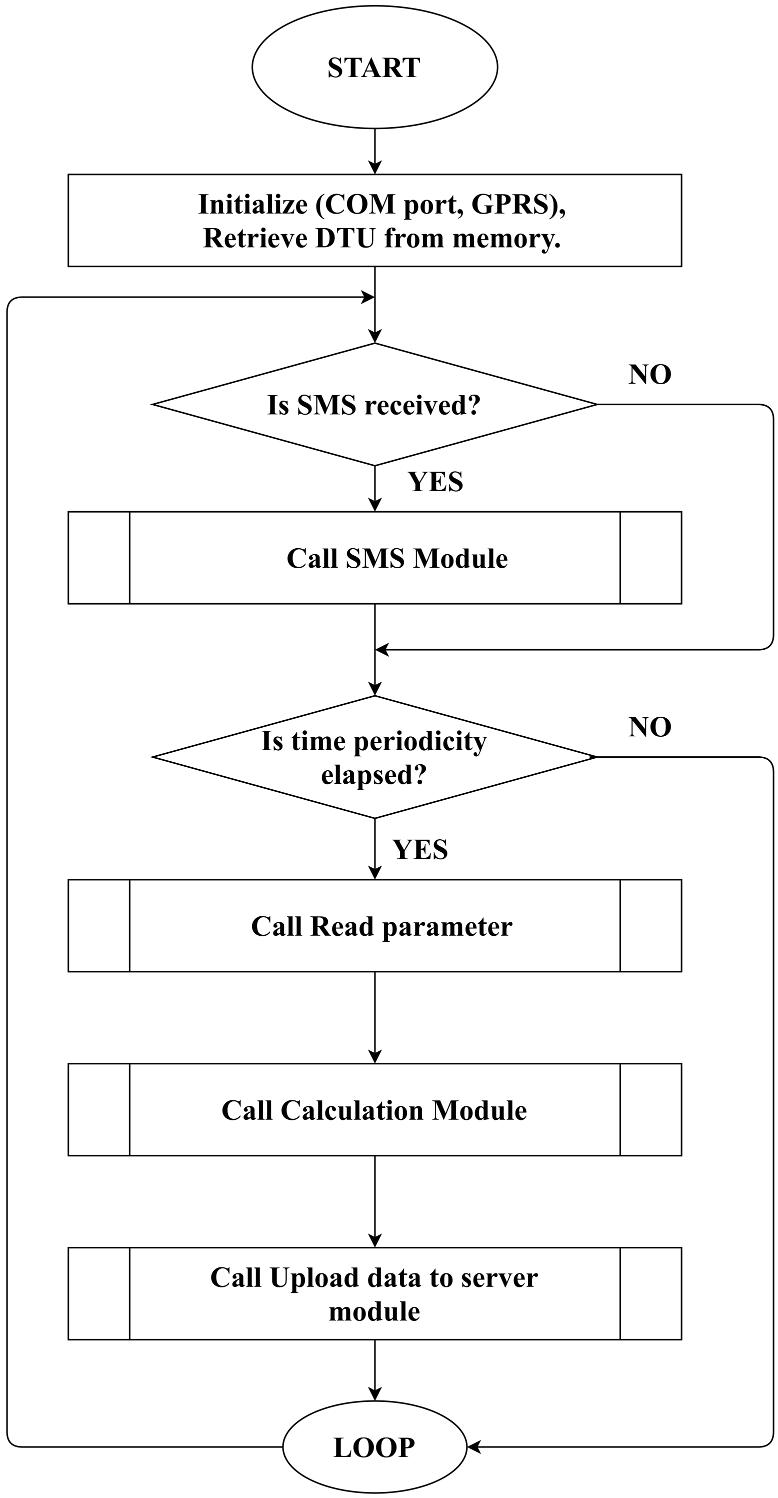


Figure 3.5: Flow chart for communication between DTU and Aggregator

Algorithm for the basic steps of the operation for Communication between DTU and aggregator shown in Figure 3.5 to 3.8 and is listed below:

- Step 1: Initializes the peripherals and retrieves the settings from memory.
- Step 2: If SMS received, goes to step 3. Else directly goes to Step 4.
- Step 3: Processing the SMS module block by checking device identification information as shown in Figure 3.6 and go to Step 4.
- Step 4: If web time periodicity is elapsed, then go to Step 5. Else goes back to Step 4 till web time periodicity is elapsed.
- Step 5: Reads the inlet and outlet parameter from each water meter as shown in Figure 3.7 and go to Step 6.
- Step 6: Calculates the total volume of inlet, outlet and net volume of total water flows in each tank as shown in Figure 3.8 and go to Step 7.
- Step 7: Sends the received data to server over GPRS and repeat the loop.

The operation of DTU $< - >$ aggregator communication :

- To read the water meter readings from aggregator and transmit the data to server, DTU acts as an interface with RS232 port connected to aggregator.
- DTU collects the readings from aggregator through RS232 port to develop a process automation.
- Firstly, DTU checks any SMS received from Master Mobile (MM) or Configuration Mobile (CM).
  - SMS module is processed by DTU when it receives SMS request from MM or CM as shown in Figure 3.6 and process Step 3 in Algorithm.
    Processing SMS Commands include:
    * Inlet Litres Per Pulse
    * Outlet Litres Per Pulse
    * Total Inlet
    * Total Outlet
    * Previous Net Volume

* Device ID
* DTU Address
* DTU configuration number
* Get MIU settings
* Get meter reading
* Reset DTU and so on.

Table 3.5 shows an example for SMS requested by MM to DTU and SMS response sent by DTU to MM.In Example1 SMS requests, get meterreport. SMS responses Inlet and outlet details. In Example 2 SMS requests, get meterreport error. SMS responses get meterreport error with error code shared in Table 3.6 (error code represents reason for error).

- Once the time periodicity is elapsed, DTU will read the received parameters through READ PARAMETER as shown in Figure 3.7 and process Step 5 in Algorithm .
  - Consecutively read water meter parameters from both Inlet and Outlet by sending commands as shown in Table 3.7.
  - Retry if no response is received from Aggregator (3times).
  - Parse the received response from Aggregator.
  - If DTU fail to read from the Aggregator or Aggregator responds with an error, then DTU reports as error in the message.
- After receiving the water parameters, next is CALCULATION MODULE which is shown in Figure 3.8 and process Step 6 in Algorithm.
  - DTU shall store all previous readings in the non-volatile memory
  - Calculates the equation from 3.1 to 3.7 to calculate the total volume of inlet, outlet and net volume of total water flows in each tank.
- Finally we are uploading the final data in the server. Send framed data to server after establishing HTTP POST over GPRS on elapse of web periodicity.

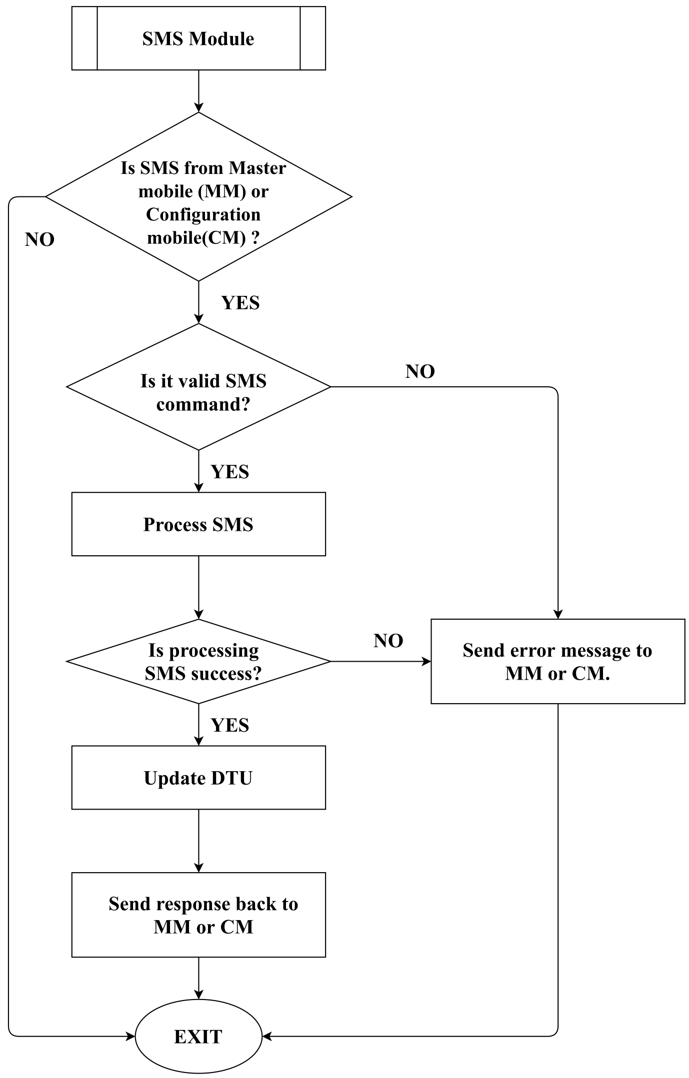


Figure 3.6: Flow chart for SMS module between mobile and DTU

Algorithm for SMS module shown in Figure 3.6 is listed below:

- Step 1: Checking whether SMS received from master or configured mobile. SMS received from non-configured number will be deleted by DTU with out any response and goes to Step 2.

- Step 2: Checks the SMS command. If command is valid goes to next Step 3 otherwise goes to Step 4.

- Step 3: Process the SMS command received as shown in Table 3.5 and goes to Step 5

Table 3.5: Command is used to get the present reading of water meters

	Format	Example1	Example2
SMS Request	get meterreport	get meterreport	get meterreport
SMS Response	I1: xxxx I2: xxxx O1: xxxx O2: xxxx O3: xxxx Or get meterreport error errorcode x: number Ix: Inlet Ox: Outlet	I1: 2421.3 I2: 1234.5 O1: 1256 O2: 1123 O3: 8765	get meterreport error 06

- Step 4: In cases of failure in processing an SMS command, the DTU will send the response message with below error codes as shown in Table 3.6 and goes to exit

Table 3.6: DTU shall send the response message with any of the following error codes.

DTU error code	Description
Success	Sent as response to successful command execution.
Failed	Sent as response if DTU is failed to execute the command.
01	Invalid SMS parameter
02	Invalid SMS Command.
03	Connection to Server Failed.
04	Failed to save in the Database
05	Failed to send data to server even after retries
06	Water meter communication failure
07	Incomplete parameters

- Step 5: Checks SMS processing?. If Success goes to Step 6 other wise goes to Step 4.

- Step 6 : Updates DTU. The DTU will send acknowledgment message to the

Configured mobile only after fulfilling all the action conditions for that command and goes to next Step.

- Exit: goes to next module.

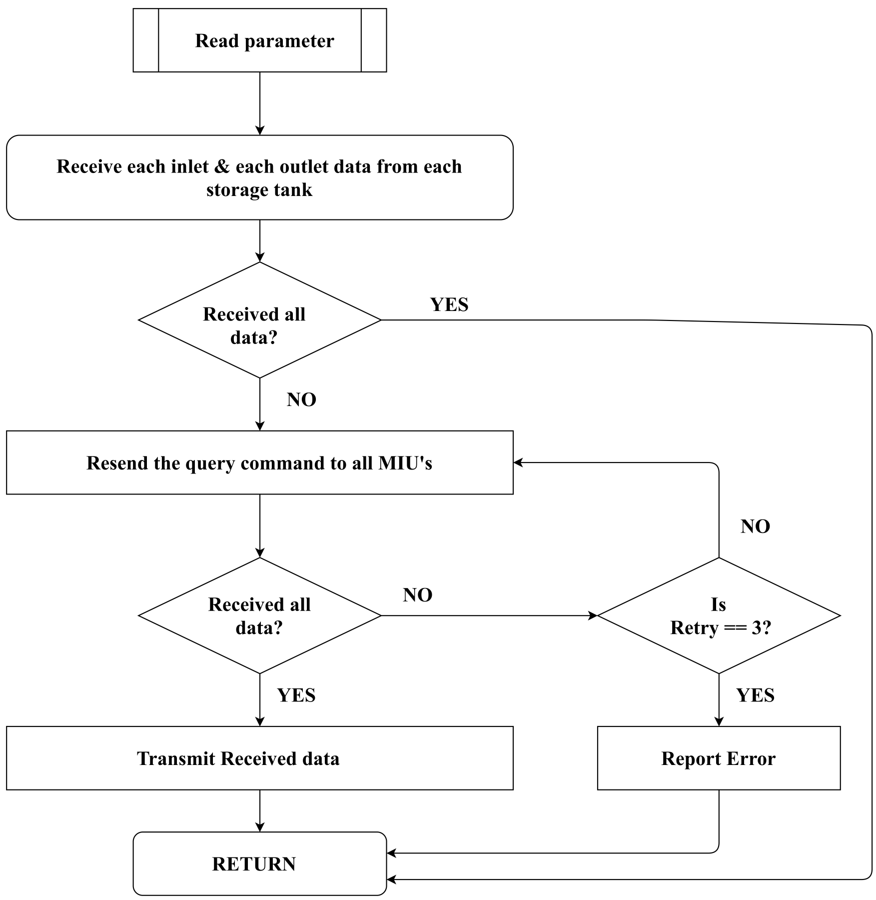


Figure 3.7: Flow chart for reading commands from MIU to aggregator

Algorithm for reading commands from MIU to aggregator as shown in Figure 3.7 is listed below:

- Step 1: MIU consecutively read water meter parameters from both Inlet and Outlet by receiving commands from aggregator as shown in Table 3.7.

- Step 2: Checks the received data?. If success goes to Return other wise goes to Step 3:

- Step 3: Resends the command to all MIU's and goes to Step 4.

- Step 4: Checks the received data?. If success, transmit the data to aggregator and goes to Return otherwise retry 3 times if no response is received from Aggregator and send error report and goes to Return.
- Return: Goes to next module.

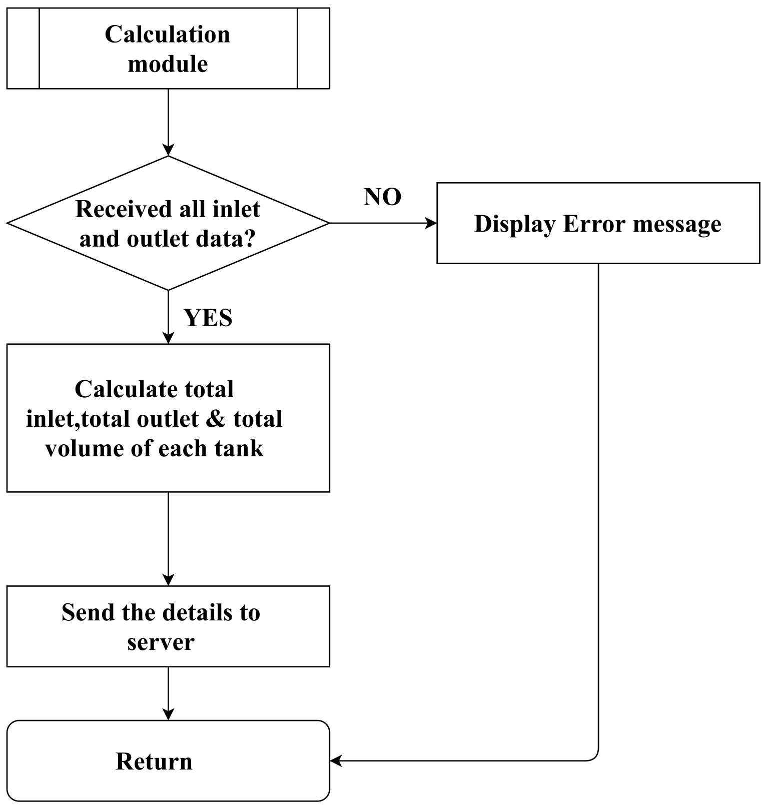


Figure 3.8: Flow chart for calculation module

Algorithm for Calculation module as shown in Figure 3.8 is listed below:

- Step 1: DTU shall store all previous readings in the non-volatile memory and checks whether received all data. If success goes to Step 2 other wise goes to Step 3.
- Step 2: Calculates the equation from 3.1 to 3.7 and goes to Step 4.
- Step 3: Display the error message and goes to Return.
- Step 4: Send the results to the server and goes to Return.
- Return : Goes to next module.

3.5 COMMUNICATION BETWEEN DTU AND ACE WEBSERVER

The DTU will communicate with ACE server over GPRS and establish HTTP POST connection with the server to send data. If GPRS connection is not available, then DTU shall retry three times to connect. If it is not connected, then error connection will be sent. It will send three types of message packets to the web namely,

3.5.1 Device Identification

Device Identification message will be sent to server after successful configuration of the application ID or Device type or DTU ID through SMS command from the configuration mobile.

Table 3.7: Pre-defined MIU Address table

MIU	ADDRESS
Inlet 1	0x00000001
Inlet 2	0x00000002
Inlet 3	0x00000003
Inlet 4	0x00000004
Inlet 5	0x00000005
Outlet 1	0x00000006
Outlet 2	0x00000007
Outlet 3	0x00000008
Outlet 4	0x00000009
Outlet 5	0x0000000A

- Request Byte Stream Format
  3CA1000000011310000000000000014C02EF0F001000100000002 5B3E

- Byte Stream Format Split View
  3C|A100|00|00|01|13|1000000000000014|C02E|F0|F001|0001|00000002|5B|3E

Table 3.8: Device Identification- Request Byte Stream Format

Message header	**Description**	**Length (bytes)**
3C	Start byte (constant)	1
A100	Message code (constant)	2
0	Status, 0 – Success, >0 – Error	1
0	Indicates the total number of responses that shall follow the current message stream	1
1	No of parameters in this message (constant)	1
13	Total number of bytes following this (constant)	1
1E+15	Device ID	8
C02E	Device type	2
F0	Constant Value for Device Identification Data Packet	1
F001	Constant	2
1	Parameter code (constant)	2
2	Application ID	4
56	FCS (ˆoperation on all message heads excluding the Start and Stop Bytes)	1
3E	Stop byte	1

3.5.2 Water meter parameters

Water meter parameters shall be sent to server on expiration of web reporting period.

- Request Byte Stream Format
  3CA10003000D0B07DC0F05050007371000000000002001C02E01F02000010002000B
  000C00150018001B000043210000123400004563000032140000231300000ACFD000
  A4321xx3E

- Byte Steam Format Split View
  3C|A100|03|00|0D0B07DC0F0505|00|07|37|1000000000002001|C02E|01|F020|0001|0002
  |000B |000C|0015|0018|001B|000043210000123400004563000032140000231300000ACFD000
  A4321xx3E

Table 3.9: Water meter parameters - Request Byte Stream Format

Message header	Description	Length (bytes)	Type
3C	Start byte (constant)	1	Int
A100	Message code (constant)	2	Int
3	Status: 3 { Message with Retry counterand Timestamp	1	Int
0	Retry counter	1	Int
0D0B07D C0F0505	Time stamp in format DDMMYYYYHHMMSS (all in hex)	7	Int
0	Indicates the total number of responses that shall follow the current message stream	1	Int
7	No of parameters in this message = 4+ Inlet 1+ Inlet 2 + Outlet 1 + Outlet 2 and so on .	1	Int
37	Total number of bytes following this (Sum of bytes from Device ID to Meter Alarm status value)	1	Int
1.00E+15	Device ID	8	Int
C02E	Device Type	2	Int
1	Constant	1	Int
F020	Constant	2	Int
1	Parameter code for Inlet 1	2	Int
2	Parameter code for Inlet 2	2	Int
And so on.			
3	Parameter code for Outlet 1	2	Int
4	Parameter code for Outlet 2	2	Int
And so on.			
5	Parameter code for Total Inlet	2	Int
7	Parameter code for Total Outlet	2	Int
8	Parameter code for Net Volume	2	Int
101	Parameter code for Meter Alarm value	2	Int
004A1C44	Inlet 1	4	Float
0000DE44	Inlet 2	4	Float
And so on.			
0080EFD0	Outlet 1	4	Float

Table 3.9 continued from previous page

Message header	Description	Length (bytes)	Type
00088B80	Outlet 2	4	Float
And so on.			
36	Total Inlet	4	Float
3296	Total Outlet	4	Float
0000FF69	Net Volume	4	Float
0	Meter Alarm status value	1	Int
8F	FCS (XOR ˆoperation on allmessage heads excluding the Startand Stop Bytes)	1	Int
3E	Stop byte	1	Int

3.5.3 Water meter communication

Water communication failure message shall be sent to server when DTU failed to read parameters from Aggregator or received parameter value with error code. DTU shall wait for the acknowledgment from the server after sending the message. The server sends either a SUCCESS or FAILURE message. The Response sent out by Web server shall be either Success or Failure as shown in Table 3.10 and Table 3.11 respectively. The Response Byte Stream Format shall be as below

- Success: 3C|A100|00|00|A1|3E

Table 3.10: Success Response Byte Stream Format

Message header	Description	Length (bytes)
3C	Start byte (constant)	1
A100	Message code (constant)	2
0	Status: 0 – Success	1
0	Indicates the total number of responses that shall follow the current message stream	1
A1	FCS (ˆoperation on all message heads excluding the Start and Stop Bytes)	1
3E	Stop byte	1

- i) Failure: 3C|A100|02|00|A3|3E

Table 3.11: Failure Response Byte Stream Format

Message header	Description	Length (bytes)
3C	Start byte (constant)	1
A100	Message code (constant)	2
02	Status: >0 – Error	1
00	Indicates the total number of responses that shall follow the current message stream	1
A3	FCS (ˆoperation on all message heads excluding the Start and Stop Bytes)	1
3E	Stop byte	1

3.6 SUMMARY

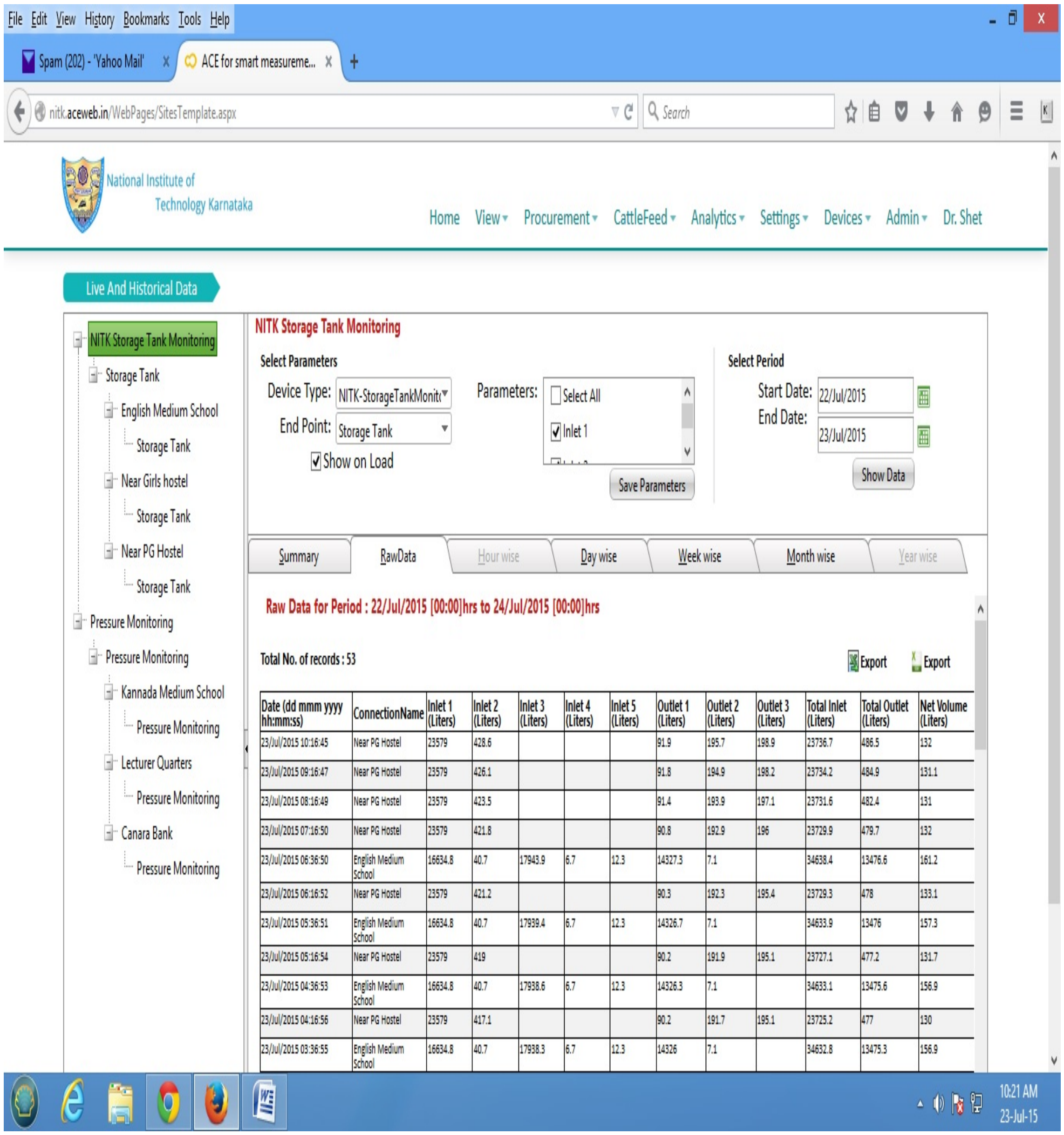


Figure 3.9: Screen Shot displaying water in and out flow details of respective Storage tanks on the ACE web server

The communication in the monitoring unit has been achieved successfully. The connection between DTU, aggregator and web server are working as needed. Here, we have developed a process automation solution that counts the number of pulses from water meter and send to the Aggregator, when Aggregator read command interrupt and also count variable should be initialized to 0 after it reaches to maximum count value or after getting reset command from Aggregator.

The request and response byte stream format for success and failure messages are communicated accurately between RF module,aggregator, DTU and web server. Finally, we are able to display the water flow details from the inlet and outlet and also net volume of the storage tank in the web server as shown in the Figure3.9 . With the available water flow data, we have the following strategies towards data analytics. Since data analytics plays a significant role in computing and actuation part of CPS. We hope that these challenges and issues bring enough motivation for future discussions and interests of research work on CPS.

Chapter 4

COMPUTATION UNIT

Outline of the Chapter

- **Section 4.1** gives the importance of Forecasting in Computation Unit in CPS
- **Section 4.2** summarizing the first impediment of predictive model in computation unit of CPS and explains in detail regarding investigational protocols implied for imputing the data.
- **Section 4.3** largess an in-depth conversation concerning numerous techniques and methods suggested in forecasting and briefs in about the results acquired through comparison graphs and evaluation tables.
- **Section 4.4** delivers an all-inclusive inference and discussion of the research work.

4.1 INTRODUCTION TO FORECASTING

Data is information, Information is knowledge and knowledge is power. Advantages provided by insights derived from data that has been transformed into useful information is vast. Information derived from measurement data can be used to drive optimization and innovation. The biggest challenge lies in effectively managing acquired data to give useful insights to what is otherwise just a pile of data dump.
Following this brief introduction to the significance and challenges of data in the todays world, this Snyder and Burress (2011) will inform the audience on several technologies and techniques that provide a more integrated, scalable approach to data management

and yield the following benefits: Advance insights: descriptive information throughout the data acquisition and processing life cycle makes data traceable, trackable, usable, and logical to store. Searchable: data mining makes it easy to locate and access as and when needed. Management and processing: from analysis to reporting, data management reduces the many barricades in data processing, particularly as data sets grow in size. Researchers can be more efficiently predict the automated result by selecting the correct data storage, data mining, and data processing tools and techniques. This enables faster and more reliable results of experiments and discoveries.
Fernandez-Delgado et al. (2014) evaluate 179 classifiers arising from 17 families (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today.

In our previous work, RF modules were also used to analyze the trend (behavior) of water flow that is used to plan the scheduling of pump, and in future it can be used to detect the leakage and water wastage using different methods. In addition, the data also help us to predict the water consumption for optimizing pumping schedules of each storage tank depending on the water flow rates.
Later, the analytical tools have been deployed in each storage tank to provide a decision support system. Decision support system involves instance predictive model that helps in handling missing data, forecasting the data and control the valve operation. Moreover, it helps in enhancement of network management and planning. This enables administering the water supply network effectively, utilizing minimal wireless technology and ensuring all the storage tanks in the campus to offer a reliable and sustainable water supply within all storage tanks. In this paper, we have discussed about computation unit of CPS that has decision support system. The decision support system is used to forecast the water flow rate in a storage tank. Water demand forecasts are utilized by water authorities and water service providers for the planning, designing the infrastructure, operating and managing water supply systems depending on the precise goals; different forecasting time frames are utilized and diverse forecasting methods are applicable Lopez Farias *et al.* (2018).
Water demand forecasting is basically regular, real-time domain demanding temporal

calculations that may vary on an hourly basis on the micro-level, wherein the macro level or long-term forecasting parameters such as weather need to be incorporated. The objective of such forecasting is to ensure water management and load distribution towards costs and resource optimisation Alvisi *et al.* (2007) in water-rich regions such as the European –North or scarcity regions such as near the Mediterranean region. Chen and Boccelli (2014) The demand calculations involve short-term as well as long-term variables. The predictive effort in water-demand becomes complex because of the randomness and variability inherent the data of end-user requirement as well as in the weather or climatic conditions. The Chen and Boccelli (2014) reveals both short-term as well as long-term predictive calculation methods to match demand and supply. Time series analysis provides for such forecasting within a resolution of as less as 15 minutes using conventional systems such as SCADA, (Supervisory Control and Data Acquisition) that is based on statistical framework. Seasonal ARIMA AutoRegressive Integrated Moving Average) is a classical Probabilistic time series model used in projecting water demand over varying timescales (half-hourly: Sterling and Bargiela (1985) over a 48 day period; daily demand forecasting: Jowit and Xu (1992); and Caiodo (2009) for weekly and yearly bases to forecast daily demand) [2]. The limitation of ARIMA used in abovementioned cases is the lack of matching per capita use with resource. The variables not taken into account are population changes, lifestyle changes, town planning measures, environmental warming, and similar factors that affect long-term projections.
In Studies Bakker *et al.* (2014), have found that using end-use water demand for forecasting is an efficient framework. However, the variations in predictions have been observed to be error-prone when weather conditions are taken into account. The current paper evaluates the performance of three forecasting models with and without weather input. The models under investigation are Multiple Regression model, transfer noise model, and Adaptive heuristic Model. The results of the tests reveal that using the weather input as an input has a correction value of 11% on the maximum and 7% on the average forecast over the single input demand data.

Forecasting plays an important role in CPS. Forecasting has become the necessary action to complete the decision-making activities in CPS. Forecasting is a statistical task, where it guides to inform the decisions about scheduling and strategic planning. Three types of forecasting can be done, namely, short-term forecasting, medium-term forecasting and long-term forecasting (Jain *et al.* (2001)). Depending on the type of application, specific forecasting technique has to be chosen. For instance, planning for large-scale investments typically requires long term forecasting, at a yearly or monthly resolution with limited accuracy. Conversely, for optimum allocation of water supply among users, medium term forecasting at a monthly or weekly resolution with somewhat higher accuracy might be required (Rinaudo (2015)). A selection of water demand forecasting methodologies suitable for such purposes exist, comprising of unit use/end-use models, regression models, time-series analysis, etc. Forecasting can be done for future demands that require several years in advance or only a few minutes depending on the specification of the application. In our research work, we develop a forecasting system that involves decision support systems that predicts the events based on historical events. Here, we have selected appropriate forecasting methods, evaluating and refining them with respect to time.

In (Ahmed *et al.* (2010)), an outsized scales assessment study in chief machine learning models for time series forecasting is offered. Models reflected in this research work are Bayesian Neural Networks, multilayer perceptron, Radial Basis Functions, generalized regression neural networks (also called kernel regression), K-nearest neighbor regression, support vector regression, CART regression trees, and Gaussian processes. The study discloses momentous dissimilarities amongst models and correspondingly that pre-processing will have a substantial influence on enactment. Pre-eminent approaches curved out to be the multilayer perceptron and the Gaussian process regression. Apart from model assessments, numerous pre-processing approaches are also verified, and outcomes reveal that they have numerous influences on performance too. A research study in (Smith and Agrawal (2015)) dedicates on demonstrating possibilities to form technology forecasting models by means of patent groups. Concentration is on smearing time series modeling practices to an assortment of USPTO patents from 1996 to 2013. Holt-Winters Exponential Smoothing and ARIMA procedures are implied for this study. Cross Authentication approaches are also executed in governing regarding the finest fitting model, and eventually, a patent data might be exhibited as a time series or not. Literature assessment study in (Khashei and Bijari (2012)) platforms saying that numerous large-scale forecasting competitions with huge quantity and frequently castoff time series forecasting models determine that merging forecasts from more than one model every so often clues to enhanced performance, particularly the minute when the models in the communal are moderately not the same. In this broadsheet, a novel technique is projected to build a hybrid model by means of a time

series model as basis model and a probabilistic neural network as a classifier model. Experimental fallouts with three well known real data sets point towards the recommended technique, which is an operational approach to build an additional precise hybrid model compared to basis time series model.
The initiative in (Makridakis *et al.* (2018)) is to assess routines through numerous forecasting horizons through a large subset of 1045 monthly time series used in the M3 Competition. In this study, results are conversed, clarification regarding why the accuracy of ML models is underneath that of statistical ones are specified, and also certain likely techniques forward are proposed. ML approaches must be extra precise, necessitating a smaller amount of computer time. Foremost involvement of this paper is in presenting that traditional statistical methods are more precise compared to ML methods and also sockets out the necessity to ascertain motives in it, and simultaneously formulating methods to converse the circumstances.
The UK Environment Agency and the California Bay-Delta Authority suggest planners to use hybrid tools (an ensemble of choice of the following: Temporal Extrapolation, Unit Water Demand, Multivariate Statistical Modelling, Micro-component Modelling and Land Use Based Model) Christian-Smith *et al.* (2012). Another study Kourentzes *et al.* (2014), reports that an ensemble of neural networks performs better in predictive analysis than a single 'best-fit" neural model, owing to the fact that the Mean operator commonly used in NN based tools is an inefficient parameter and that using Mode has been found to provide more accuracy using lesser number (25% to 300% lesser) of models in ensemble. The application of water demand forecasting methodology used by in Barcelona Water Distribution Network reveals a multi-model approach that combines temporal time-series data qualitatively as well as quantitatively to deliver the most accurate results. The Nearest Neighbour (NN) classifier using the Mode operator of an ensemble is operated in conjunction with the daily demand forecaster. The model thus developed was called QMPP+ (Qualitative Multi-Model Predictor Plus), which when compared to ARIMA and DSHW (Double Seasonal Holt-Winters) improved the forecasting efficiency Lopez Farias *et al.* (2018). Multi modelling approaches are known to work efficiently and accurately in other human activities, too, the study reports.

4.2 HANDLING MISSING VALUE

Water forecasting is crucial for planning, designing the infrastructure, and also for operating and managing water supply systems. Forecasting in the computation unit plays a very significant role in Cyber-Physical System. Real-time monitoring of water flow rates information helps us to conserve water when it is needed the most. Here, we are considering the first impediment of predictive model that is handling missing data in the data sets.

Water demand forecasting involves not only immediate human activity estimation but also variables such as climatic and weather variations, land use planning implements amongst other factors such as resources and its allocation. Data used for forecasting is accessed through historical encapsulation (Calendar data) as well as through sensors that provide real-time inputs. Real-time Missing data occur in water management systems either due to sensor faults or due to lapses in communication systems. Estimation of Missing data from either sources is under active research in all fields of human activity towards innovating efficient forecasting tools.

Imputation generally involves calculations using missing-indicator method, overall mean imputation, and complete case analysis. These have, however been observed to deliver bias. Variations of imputing methods can involve multiple imputation calculations to avoid bias in single imputation method. The current paper establishes the biases observed in Missing-Indicator and overall Mean Imputation methods Donders *et al.* (2006).

The mitigation methodologies involved consider two approaches: conventionally, imputation techniques are used to find the most accurate estimated value, whereas recent developments ignore the missing values and train the predictor on available values. In Imputation methodology, the predictor is supplied with the completed vector, where the missing value is calculated using statistical tools in the instance vector. In the Ignore methodology, the missing value is not taken into account, and the reduced vector is fed to the predictor to execute forecasting process. Imputation technique using the mean, k-nearest neighbour (k-NN), and the regression are the ones most commonly used in practice Zhang *et al.* (2013). Ignore techniques mainly either delete entire training sets of corresponding data with missing entries or use and Ensemble treatment that trains the classifier for each missing value possible and deploy corresponding training model when encountered. Thus ensemble methodology in Ignore has trained treatment for each missing value case. However, the repository required and corresponding storage demands increases with increase in the real-time or calendar input variables.

In Nelwamondo *et al.* (2007), two recently developed imputation techniques, EM (Expectation Maximisation and ANN (auto-associative Neural Network) in conjunction with Genetic Algorithms. The current paper evaluates the techniques and compares the characteristics to help users decide the applications and conditions under which each offers better results. The evaluation was done on data accessed from disparate fields: HIV sero-prevalence data, Industrial wire-winding application, and industrial power plant data. The EM performed better in cases where input variables were not related to or are dependent on each other, whereas the ANN and Genetic algorithm combination produced better results for calculating missing values' imputation where the input variables indicate interdependency of correlation of any magnitude.

In this work Laña *et al.* (2018) the traffic forecasting was considered for evaluation and development of techniques to overcome missing data instances. The evaluation considers both high and low frequency missing data instances. Further the evaluation considers real-time context as well as historical data for evaluation as a further dimension of evaluation of the the techniques under investigation. This work also considers missing data from sensors as well as that from external factors to provide users with better characteristics when dealing with either of the data sets or when in combination.

In this study, we collected and utilized a dataset from Bangalore Water Supply and Sewerage Board (BWSSB),Government of Karnataka, India in addition, analyzed water flow information values repeatedly at every 15 minutes interval . Before building the model, it was critical to pre-process the data sets received from the BWSSB in a proper method. With the data available, we were required to replace the missing data by adjusting it with the algorithms, so that data become constant and have some regularity to it. In a predictive model, the first obstacle to be contemplated as missing values in the datasets. When the missing values are confirmed in the datasets, it is very much required to view the data and do further analysis with the data with the standard assumptions of the algorithm. In our experiment setup, we synthetically omitted some missing values in the data sets for the rationale of finding omitted values in the data sets .

4.2.1 Different types of missing data

These methods are used to improve the identification of the absence of the data of each observation (Pires *et al.* (2016)).

- MCAR(Missing Completely at Random): If the likelihood of a data to have missing values is not dependent on either the missing values or the observed data. Data missing due to technical reasons cannot be considered MCAR (Lakshminarayan *et al.* (1999)).The missing value in a dataset does not have anything to do with its assumed value and the values of other attributes.

- MAR(Missing at Random): If the likelihood of data to have missing values is not dependent on the missing values but on the values of the observed data. Data missing due to technical reasons only are regarded as MAR. (Lakshminarayan *et al.* (1999))

- MNAR(Missing not at Random):MNAR is data that is not MCAR or MAR. If the likelihood of data to have missing values is dependent on the values in the dataset. To extend the previous example under MAR, this would occur if males failed to complete a depression survey because of their level of depression. (Lakshminarayan *et al.* (1999)).

In the first two instances, removing data with missing values depending on their occurrences, is safe, while in the third instance, removing data with missing values may result in a bias in the model. So before removing observations, we have to be really cautious.

4.2.2 Experimental Procedure to handle missing data

A series of 0 to 60 percent of missingness is introduced in the data sets. Zero percent missingness means no data is omitted in the data sets (Zhang *et al.* (2013) Wubetie (2017)). Here we introduced two types of missingness in the data sets i.e., randomly and Burst nature missingness. Randomly, when X percent missingness is introduced in the data sets, each column (attribute) has an X percent chance of value being missed, regardless of its position. The values are deleted randomly. Likewise in Burst nature,

when X percent missingness is introduced in the data sets, each column (attribute) has an X percent chance of value being missed, endlessly from any of its position. The values are deleted constantly with respect to X percent is known as Burst nature. The data in the data sets are in agreement with the MAR principle. The performance of the algorithm method on the full data set is used as a reference point for comparing it with the algorithm for the missing data sets.
Here we are examining daily model, we will consider as each day from respective month in each column and perform the experiment. Ultimately, we compare the daily model using performance metrics for diverse imputation techniques. In the Figure 4.1, the experimental model set up is shown.

Algorithm for imputation set up for Instantaneous (one step imputation) shown in Figure 4.1 is listed below:

- Step 1: Duplicating the training data sets, keeping original training data set as a reference.

- Step 2: Introducing missing values (5 percentage to 60 percentage) both randomly and burst nature into the testing set.

- Step 3: Impute the missing values using different imputation techniques into the testing data set.

- Step 4: Comparing the original training data set with testing data set with imputed values.

- Step 5: Performance evaluation using RMSE, MAE, and MAPE is noted.

- Step 6: Comparing different imputation techniques and choosing imputation technique with minimum RMS, MAE and MAPE values

The missing value calculation is evaluated by comparing calculations using the data from the computed value table and that of an earlier complete data-set. The arrangement in the data tables follow the rules stated in the Missing at Random principle. Since in-depth elaboration of statistical computation is essential in the processing of accessed data, the tools offered by 'R' package are used in the calculation of missing data. The trials are conducted through intentionally injecting 'missing values' and allowing the different imputation algorithm techniques to compute and predict the

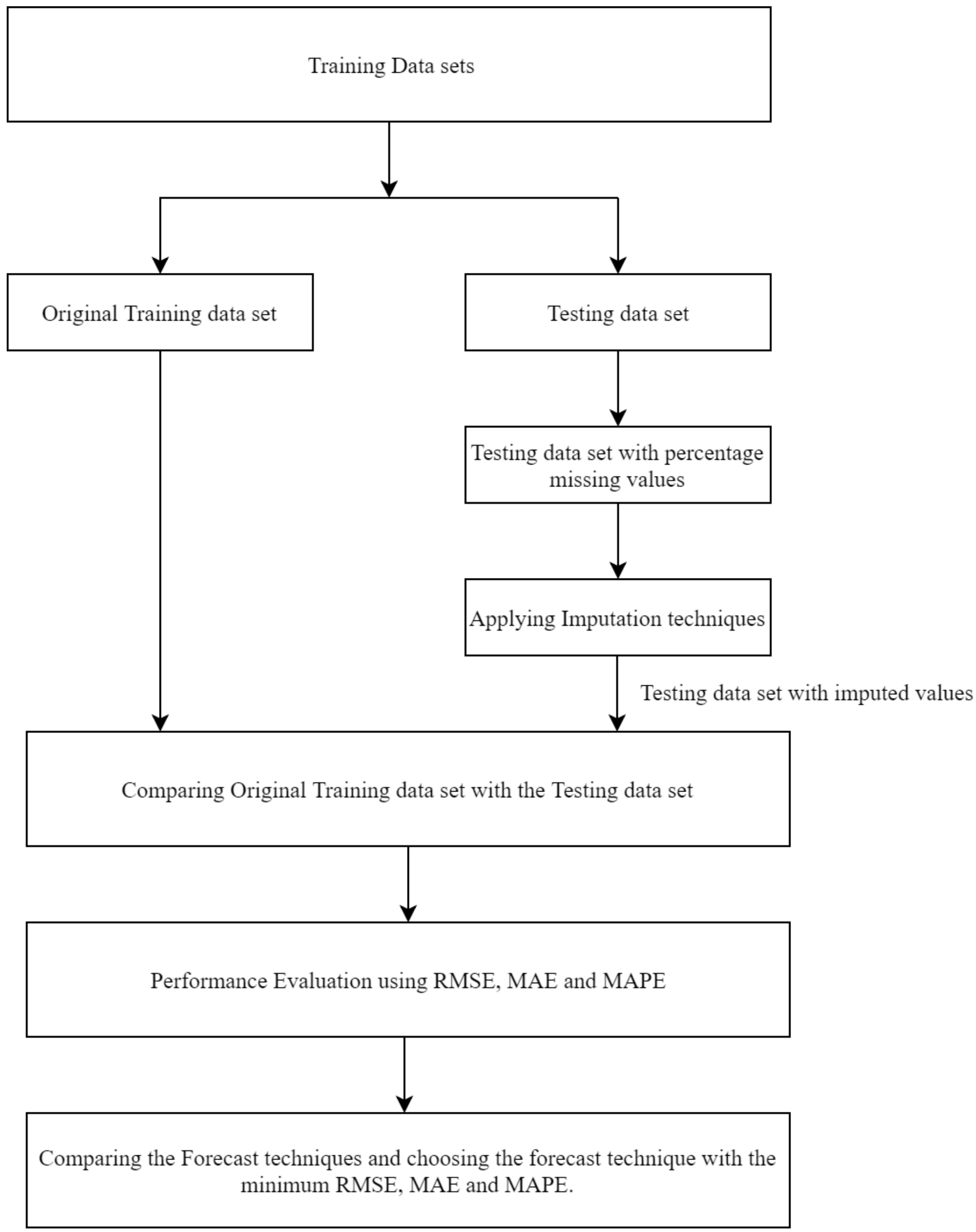


Figure 4.1: Experimental procedure for imputation set up for Instantaneous (one step imputation).

values required. The validation is carried out by computing for a range of 'missingness' varying from 0% to 60%. (Zhang *et al.* (2013)).

4.2.3 Imputation method techniques

Here we are considering the problem of completing the missing values of the water flow rate in storage tanks. The water flow rate is taken in each storage tanks. The composed water flow rate values are assembled into excel format i.e., into a data matrix (In excel format, we consider as data matrix) where each column refers to daily flow rates of each storage tanks and each row refers to time with sampling epoch at every fifteen or thirty minutes.

- Mean- Many functions are added implicitly in R programming tool, to perform statistical analysis. Mean of a data value is a statistical analysis of the middle value of the data i.e., the sum value (water level data entries in the present paper) of all water levels normalized by the number of tanks from which the level values have been recorded. Mean is measured as shown in the equation 4.1

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.1}$$

  Where, xi is the input values, n is the number of values in the data series and M is the mean. Where xi is the input values, n is the no. of values in the data sequence and M is the mean.

- K-Nearest Neighbors (K-NN): The methodology suggested under this premise is to capture the nearest values in the proximal cells of the missing value and computing the average of such values to predict 'missing value'. This form of computation is based on the Gower distance postulate that adds weight to the values based on their proximity to the 'missing cell'(Kowarik and Templ (2016),Templ *et al.* (2011) and Prantner (2011)). Studies rate method enabled to data sets that carry values of different characteristics. In addition, the authors also assert that this method of finding 'missing values' is capable of offering solutions even when the number of missing values is high.
  Perform Imputation of A Data Frame Using K-NN.
  k-Nearest Neighbour algorithm is used to frame the missing data imputation. The model is used for discrete variables, instead of for continuous variables median used.
  "knnimpute(Data) replaces NaN s in Data with the relating an incentive from the closest neighbor section. The closest neighbor section is the nearest segment

in Euclidean separation.
knnimpute(Data, k) replaces NaN s in Data with a weighted mean of the k closest neighbor sections."

- Expectation-Maximization (EM): EM algorithm is a method in which statistical values to be computed are measured (imputed) if a complete dataset is available, taking into account missing data patterns. In this method, values are not generally imputed for individual missing data items. In statistics, the EM is an algorithmic approach to finding maximum likelihood estimates of parameters in statistical models where the model depends on unnoticed latent variables. The EM iteration alternates between expectation (E) step that creates a function for the log-likelihood that is evaluated using the current parameter estimate and maximization (M) step that computes parameters to maximize the expected log-likelihood found in the (E) step. These parameter estimates are then used in the next (E) step to evaluate the distribution of the latent variables.
  EM imputation is the better way for finding the singe imputing values and it's also used to find all the missing responses, which is slightly hard than mean imputation. It is an iterative system in which it utilizes different factors to ascribe esteems (Expectation), at that point checks whether that is the esteem in all likelihood (Maximization). If not, it re-credits more probable esteem. This goes on until it achieves the no doubt esteem (Honaker *et al.* (2010)).

- Matrix Completion (MC): Matrix completion is the process of filling out missing matrix entries of an incomplete matrix. A large collection of datasets is naturally organized in matrix form. (Pillai *et al.* (2012)) explain this method as a low-rank approximation technique through nuclear normalization to impute desired missing values. (Hastie *et al.* (2015)) This process is iterative and similar to the EM method in trying to explore the probability of achieving the utmost proximity to the expected value.

4.2.4 Performance evaluation of an Imputation technique

The performance evaluation of these data sets helps us understand the condition when there is a shortage of water and excess of water supplied to diverse areas. With these details, we can arrive at a decision when water is required and when water should be pumped. Further research is needed to resolve the water scarcity issue.

The performance and efficiency of Imputation Techniques using R software (that are basically predictive and probabilistic in the procedural inquiry) are tested by the following error measures: "RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error)". These measures are accepted in scholastic pursuits as delivering a just comment on the validity of specific technique utilized for accusation under a defined environment (Zheng *et al.* (2015), Wu *et al.* (2017) and Herrera *et al.* (2010)). The accuracy of the implemented technique is tested against three models defined for the purpose of testing- Overall, monthly, and weekly model.

As stated earlier, the randomly generated missing values increase linearly from 00 to 60 percentage, and the comparison of the stated Error measures (RMSE, MAE, and MAPE) serves as the base of the quality of the imputation method selected. The performance contrast for each model with the best accuracy is as shown in the Figures 4.2 to 4.4.

- RMSE: (Torgo and Torgo (2013), Nookhong and Kaewrattanapat (2015)) The Root Mean Square Error is an indicator of the proximity of the imputed value to the actual desired value. If RMSE is zero, then the imputed value matches the missing value, and an accurate method for the application; as the indicator (error) goes on increasing, the imputed value moves farther away from the actual missing value, indicating a weak tool. It is used to test the accuracy of the technique deployed to find the missing value of a data set. RMSE is measured as shown in the equation 4.2

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(Y_i - \hat{Y}_i)^2}{n-1}} \tag{4.2}$$

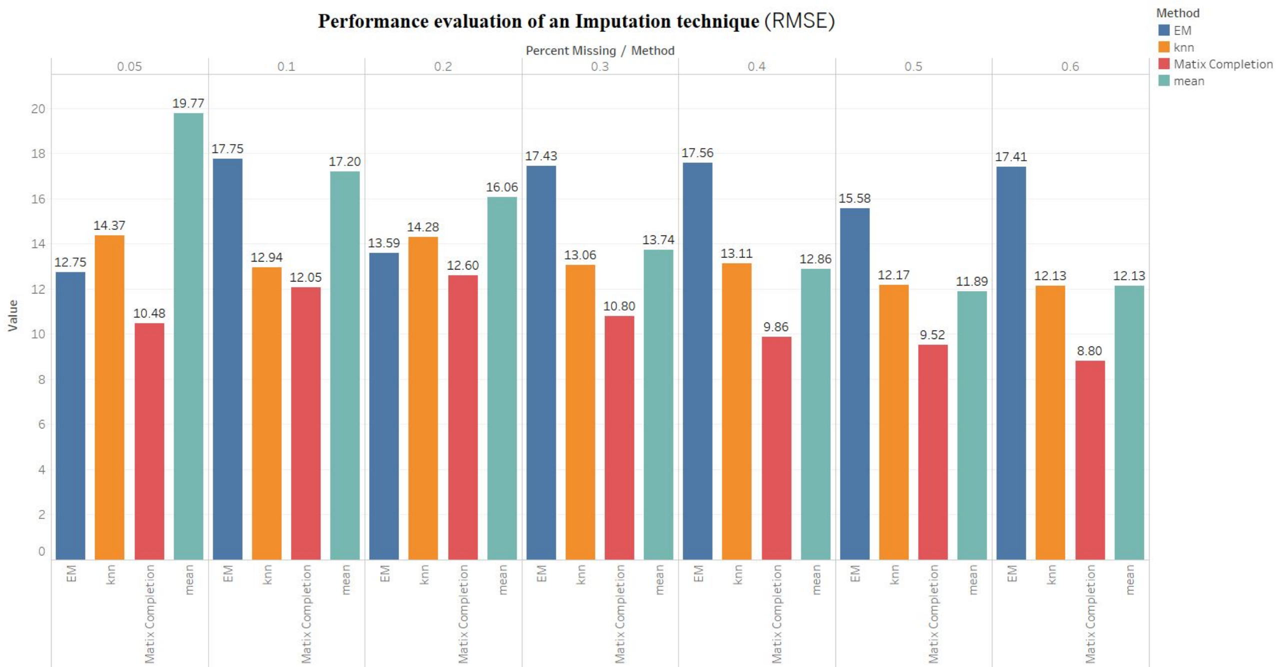
As: Y_i= Approximation from data value model from forecasting, $\hat{Y}_i$ = Actual value of actual data obtained from calculation, n= Number of sample size using in model estimation (Nookhong and Kaewrattanapat (2015), Torgo and Torgo (2013)) .

Table 4.1: Comparing performance evaluation (RMSE) of imputation techniques (a) Random

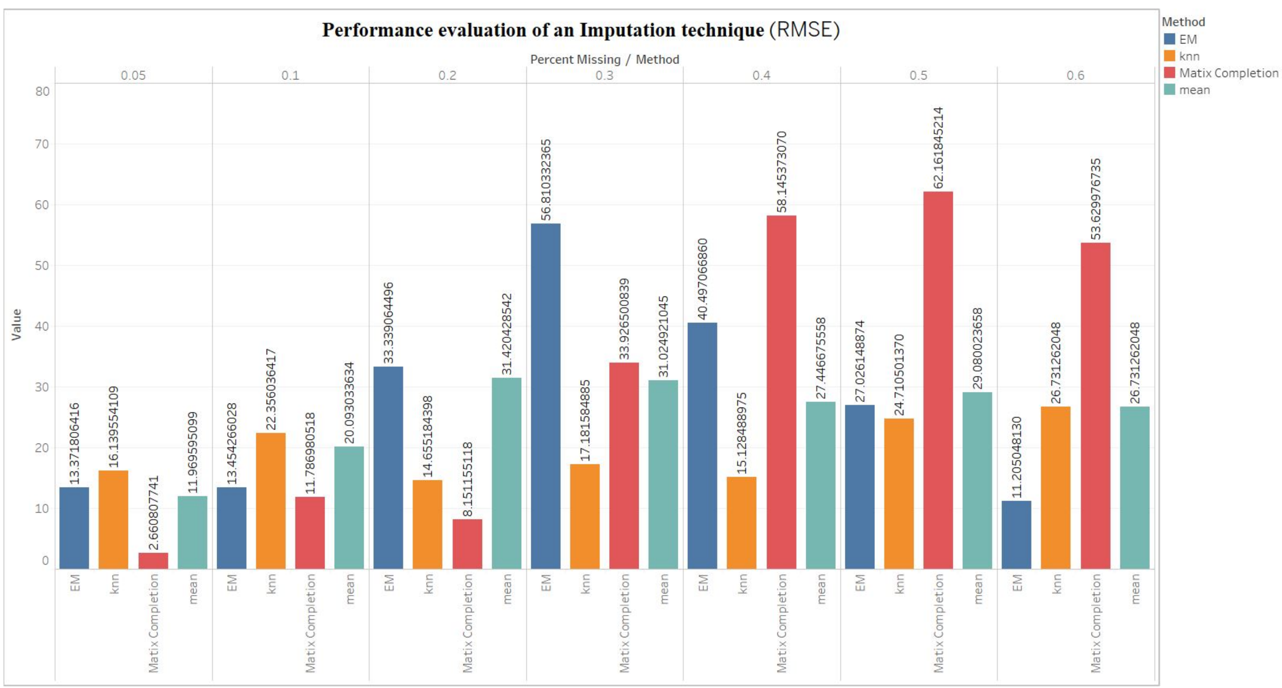
RMSE							
Methods	**Missing Percentage (Randomly)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	14.36781	12.93671	14.28008	13.06095	13.11013	12.17254	12.1282
Mean	19.77244	17.2027	16.06107	13.73778	12.85795	11.88778	12.1282
EM	12.74834	17.7518	13.5912	17.43311	17.56159	15.57518	17.40782
MC	10.47532	12.04623	12.60041	10.80096	9.85614	9.521413	8.802164

Table 4.2: Comparing performance evaluation (RMSE) of imputation techniques(b) Burst

RMSE							
Methods	**Missing Percentage (Burst)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	16.13955	22.35604	14.65518	17.18158	15.12849	24.7105	26.73126
K NN	11.9696	20.09303	31.42043	31.02492	27.44668	29.08002	26.73126
EM	13.37181	13.45427	33.33906	56.81033	40.49707	27.02615	11.20505
MC	2.660808	11.78698	8.151155	33.9265	58.14537	62.16185	53.62998



(a) Random



(b) Burst

Figure 4.2: Comparing performance evaluation (RMSE) of imputation techniques (a) Random (b) Burst

- MAE(Mean Absolute Error): MAE measures the average of differences between the estimated value and the real value for multiple instances (Torgo and Torgo (2013) Nookhong and Kaewrattanapat (2015)). The error indicator reflects the output similarly as the RMSE. If the value of MAE is zero then the imputed value matches the real value. MAE can be measured as shown in the equation 4.3

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \tag{4.3}$$

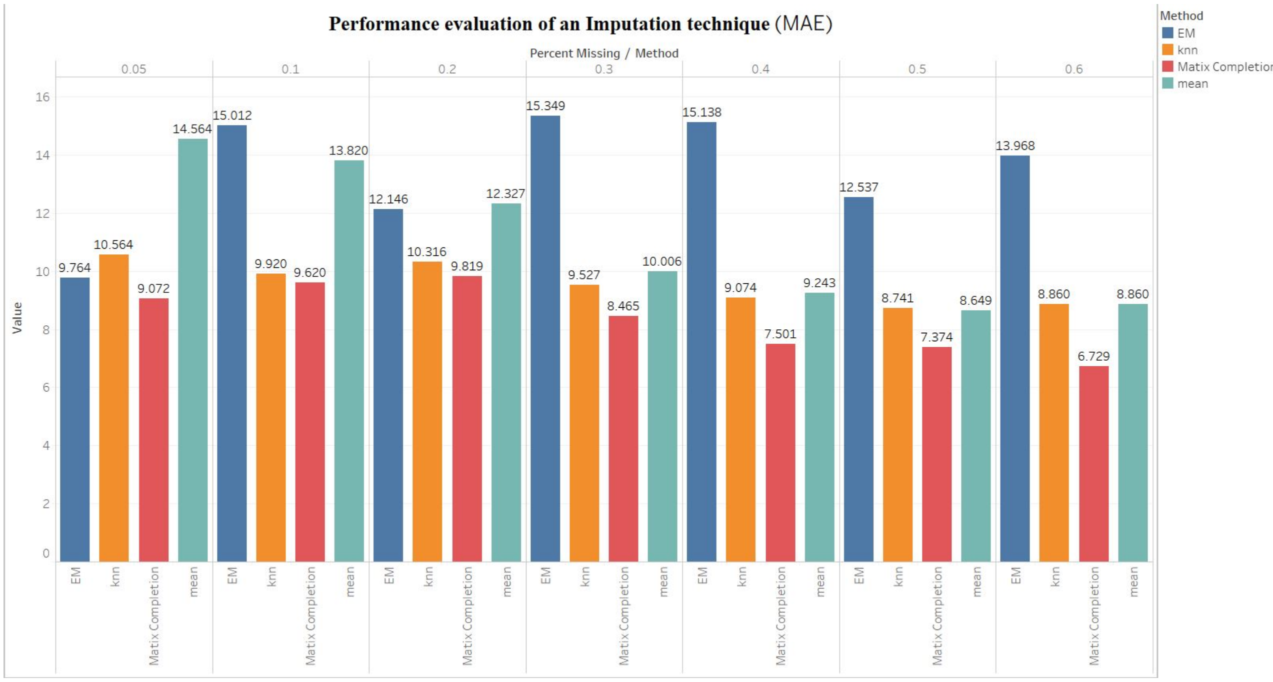
As, f_i = Approximation from data value model from forecasting , y_i = Actual value of actual data obtained from calculation , n = Number of sample size using in model estimation.(Nookhong and Kaewrattanapat (2015),Torgo and Torgo (2013)).

Table 4.3: Comparing performance evaluation (MAE) of imputation techniques (a) Random

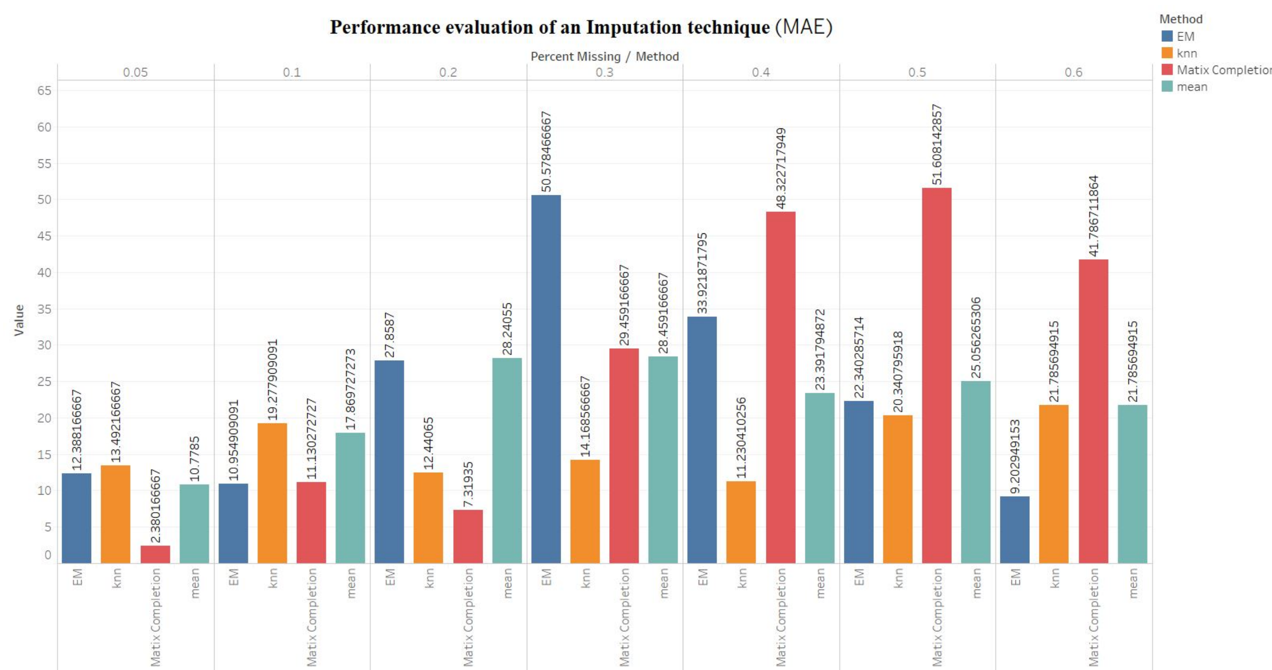
MAE							
Methods	**Missing Percentage (Randomly)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	10.5638	9.9204	10.31595	9.526586	9.074342	8.740938	8.860086
Mean	14.5638	13.8204	12.32732	10.00555	9.242658	8.649354	8.860086
EM	9.7638	15.0117	12.14579	15.34914	15.13842	12.5366	13.96753
MC	9.072	9.6204	9.818789	8.465483	7.5005	7.374313	6.728638

Table 4.4: Comparing performance evaluation (RMSE) of imputation techniques (b) Burst

MAE							
Methods	**Missing Percentage (Burst)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	13.49217	19.27791	12.44065	14.16857	11.23041	20.3408	21.78569
K NN	10.7785	17.86973	28.24055	28.45917	23.39179	25.05627	21.78569
EM	12.38817	10.95491	27.8587	50.57847	33.92187	22.34029	9.202949
MC	2.380167	11.13027	7.31935	29.45917	48.32272	51.60814	41.78671



(a) Random



(b) Burst

Figure 4.3: Comparing performance evaluation (MAE) of imputation techniques (a) Random (b) Burst

- MAPE :(Torgo and Torgo (2013)) MAPE is a process to check the accuracy of a prediction of forecasting methods in statistics . It is typically calculated as the absolute value for every difference between the actual value and forecasted value and divided by the actual value with respect to time. Multiplying by hundred and dividing it by the figure of values in the data makes it a percentage error. The MAPE (Mean Absolute Percent Error) measures the extent of the blunder in rate terms. It is determined as the normal of the unsigned rate mistake, has appeared in the precedent underneath. MAPE is measured as shown in the equation 4.4

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - i_t}{A_t} \right| \tag{4.4}$$

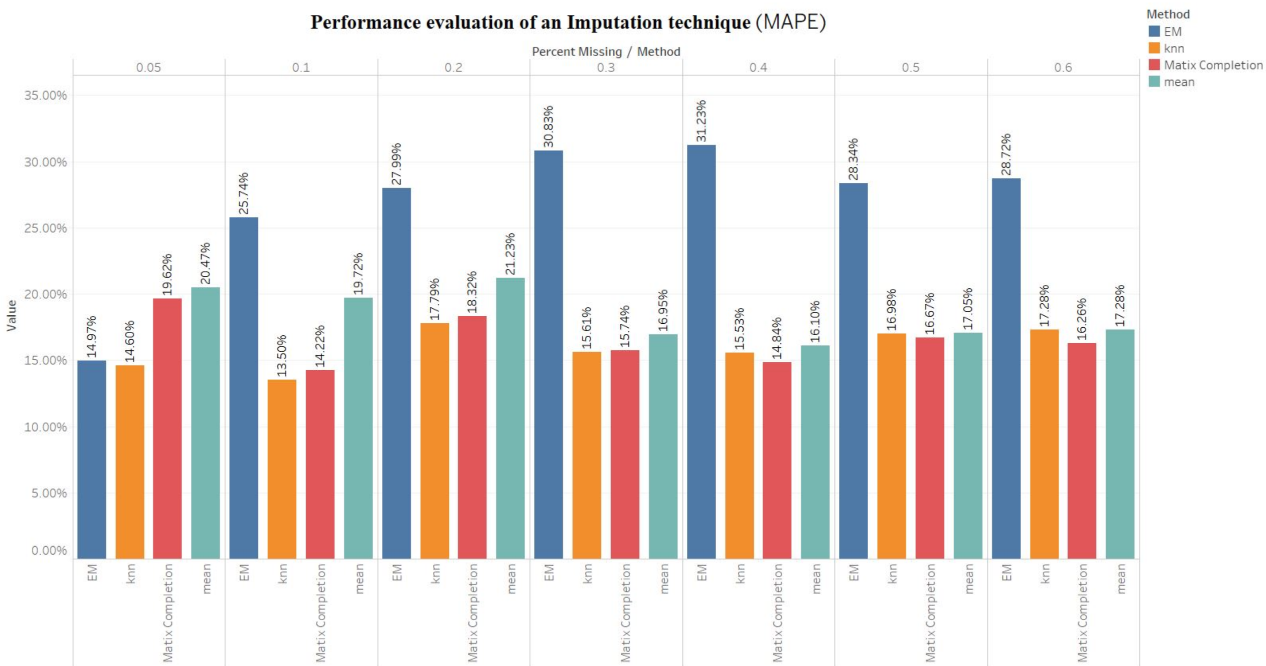
As, A_t= Actual Value,
i_t= imputed value,
n= Number of values in the data sets.Torgo and Torgo (2013).

Table 4.5: Comparing performance evaluation (MAPE) of imputation techniques (a) Random

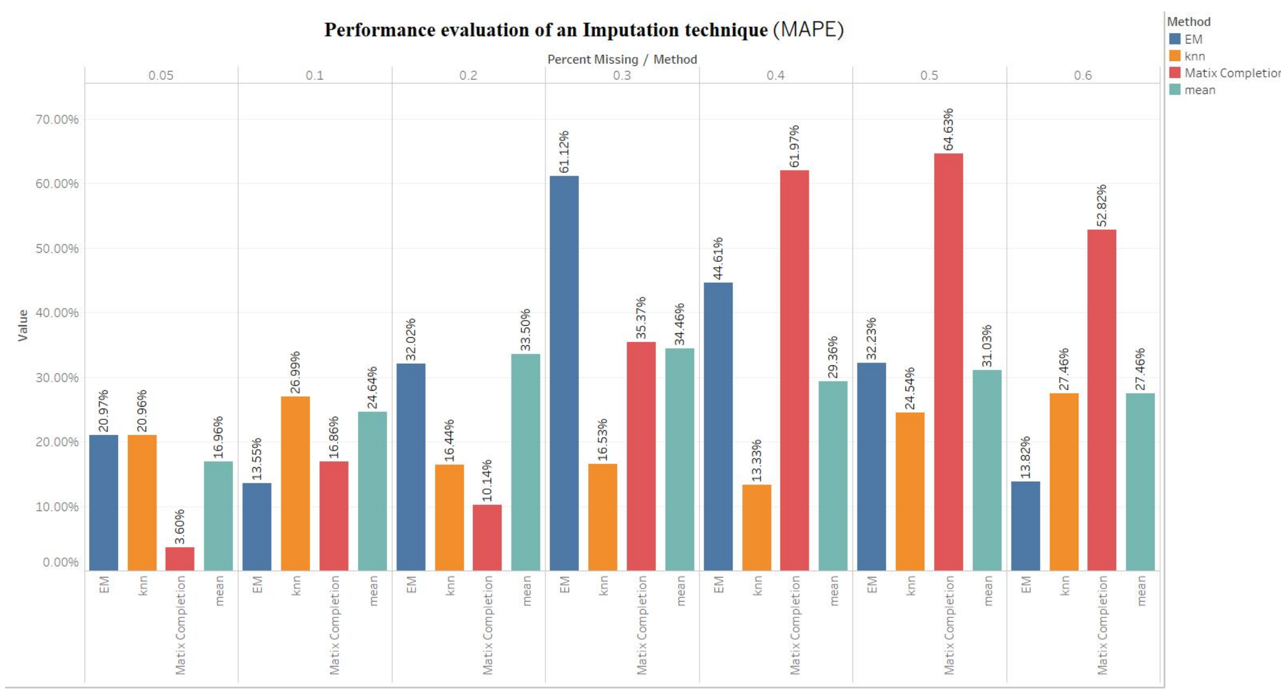
MAPE							
Methods	**Missing Percentage (Randomly)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	0.145999	0.135015	0.177893	0.15605	0.155347	0.16979	0.172782
Mean	0.204689	0.197152	0.212264	0.169497	0.161033	0.17053	0.172782
EM	0.149696	0.25742	0.279875	0.308284	0.312343	0.283405	0.287193
MC	0.196152	0.142209	0.183245	0.157353	0.148369	0.166713	0.162638

Table 4.6: Comparing performance evaluation (RMSE) of imputation techniques (b) Burst

MAPE							
Methods	**Missing Percentage (Burst)**						
	5%	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
K NN	0.209616	0.269885	0.16441	0.165296	0.133301	0.245399	0.27456
K NN	0.169606	0.246371	0.33497	0.344556	0.293558	0.310283	0.27456
EM	0.209707	0.135514	0.320189	0.611168	0.446067	0.32226	0.138219
MC	0.036032	0.168622	0.101399	0.353688	0.61967	0.646271	0.528231



(a) Random



(b) Burst

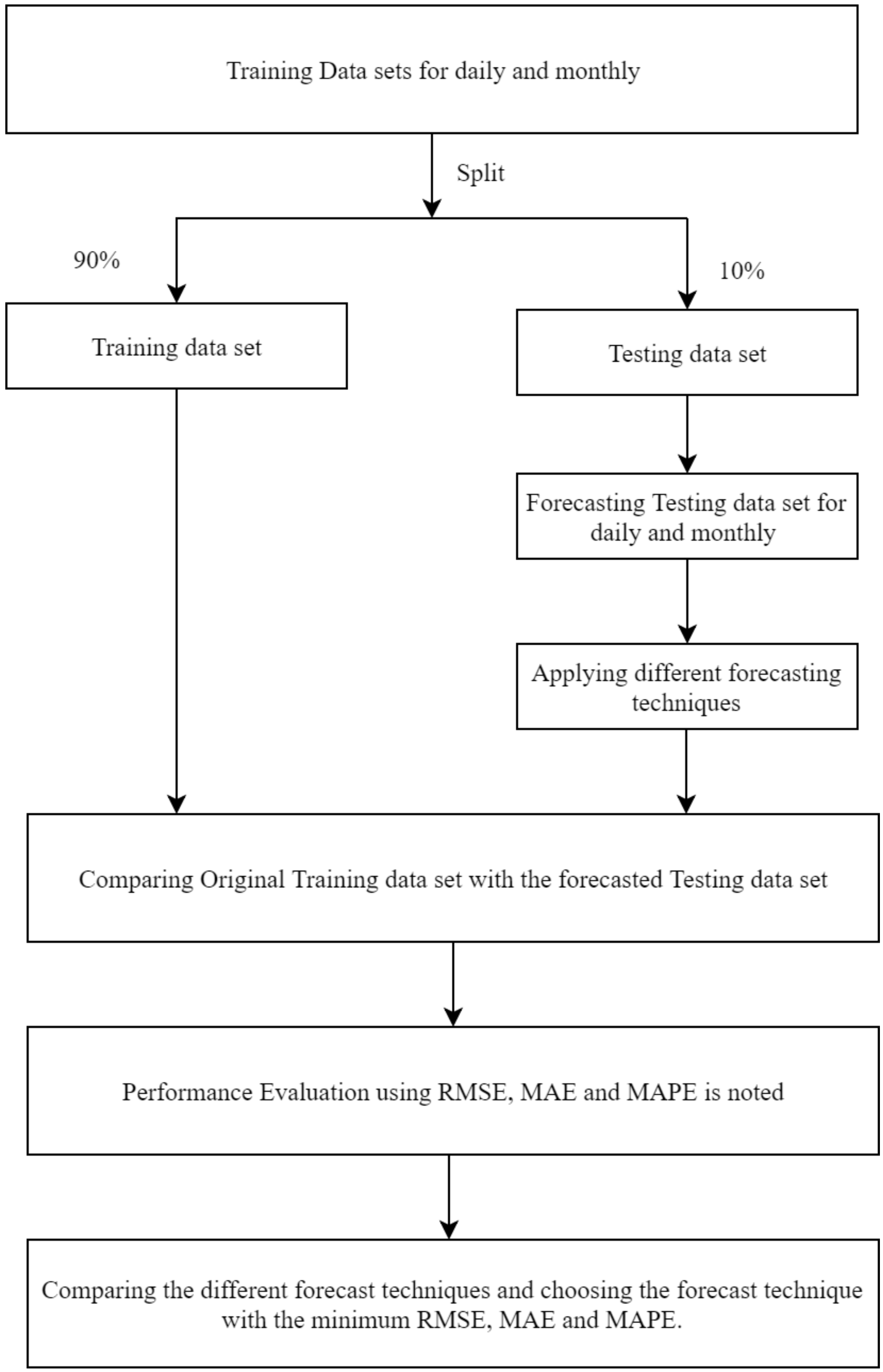
Figure 4.4: Comparing performance evaluation (MAPE) of imputation techniques (a) Random (b) Burst

4.2.5 Summary

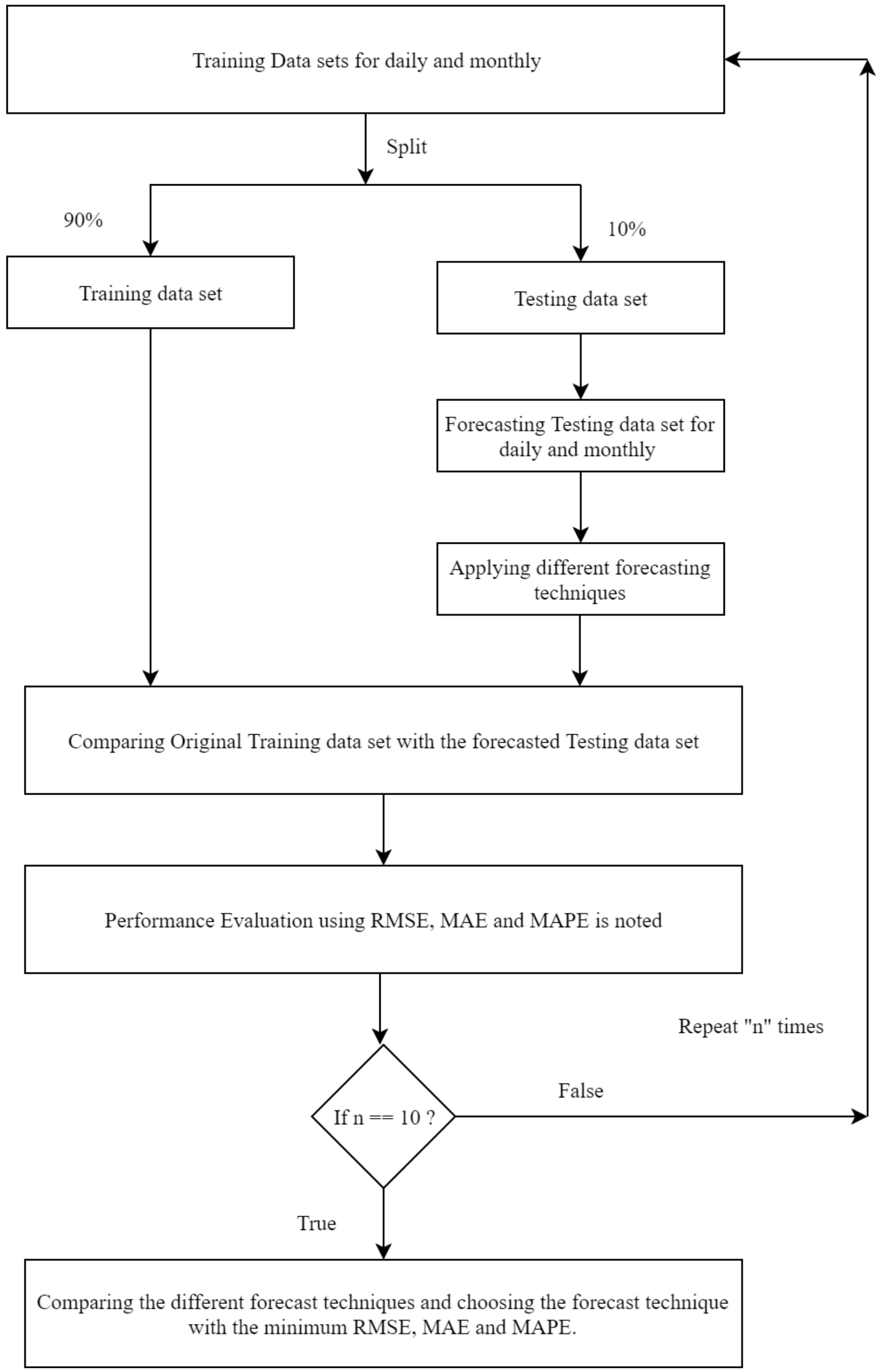
In this section, we are summarizing the first impediment in computation unit of CPS. This research,seeks to develop a set of principles that guide the selection of the best algorithm for the classification based on the datasets and imputation techniques. Comparing performance evaluation (RMSE) of imputation techniques is shown in Table 4.1 to Table 4.6 The performance evaluation factors were determined as missingness, datasets, and imputation techniques. This enables us to find out if water is under supplied and if excess water is being supplied in areas where the imputation method is used to complete missing data observed in the datasets obtained from various areas. In this study, we made use of various imputation techniques such as K-Nearest Neighbors (K-NN) imputation method, Expectation-Maximization imputation method, and the Matrix Completion imputation method. It can be easily understood from the plots that the Matrix completion method is best for imputing missing values at random. The best way to select the best imputation model is to find the model with the smallest RMSE, MAE and MAPE computed. Different performance measures will lead to unlikely results and determine the best method for forecasting. In this case, the performance assessment techniques used in the matrix completion is the best when compared with other imputing methods for these datasets. In the same way, for imputing missing values at burst, matrix completion will be enhanced for less absent percentage imputation. As the percentage increases, matrix completion is not effective, so expectation-maximization is best for imputing missing value at burst. The results of this study suggest that the optimal imputation method that improves the accuracy of computing applications according to the characteristics of the datasets can be Matrix Completion or Expectation-Maximization imputation techniques.

4.2.6 Experimental procedure for daily and monthly model

In this study, we have collected and utilized a dataset from the BWSSB, Government of Karnataka, India and analysed water flow rate values reportedly at every 15 minutes. Before building the model, it was critical to pre-process the data sets received from the BWSSB in the proper method.
The experimental set up of forecasting model for instantaneous and average procedure is presented in Figure 4.5.



(a) Instantaneous (one step forecasting)



(b)Average (10 cross validation)

Figure 4.5: Experimental procedure for forecasting computation set up for (a) Instantaneous (one step forecasting) and (b) Average (10 cross validation).

Algorithm for Experimental procedure for forecasting computation set up for Instantaneous (one step forecasting) and Average (10 cross validation) shown in Figures 4.5 (a) and (b) is listed below:

- Cleaning the data sets.
- Keeping original training data set as reference. Splitting the original data set into a training data set containing the first 90% of instances, and the remaining 10% forms the testing data set.
- Apply different forecasting techniques for the 10% testing data set.
- Comparing the original training data set with testing data set with forecasted values.
- Performance is measured using root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).
- Comparing different forecasting techniques and choosing forecast technique with minimum RMSE, MAE and MAPE values (Figure. 4.5 (a)).
- Repeat the whole procedure for 10 times, and record the average RMSE, MAE and MAPE (10 cross-validation).
- Comparing different forecasting techniques and choosing forecast technique with minimum RMSE, MAE and MAPE values (Figure. 4.5 (b)).

4.3 METHODS FOR FORECASTING

The suitable utilization of forecasting methods normally depends on the type (classification) of data used (Hyndman *et al.* (2007)). In our research work, there was a need to extend a model of forecasting for water flow rate that is altering over time in each storage tank. Time series forecasting is helpful when we are forecasting data that are changing over time (Adhikari and Agrawal (2013)) . Time series forecasting is a set of numerical data obtained at a regular period of time. Here, period refers to quarterly, yearly, monthly and daily.Forecasting would be carried out using R Software.

The quantitative component (i.e. the daily consumption prediction) is forecasted and the pattern mode is estimated using the following techniques as discussed below.

4.3.1 Simple Moving Average

Simple moving average (SMA) is based on a simple idea that is to forecast future data points by using an average of several past data points (Hansun (2013), Lauren and Harlili (2014), Ulrich *et al.* (2018)). This method is appropriate when there is no trend or seasonality. The SMA for time period t is the mean of the "k" most recent observations. At the beginning, the constant number k is set. The smaller the number k, the greater weight is given to recent periods. The higher the number k,, the less weight is given to more recent periods.When large, rare fluctuations in the series are present, a large k is desirable.If there are sudden series shifts, a small k is the most desirable. A moving average of order k, MA(k) is the value of k consecutive observations

$$F_{t+1} = \hat{y}_{t+1} = \frac{(y_t + y_{t-1} + y_{t-2} + ... + y_{t-k+1})}{k} \tag{4.5}$$

where, k is the number of terms in the moving average Hansun (2013).

4.3.2 Holt and –Winters

Holt–Winters (HW) is also known as triple exponential smoothing. It is the advanced exponential smoothing superior to simple exponential smoothing (Ekberg *et al.* (2011),Elmunim *et al.* (2015), Rubab *et al.* (2015), Yang *et al.* (2017), Hyndman *et al.* (2019)). The idea behind SES is to forecast future values by using a weighted average of all previous values in the series. The main idea in advanced exponential smoothing is taking the simple exponential smoothing set set-up and expanding it to capture trend and/or seasonality.
Forecast = level + trend + seasonality.

$$F_{t+k} = L_t + kT_t + S_{t+k-M} \tag{4.6}$$

where, L_t is the most recent estimate level,
T_t is the most recent trend,
k is how many steps into the future trying to forecasts,
S_t is the Seasonality with M seasons Ekberg *et al.* (2011),Elmunim *et al.* (2015), Rubab *et al.* (2015).

4.3.3 Autoregressive Integrated Moving Average Model

Autoregressive integrated moving average (ARIMA) is a forecasting technique that projects the future values of a series based entirely on its own inertia (Akpinar and Yumusak (2013), Skopal (2015), Hyndman *et al.* (2019), Du (2018)). It is one of the finest methodologies for doing the time series forecasting which means, forecasting for the data that is distributed over time. ARIMA has become powerful because it is based on the assumption that over a period of time the current values are related or correlated with their immediate previous or "n" previous values.
ARIMA (p, d, q): where p = values from partial auto correlation, d = lagged difference between current and previous values and q = value from auto correlation Hyndman *et al.* (2019).

4.3.4 Neural Network

Neural network (NN) is basically a computing system which is designed to simulate the way the human brain analyses and processes the information (Crone and Kourentzes (2010)). NN has self-learning capabilities that enable it to produce better result when more data are available. NN consists of three types of nodes.

- Input node: also known as "input layer," is used to collect the data from the user to the network. Input layer transmits the data to the hidden nodes. No computation function is carried out in any of the input layers.

- Hidden node: has no interconnection with the user, hence the name "hidden" is used. It is also referred as "hidden layer." Hidden layer executes computation function and transfers the data to the next node or layer. Feedforward network will only have a single input layer, multiple hidden layer and single output layer.

- Output node: also known as "output layer," is used to receive the computation function result from the hidden node and transfer that result from the network to the user.

4.3.4.1 Multilayer Perceptron

Multilayer perceptron (MLP) contains one or more hidden layers which emphasize on the computation function, input layer which gives prominence to the user to receive the data and an output layer which generates the resulting outputs (Ord *et al.* (2017)). Any layers connecting between input and output layer are referred as hidden layers.

4.3.5 Proposed Hybrid Model

In this work, the proposed hybrid model is designed using Package opera (Online Prediction using Expert Aggregation) in R studio software (Gaillard).The mixture function and predict function of package opera are used. Mixture function is used to create the algorithm object and the function predict is used to predict the forecast values using the created algorithm. The proposed hybrid model (ARIMA + HW) is designed to take into consideration the instantaneous and cross-validation performance combining both HW and ARIMA. To make time series prediction more accurate, the model blends the ARIMA model with HW model into an integrated structure.The combination (ensemble) of ARIMA and HW is done by the MLpoly algorithm (Gaillard). The final forecast can be modeled using the following equation

$$HM = Output(A, HW) \tag{4.7}$$

Where HM is the output of the proposed hybrid model, A is the forecast output value from auto.arima, HW is the forecast output value from HW.

4.3.6 Result

The performance evaluation of forecasting technique using R software is tested by the following error measures: RMSE, MAE and MAPE (Herrera *et al.* (2010),Zheng *et al.* (2015), Wu *et al.* (2017))

- RMSE
  RMSE can be measured as,

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(Y_i - \hat{Y}_i)^2}{n-1}} \tag{4.8}$$

  where: Y_i= Forecast value,
  $\hat{Y}_i$ = Actual value,
  n= Number of values in the data sets.(Nookhong and Kaewrattanapat (2015), Torgo and Torgo (2013)) .

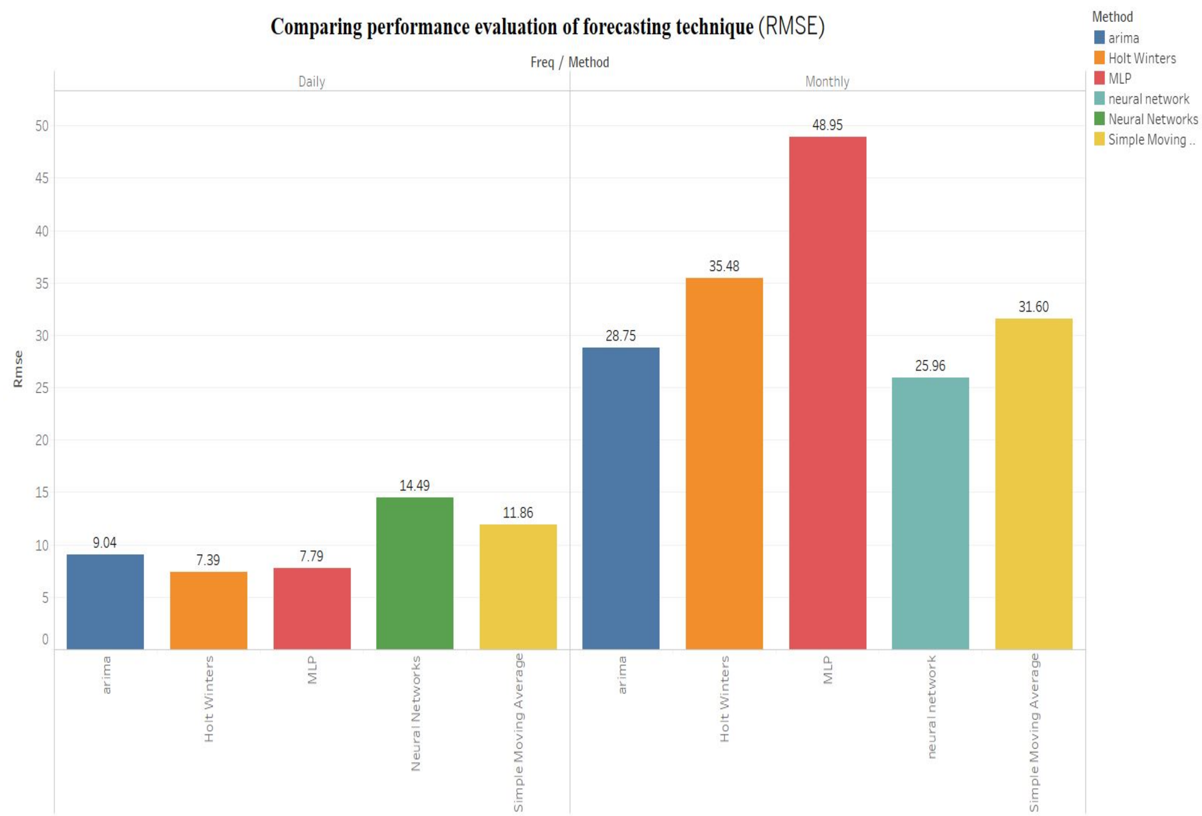


Figure 4.6: Comparing daily versus monthly performance evaluation (RMSE) of forecasting techniques(Instantaneous)

- MAE

  MAE can be measured as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \tag{4.9}$$

where, f_i = Forecast value ,

y_i = Actual value,

n = Number of values in the data sets.(Nookhong and Kaewrattanapat (2015),Torgo and Torgo (2013)).

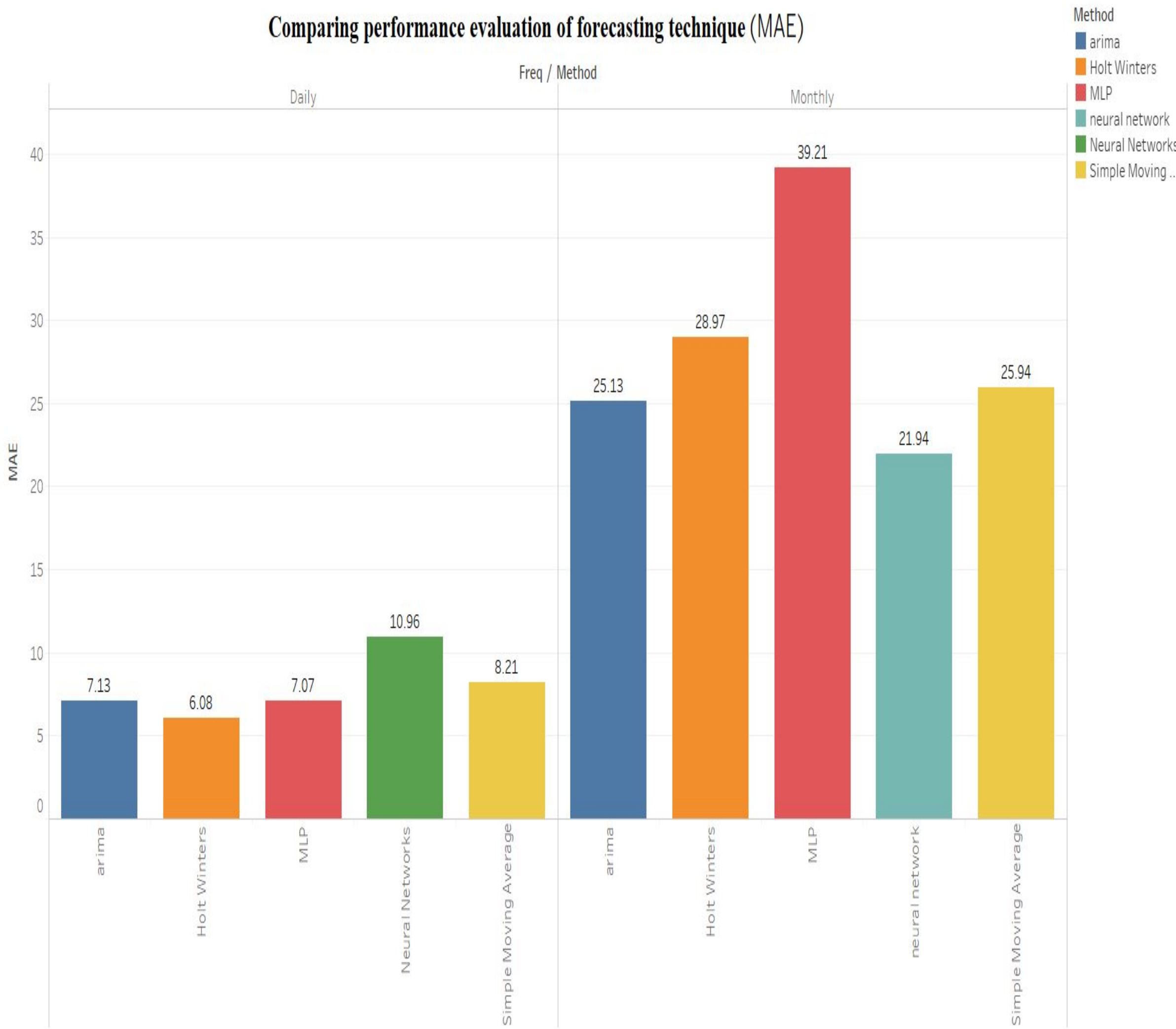


Figure 4.7: Comparing daily versus monthly performance evaluation (MAE) of forecasting techniques(Instantaneous)

- MAPE

  MAPE is measured as,

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - f_t}{A_t} \right| \tag{4.10}$$

  As, A_t= Actual Value,

  f_t= Forecast value,

  n= Number of values in the data sets.(Torgo and Torgo (2013)).

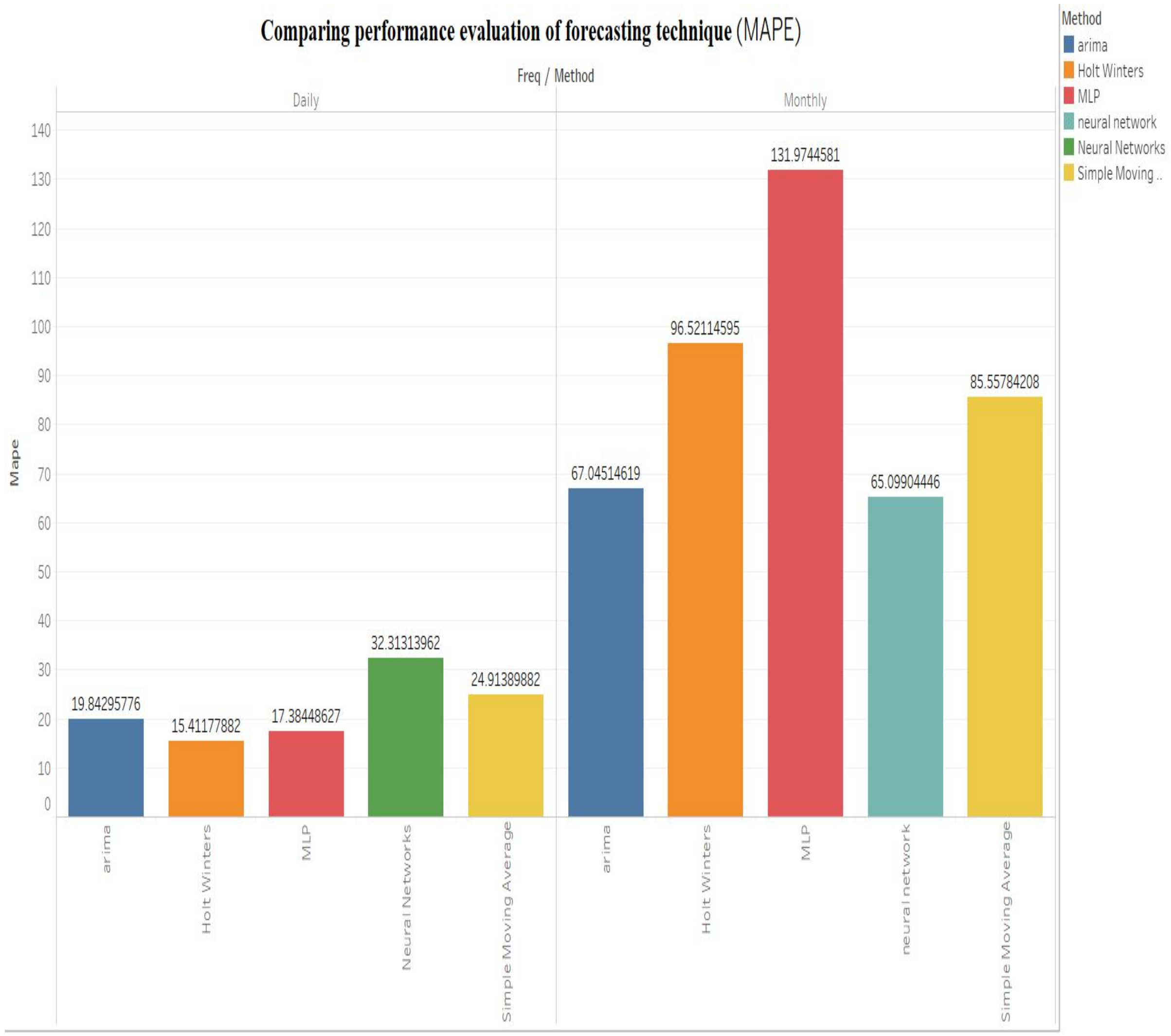


Figure 4.8: Comparing daily versus monthly performance evaluation (MAPE) of forecasting techniques (Instantaneous)

The results shown in Figures 4.6 to 4.8 represents the performance evaluation of forecasting techniques (RMSE,MAE and MAPE) for daily and monthly model (Instantaneous). These graphs portray the collective data for different values of SMA, HW, ARIMA, NN and MLP. In the graph, we compare daily and monthly model for different forecasting techniques against different performance evaluation RMSE, MAE and MAPE respectively. We observe that in Table 4.7, for monthly model RMSE, MAE and MAPE is higher when compared with daily model. So for our application daily model is best suited.

When (10 cross validation) averaging is done, ARIMA does well compared with

Table 4.7: Performance evaluation (Instantaneous) of forecasting techniques

Methods of forecasting	Performance evaluation (Instantaneous)					
	RMSE		MAE		MAPE	
	Daily	Monthly	Daily	Monthly	Daily	Monthly
ARIMA	9.04	28.75	7.13	25.13	19.84	67.04
HW	7.39	35.48	6.08	28.97	15.41	96.52
SMA	11.86	31.60	8.21	25.94	24.91	85.55
NN	14.49	25.96	10.96	21.94	32.31	65.09
MLP	7.79	48.95	7.07	39.21	17.38	131.97

Table 4.8: Performance evaluation (10- fold validation) of forecasting techniques

Methods of forecasting	Performance evaluation (10- fold validation)		
	Average RMSE	Average MAE	Average MAPE
ARIMA	17.89	15.08	30.91
HW	32.06	27.03	52.92
SMA	18.88	15.28	31.99
NN	20.07	16.51	37.57
MLP	20.40	16.81	37.42

other time series forecasting techniques as seen in Table 4.8 . RMSE, MAE and MAPE is comparatively low in ARIMA as compared with other techniques . Cross validation is the measurement of performance calculated average classification accuracy of "n" rounds of random selection of different data. When So for our application we will consider ARIMA forecasting model.

From Fig. 4.6,4.7,4.8 we see that HW model is best suited for instantaneous performance. Fig.4.9 shows the performance of the proposed ARIMA-HW model in comparison with ARIMA, HW, SMA, NN and MLP models. One can observe that in case of daily data the proposed ARIMA-HW model is more accurate than its basis time series models. Thus, ARIMA-HW model can be used as an appropriate alternative model when higher forecasting accuracy is needed.

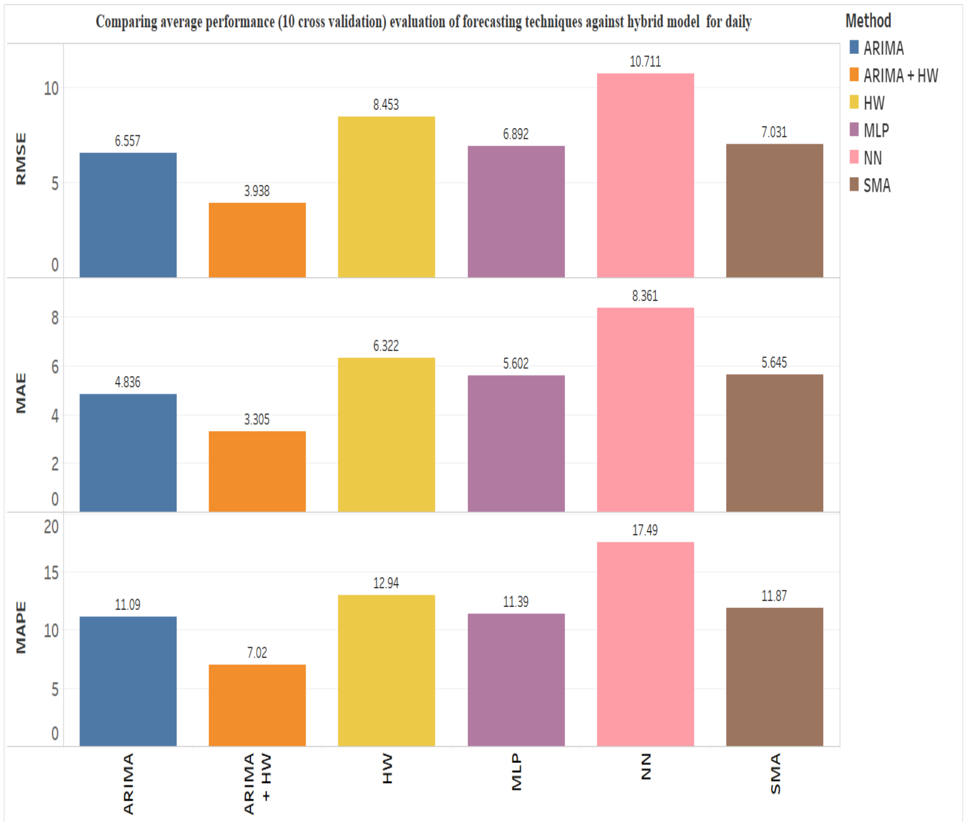


Figure 4.9: Comparing average performance evaluation of forecasting techniques (10 cross validation) for daily

4.4 SUMMARY

Here our work, is intended for time series data based water flow rate forecasting, utmost prominent predictive models implied in Cyber-Physical Systems are compared. Data forecasting concern was investigated through simulations. At this juncture, we have considered and compared water flow rate data by both daily (one day) forecasting and monthly forecasting in a storage tank. Data of both daily and monthly forecasting were compared through instantaneous one step forecasting and the consequences of the same showcased daily forecasting as precise and accurate for our application.

Performance of traditional forecasting methods like ARIMA, HW, MLP, NN, and SMA, is evaluated on the considered time series data sets. Simulation results show that Holt-Winters method performs better when compared with other existing methods (in both daily and monthly data).In illustration and comparison of instantaneous and multistep validation (10 fold) forecasting's, HW model and ARIMA models are perceived to foretell with low RMSE, low MAE and low MAPE, as compared to other models. An innovative hybrid model that is collaborative of ARIMA and HW has been projected. Simulation results of the same point toward that the hybrid model is more precise in comparison with its basis time series model. This puts forward that the anticipated and recommended ARIMA-HW forecasting technique aids us in taking essential acts to popular supervisory activities (Computation unit) in CPS.

Chapter 5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Conclusion

The objective of the work undertaken was to provide a scalable and robust water supply management framework through Cyber Physical System (CPS) perspective. The major issue observed in the existing water distribution scheme was that of water wastage arising on everyday basis without anyone even being conscious of it. As an overall cascading effect, improper water distribution was leading to shortage. The pilot project under which the present research has been carried for mitigating the wastage still uses the existing storage tanks in the campus as the supply to end-user. The involved CPS framework consists of four major functional work flow: The Monitoring Unit, the Networking Unit, and the Computation Unit and Actuation unit.
The core utility of the CPS in our research work is to monitor and collect data (inflow and outflow quantities) from each storage tank through Meter Interface Unit on a regular basis, estimate the requirement through predictive analytics, and manage the distribution based on real time data. This is done by utilizing "minimal" wireless technology to screen and control the water flow rate in a storage tank and stretching from scale set up to large region efficiently.
The following are some of the contributions of the work.

- Monitoring Unit
  Systematic approach of customizing the existing water pipeline infrastructure

using minimum wireless technology to monitor the water distribution through a storage tank. Customization includes a set up of RF transceiver modules interfaced with water meter which has been installed to get the water inflow and outflow details for each pipe.

- The Networking Unit
  We have developed a process automation solution that counts the number of pulses from water meter and send to the Aggregator through RF module; Aggregator read command interrupt and also count variable are initialized to 0 after it reaches to maximum count value or after getting reset command from Aggregator. The connection between DTU , Aggregator and the web server were integrated to function as needed. The transmission of data between the RF modules, DTU, aggregator, and finally the web server is established for success and failure upon request and response signals (byte stream). Finally, we are able to display the water flow details from the inlet and outlet and also net volume of the storage tank in the web server; the necessary functionality is built during the execution of research.

- The Computation Unit
  The Computation Unit of the proposed CPS framework is mainly meant for data analysis among others. Through elaborate and systematic simulation study, missing values in the data and the flow forecasting has been investigated. In Handling missing values, different techniques are tried to impute the values, by introducing random and burst-natured (continuously) missing values for different percentage missingness. The performance is assessed based on the popular measures available in the literature, and it is found through studies that existing techniques are reasonably good towards handling the missing data. This research, seeks to develop a set of principles that guide the selection of the best algorithm for the imputation of missing values. This enables us to find out if water is under supplied and if excess water is being supplied in areas where the imputation method is used to complete missing data observed in the datasets obtained from various areas.
  In the next phase, Performance of traditional forecasting methods like ARIMA, HW, MLP, NN, and SMA, is evaluated on the considered time series data sets. Simulation results show that HW method performs better when compared with

other existing methods (for both daily and monthly data). Additionally, it was observed that the performance of all the models is better when trained on daily data as against to the data obtained for a month. In the case of cross validation (10 fold) ARIMA method has been observed to offer best results as compared to other existing methods, implying that ARIMA model is best suited for the forecasting application. A new hybrid model that is obtained from ARIMA and HW is observed to work more accurately than its time series model; this is another novel contribution of the research. This forecasting technique helps us to take necessary action to complete the decision-making activities (Computation unit) in CPS to conserve water, scheduling of pump, leakage detection and distribution of water when needed most.

5.2 Future work

The future work can be aimed towards CPS simulation to work out better Water Management Strategies. This work estimates the methods used on a small-scale three storage tanks scenario. The scalability and sustainability needs to be tested using multiple storage tanks that would require collating larger data sets and conveying the information about water supply schedules to end-users.

- In computation unit, Anomalies in expected water distribution data such as missing data and forecasting may require concerted effort based on the model presented in this work. We anticipate rigorous ML algorithm, other sophisticated methods as an effective methodology to be deployed based on the processes and tools provided in this work. Also we can upgrade the decision support system to alert the consumers about the water consumption for optimizing pumping schedules of each storage tank.

- Event Detection.
  Events correspond to interesting and useful information and may correspond to anomalies and leakage. There is a good scope for working out novel algorithms. Further, classifying events involve working out appropriate machine-learning algorithms, specific to our scenario.

- Leakage Detection
  Leakage detection can be carried out through flow balancing to start with. There

is a further scope to arrive at methods based on statistical inference.

Actuation unit is used to actuate the control valve determined by the computing phase.

- Generating Control input to electro-mechanical valve (for Automatic control). Using the events appropriate control inputs can be generated. Here, a state-space formulation can be explored to bring the control aspect.

Towards getting some understanding of water distribution and management for the entire campus, we would be considering the simulation studies, knitting together different models. For example, interaction between different tanks and optimizing the distribution for the entire campus can be considered in the future. Our aim is to implement CPS to reduce water scarcity issue in a campus and later extending to large area efficiently. So that, Water wastage can be reduced and utilization of water can be improved.

Bibliography

2b1 consulting (2018). Cyber physical systems (cps). URL `https://www.2b1stconsulting.com/cyber-physical-systems-cps/`.

Adhikari, R. and **R. K. Agrawal** (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.

Ahmed, N. K., **A. F. Atiya**, **N. E. Gayar**, and **H. El-Shishiny** (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, **29**(5-6), 594–621.

Akpinar, M. and **N. Yumusak**, Forecasting household natural gas consumption with arima model: A case study of removing cycle. *In Application of Information and Communication Technologies (AICT), 2013 7th International Conference on*. IEEE, 2013.

Alvisi, S., **M. Franchini**, and **A. Marinelli** (2007). A short-term, pattern-based model for water-demand forecasting. *Journal of hydroinformatics*, **9**(1), 39–50.

Bakker, M., **H. Van Duist**, **K. Van Schagen**, **J. Vreeburg**, and **L. Rietveld** (2014). Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering*, **70**, 93–102.

Baroni, L., **L. Cenci**, **M. Tettamanti**, and **M. Berati** (2007). Evaluating the environmental impact of various dietary patterns combined with different food production systems. *European journal of clinical nutrition*, **61**(2), 279.

Bralower, T. (). Earth 103: Earth in the future. distribution of water on the earth's surface. URL `https://www.e-education.psu.edu/earth103/node/701`.

Chen, J. and **D. Boccelli** (2014). Demand forecasting for water distribution systems. *Procedia Engineering*, **70**, 339–342.

Christian-Smith, J., **M. Heberger**, and **L. Allen** (2012). Urban water demand in california to 2100: incorporating climate change. *Pacific Institute*.

Crone, S. F. and **N. Kourentzes** (2010). Feature selection for time series prediction–a combined filter and wrapper approach for neural networks. *Neurocomputing*, **73**(10-12), 1923–1936.

Donders, A. R. T., **G. J. Van Der Heijden**, **T. Stijnen**, and **K. G. Moons** (2006). A gentle introduction to imputation of missing values. *Journal of clinical epidemiology*, **59**(10), 1087–1091.

Du, Y., Application and analysis of forecasting stock price index based on combination of arima model and bp neural network. *In 2018 Chinese Control And Decision Conference (CCDC)*. IEEE, 2018.

Ekberg, J., **J. Ylinen**, and **P. Loula**, Network behaviour anomaly detection using holt-winters algorithm. *In Internet Technology and Secured Transactions (ICITST), 2011 International Conference for*. IEEE, 2011.

Elmunim, N., **M. Abdullah**, **A. M. Hasbi**, and **A. Zaharim**, Forecasting ionospheric delay during quiet and disturbed days using the holt-winter method. *In Space Science and Communication (IconSpace), 2015 International Conference on*. IEEE, 2015.

Eswaran, P. and **A. Kumar**, Design and development of automated water distribution and metering system for residential buildings. *In 2012 IEEE International Conference on Computational Intelligence and Computing Research*. IEEE, 2012.

Experts, W. P. (). Facts.

Gaillard, P. (). Opera, year = 2016, url = http://pierre.gaillard.me/opera.html, urldate = 2019.

GAO, U. (2014). Freshwater supply concerns continue, and uncertainties complicate planning.

Geographic, N. (). Freshwater crisis.

Hansun, S., A new approach of moving average method in time series analysis. *In New Media Studies (CoNMedia), 2013 Conference on*. IEEE, 2013.

Hastie, T., **R. Mazumder**, **J. D. Lee**, and **R. Zadeh** (2015). Matrix completion and low-rank svd via fast alternating least squares. *The Journal of Machine Learning Research*, **16**(1), 3367–3402.

Herrera, M., **L. Torgo**, **J. Izquierdo**, and **R. Pérez-García** (2010). Predictive models for forecasting hourly urban water demand. *Journal of hydrology*, **387**(1-2), 141–150.

Honaker, J., **G. King**, **M. Blackwell**, and **M. M. Blackwell** (2010). Package 'amelia'.

Hyndman, R. J., **G. Athanasopoulos**, **C. Bergmeir**, **G. Caceres**, **L. Chhay**, **M. O'Hara-Wild**, **F. Petropoulos**, and **S. Razbash** (2019). Package 'forecast'. *Online] https://cran. r-project. org/web/packages/forecast/forecast. pdf*.

Hyndman, R. J., **Y. Khandakar**, *et al.*, *Automatic time series for forecasting: the forecast package for R*. 6/07. Monash University, Department of Econometrics and Business Statistics, 2007.

Igor, S. (1993). World fresh water resources. *Water in crisis: a guide to the world's. Oxford University Press, Inc, Oxford*.

Imen, S. and **N.-B. Chang**, Developing a cyber-physical system for smart and sustainable drinking water infrastructure management. *In 2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC)*. IEEE, 2016.

Institute, T. W. R. and **C. W. Centre** (2016). The World Resources Institute and Columbia Water Centre the india water tool. URL `http://www.indiawatertool.in`.

Jain, A., **A. K. Varshney**, and **U. C. Joshi** (2001). Short-term water demand forecast modelling at iit kanpur using artificial neural networks. *Water resources management*, **15**(5), 299–321.

Khashei, M. and **M. Bijari** (2012). A new class of hybrid models for time series forecasting. *Expert Systems with Applications*, **39**(4), 4344–4357.

Kourentzes, N., **D. K. Barrow**, and **S. F. Crone** (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, **41**(9), 4235–4244.

Kowarik, A. and **M. Templ** (2016). Imputation with the r package vim. *Journal of Statistical Software*, **74**(7), 1–16.

Lakshminarayan, K., **S. A. Harp**, and **T. Samad** (1999). Imputation of missing data in industrial databases. *Applied intelligence*, **11**(3), 259–275.

Laña, I., **I. I. Olabarrieta**, **M. Vélez**, and **J. Del Ser** (2018). On the imputation of missing data for road traffic forecasting: New insights and novel techniques. *Transportation research part C: emerging technologies*, **90**, 18–33.

Lauren, S. and **S. D. Harlili**, Stock trend prediction using simple moving average supported by news classification. *In Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of*. IEEE, 2014.

Lee, E. A. and **S. A. Seshia**, *Introduction to embedded systems: A cyber-physical systems approach*. Mit Press, 2016.

Lin, J., **S. Sedigh**, and **A. Miller**, Integrated cyber-physical simulation of intelligent water distribution networks. *In Scientific and Engineering Applications Using MATLAB*. IntechOpen, 2011.

Lopez Farias, R., **V. Puig**, **H. Rodriguez Rangel**, and **J. J. Flores** (2018). Multi-model prediction for demand forecast in water distribution networks. *Energies*, **11**(3), 660.

Makridakis, S., **E. Spiliotis**, and **V. Assimakopoulos** (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, **13**(3), e0194889.

Maxwell-Gaines, C. (2012). Water infographic: Alarming statistics about wasted water in our communities. URL `https://www.watercache.com/blog/2012/07/water-infographic-with-water-waste-statistics-from-leaky-pipes`.

McNabb, D. E., Pathways to supplemental water resources. *In Global Pathways to Water Sustainability*. Springer, 2019, 233–250.

Mutikanga, H. E., **S. K. Sharma**, and **K. Vairavamoorthy** (2012). Methods and tools for managing losses in water distribution systems. *Journal of Water Resources Planning and Management*, **139**(2), 166–174.

Nelwamondo, F., **S. Mohamed**, and **T. Marwala** (2007). Missing data: A comparison of neural network and expectation maximisation techniques, eprint. *arXiv preprint arXiv:0704.3474*.

Nicolini, M. and **A. Patriarca**, Model calibration and system simulation from real time monitoring of water distribution networks. *In 2011 3rd International Conference on Computer Research and Development*, volume 1. IEEE, 2011.

Nookhong, J. and **N. Kaewrattanapat** (2015). Efficiency comparison of data mining techniques for missing-value imputation. *Journal of Industrial and Intelligent Information*, **3**(4).

Ord, K., **R. Fildes**, and **N. Kourentzes**, *Principles of Business Forecasting–2nd Ed*. wessex, Incorporated, 2017. ISBN 9780999064917. URL `https://books.google.co.in/books?id=JKJLtAEACAAJ`.

Organization, W. H., **W. J. W. Supply**, and **S. M. Programme**, *Progress on sanitation and drinking water: 2015 update and MDG assessment*. World Health Organization, 2015.

Pillai, U., **V. Murthy**, and **I. Selesnick**, Missing data recovery using low rank matrix completion methods. *In Radar Conference (RADAR), 2012 IEEE*. IEEE, 2012.

Pires, I. M., **N. M. Garcia**, **N. Pombo**, and **F. Flórez-Revuelta** (2016). From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. *Sensors*, **16**(2), 184.

Prantner, B. (2011). Visualization of imputed values using the r-package vim.

Ramachandran, M. (). Guidance Notes for continuous water supply (24-7 supply) - A guide to project preparation implementation and appraisal: A report by water and sanitation program, year = 2014, url = http://sanitation.indiawaterportal.org/english/node/2351., urldate = 2019.

Reddy, B. P. and **P. Eswaran**, Metro overhead tanks monitoring system using zigbee based wsn. *In 2013 IEEE Conference on Information & Communication Technologies*. IEEE, 2013.

Rinaudo, J.-D., Long-term water demand forecasting. *In Understanding and managing urban water in transition*. Springer, 2015, 239–268.

Rubab, S., **M. F. Hassan**, **A. K. Mahmood**, and **S. N. M. Shah**, Forecasting volunteer grid workload using holt-winters' method. *In Technology Management and Emerging Technologies (ISTMET), 2015 International Symposium on*. IEEE, 2015.

Skopal, R., Short-term hourly price forward curve prediction using neural network and hybrid arima-nn model. *In Information and Digital Technologies (IDT), 2015 International Conference on*. IEEE, 2015.

Smith, M. and **R. Agrawal**, A comparison of time series model forecasting methods on patent groups. *In MAICS*. 2015.

Suresh, M., **U. Manohary**, **A. G. Ry**, **R. Stoleru**, and **M. K. M. Sy**, A cyber-physical system for continuous monitoring of water distribution systems. *In 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2014.

Suri, P., **V. Yadav**, *et al.*, Design and implementation of wireless sensor network architecture for leak detection and monitoring in water supply distribution network. *In 2017 2nd International Conference for Convergence in Technology (I2CT)*. IEEE, 2017.

Templ, M., **A. Alfons**, **A. Kowarik**, and **B. Prantner** (2011). Vim: visualization and imputation of missing values. *R package version*, **2**(3).

TM, T., *Tarang-Product Manual-Melange System Pvt Ltd*. 2010. URL `www.melangesystems.com`.

Torgo, L. and **M. L. Torgo** (2013). Package 'dmwr'. *Comprehensive R Archive Network*.

UC Berkeley, D. E. (). Ptolemy project.

Ulrich, J., **M. J. Ulrich**, and **S. RUnit** (2018). Package 'ttr'.

UNESCO, W. (2006). Unesco world water assessment programme world water development report 2–executive summary.

(USEPA), U. S. E. P. A. (). Watersense.

van Lier, B. (2013). Cyberspace, cybernetics and cyber-physical systems. URL `https://www.2b1stconsulting.com/cyber-physical-systems-cps/`.

Wan, J., **M. Chen**, **F. Xia**, **D. L. 0001**, and **K. Zhou** (2013). From machine-to-machine communications towards cyber-physical systems. *Comput. Sci. Inf. Syst.*, **10**(3), 1105–1128.

Wang, E. K., **Y. Ye**, **X. Xu**, **S.-M. Yiu**, **L. C. K. Hui**, and **K.-P. Chow**, Security issues and challenges for cyber physical system. *In 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*. IEEE, 2010.

Wang, Z., **H. Song**, **D. W. Watkins**, **K. G. Ong**, **P. Xue**, **Q. Yang**, and **X. Shi** (2015). Cyber-physical systems for water sustainability: challenges and opportunities. *IEEE Communications Magazine*, **53**(5), 216–222.

WATER, S. T. (). 250 water facts.

water science school, T. U. (). Where is earth's water? URL `https://www.usgs.gov/special-topic/water-science-school/science/where-earths-water?qt-science_center_objects=0#qt-science_center_objects`.

Wiener, N., *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25. MIT press, 1965.

Wiesner, S. and **K.-D. Thoben**, Cyber-physical product-service systems. *In Multi-Disciplinary Engineering for Cyber-Physical Production Systems*. Springer, 2017, 63–88.

Wu, F.-J., **Y.-F. Kao**, and **Y.-C. Tseng** (2011). From wireless sensor networks towards cyber physical systems. *Pervasive and Mobile computing*, **7**(4), 397–413.

Wu, G.-Z., **K. Sakaue**, and **S. Murakawa** (2017). Verification of calculation method using monte carlo method for water supply demands of office building. *Water*, **9**(6), 376.

Wubetie, H. T. (2017). Missing data management and statistical measurement of socio-economic status: application of big data. *Journal of Big Data*, **4**(1), 47.

Yang, Y.-m., **H. Yu**, and **Z. Sun**, Aircraft failure rate forecasting method based on holt-winters seasonal model. *In Cloud Computing and Big Data Analysis (ICC-CBDA), 2017 IEEE 2nd International Conference on*. IEEE, 2017.

Zhang, Q., **A. Rahman**, and **C. D'este**, Impute vs. ignore: Missing values for prediction. *In Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, 2013.

Zheng, Y.-L., **L.-P. Zhang**, **X.-L. Zhang**, **K. Wang**, and **Y.-J. Zheng** (2015). Forecast model analysis for the morbidity of tuberculosis in xinjiang, china. *PloS one*, **10**(3), e0116832.

Publications based on the thesis

International Conferences

1. Abhishek MB, N. Shekar V. Shet, “Wirelessly monitoring of water usage in smart campus: A CPS perspective”, International Engineering Symposium, Kumamoto University , Japan, March 2017.

2. N. Shekar V. Shet, H. M. P., Naveen Kumar Panguluri and Abhishek MB, “Wireless monitoring of water resources in a campus”, International Engineering Symposium (IES-2017). Kumamoto University, Japan, 2017

3. Abhishek, M.B and N. S. V. Shet . ”Data transmission unit and web server interaction to monitor water distribution: A cyber-physical system perspective”, 4th International conference on Earth Science and Engineering (ICEE), Andalas University, Padang, Indonesia ,2017.

4. Abhishek MB, N.S.V.Shet , ”Data Processing and Deploying Missing Data Algorithms to Handle Missing Data in Real Time Data of Storage Tank: A Cyber Physical Perspective”,IEEE ICECIE’2019,Kuala Lumpur, Malaysia. (**Accepted**).

5. Abhishek MB, N.S.V.Shet , ”Addressing the functioning of the networking unit in Cyber Physical System for smart water management”, Global Conference On ADVANCED SMART AND SUSTAINABLE TECHNOLOGIES IN ENGINEERING (GCASSTE 2020), MITE, India.(**Accepted**).

Journals

1. Abhishek, M.B and N. S. V. Shet (2018). ”Data transmission unit and web server interaction to monitor water distribution: A cyber-physical system perspective”, International Journal on Advanced Science, Engineering and Information Technology, 8(4), 1307-1312.(Scopus Indexed Journal).

2. MB Abhishek, N. Shekar V. Shet, "Cyber physical system perspective for smart water management in a campus", Desalination and Water Treatment, volume-147 (2019) 296–307, doi:10.5004/dwt.2019.23691. (SCIE Indexed journal).

3. Abhishek MB, N. Shekar V. Shet, "Performance Evaluation of Predictive models for Water Flow Rate in Storage tank : A Cyber Physical System Perspective.",Wireless Personal Communications, Springer **(Under Review)**(SCIE Indexed journal).

Bio-data (FOR PHD)

Name	: Abhishek MB
Address	:DOOR No. 658/10, "SRISHTHI NILAYA", 4th CROSS, AREKERE, BANNERGHATTA ROAD, BENGALURU, KARNATAKA, INDIA.
Email	: abhishek.mb@gmail.com
Mobile	: +91-9886601331.
Qualification	: Bachelor of Engineering(Visvesvaraya Technological University), Kammavari Sangha Institute of Technology, Bengaluru, Karnataka, India. Master of Technology (Visvesvaraya Technological University), Dayananda Sagar College of Engineering , Bengaluru, Karnataka, India.